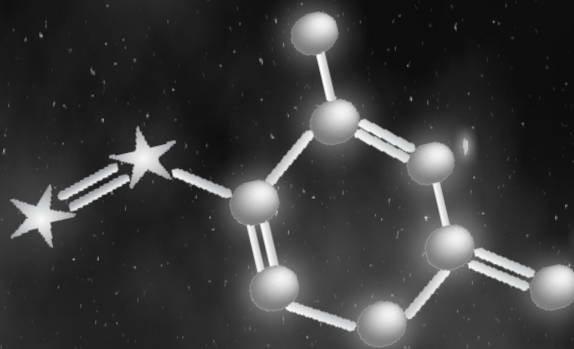
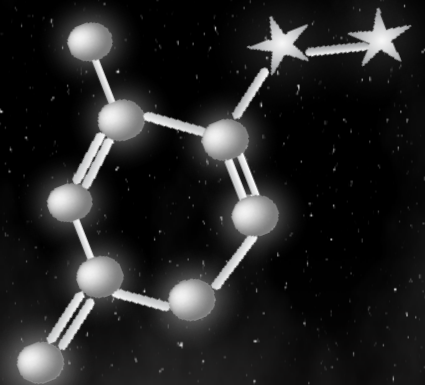
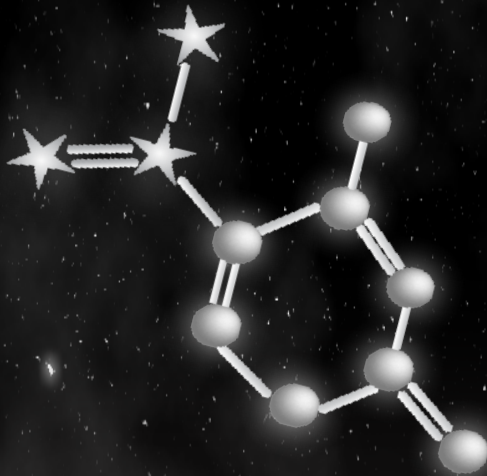


# Cytosine Base Modifications, DNA Metabolism and Development

by

Cathia Rausch





---

# Cytosine Base Modifications, DNA Metabolism and Development

---



vom Fachbereich Biologie der Technischen Universität Darmstadt  
zur Erlangung des akademischen Grades  
Doctor rerum naturalium  
(Dr. rer. nat.)

Dissertation von  
**Cathia Pascale Fabienne Rausch, M.Sc. Biologie**  
aus Luxemburg

Erstgutachterin: Prof. Dr. M. Cristina Cardoso

Zweitgutachterin: Prof. Dr. Ulrike A. Nuber

Tag der Einreichung: 05.07.2021

Darmstadt 2021

Rausch, Cathia Pascale Fabienne:

Cytosine Base Modifications, DNA Metabolism and Development

Darmstadt, Technische Universität Darmstadt,

Jahr der Veröffentlichung der Dissertation auf TUpriints: 2021

URN: urn:nbn:de:tuda-tuprints-195185

Tag der mündlichen Prüfung: 14.09.2021

Veröffentlicht unter CC-BY-ND 4.0 International

<https://creativecommons.org/licenses/>

*Le secret d'ennuyer est celui de tout dire.*

Voltaire



# Contents

<b>1 Summary/Zusammenfassung</b>	<b>1</b>
<b>2 Preface</b>	<b>4</b>
<b>3 General Introduction</b>	<b>6</b>
3.1 DNA as an Information Storage – The Mammalian Genome . . . . .	6
3.2 DNA Replication – Where, When and How . . . . .	14
3.2.1 Where To Start DNA Replication? – Origin Definition . . . . .	14
3.2.2 How To Start DNA Replication? – Origin Licensing and Activation . . . . .	15
3.2.3 How To Replicate the DNA? – The Replisome And Leading versus Lagging Strand . . . . .	18
3.2.4 Where and When To Replicate DNA? – DNA Replication in Time and Space . . . . .	20
3.2.5 How To Organize DNA Replication? – DNA Replication and Chromatin (Loop) Organization .	23
3.2.6 How to Duplicate More Than DNA? – DNA Replication and Epigenetics . . . . .	24
3.2.7 When and How To Reorganize DNA Replication? – DNA Replication and Development . . .	26
3.3 DNA Modification Readers, Writers and Their Interplay . . . . .	28
<b>4 Aims of the Study</b>	<b>46</b>
<b>5 Result Chapters</b>	<b>50</b>
5.1 CHAPTER I: Repli-FISH – Application of 3D-(Immuno)-FISH for the Study of DNA Replication Timing of Genetic Repeat Elements . . . . .	50
5.2 CHAPTER II: DNA Replication and Repair Kinetics of Alu, LINE-1 and Satellite III Genomic Repetitive Elements . . . . .	82
5.3 CHAPTER III: Developmental Differences in Genome Replication Program and Origin Activation .	118
5.4 CHAPTER IV: Cytosine Base Modifications Regulate DNA Duplex Stability and Metabolism . . . . .	192
<b>6 General Conclusions and Perspectives</b>	<b>260</b>
6.1 Replication of the Mammalian Genome – Repeat Elements . . . . .	260
6.1.1 DNA Repeat Elements Are Well Integrated Into the Mammalian Genome . . . . .	260
6.1.2 mES Cells Are Characterized by a Distinct Replication Program . . . . .	261
6.1.3 Replication of the Y Chromosome Marks the End of S-Phase . . . . .	264
6.2 Replication of the Mammalian Genome – Origin Activation and Fork Speed . . . . .	266
6.2.1 Origin Activation in Pluripotent Mouse Cells Differs From That in Somatic Cells . . . . .	266
6.2.2 Loss of DNA Methylation Enhances Replication Fork Speed . . . . .	267
6.3 Cytosine Modifications in the Mammalian Genome – DNA Duplex Stability and Metabolism . . . . .	270
6.3.1 Cytosine Modifications Change the DNA Duplex Stability . . . . .	270
6.3.2 Cytosine Modifications Influence DNA Transcription, Replication and Unwinding Rates . . .	271
6.3.3 Cytosine Modifications Modulate DNA Metabolic Processes by Changing the DNA Stability .	272
6.3.4 Cytosine Modifications and Their Readers and Writers Regulate Genome Stability . . . . .	273
<b>7 Bibliography</b>	<b>274</b>

<b>8 Annex</b>	<b>298</b>
8.1 Abbreviations . . . . .	298
8.2 List of Contributions . . . . .	304
8.2.1 DNA Modification Readers and Writers and Their Interplay . . . . .	304
8.2.2 Repli-FISH: Application of 3D-(Immuno)-FISH for the Study of DNA Replication Timing of Genetic Repeat Elements . . . . .	304
8.2.3 DNA Replication and Repair Kinetics of Alu, LINE-1 and Satellite III Genomic Repetitive Elements . . . . .	304
8.2.4 Developmental Differences in Genome Replication Program and Origin Activation . . . . .	304
8.2.5 Cytosine Base Modifications Regulate DNA Duplex Stability and Metabolism . . . . .	305
8.3 Acknowledgment . . . . .	306
8.4 Declaration - Ehrenwörtliche Erklärung . . . . .	308
8.5 Curriculum Vitæ . . . . .	310

## List of Main Figures

<b>Introduction</b>	<b>6</b>
3.1 From DNA bases to a hierarchical organization of chromatin. . . . .	7
3.2 Evolution of genomes. . . . .	9
3.3 How to initiate genome duplication: overview of origin licensing and activation. . . . .	16
3.4 How to duplicate the genome: overview of the individual steps in DNA replication. . . . .	19
3.5 Replication patterns in mouse somatic cells: visualization of the spatio-temporal organization and progression of DNA replication. . . . .	21
3.6 Replicon and chromatin loop organization. . . . .	25
3.7 Effects of epigenetic cytosine modifications. . . . .	31
3.8 Schematic representation of selected 5mC reader and modifier proteins. . . . .	33
3.9 Role of MBD proteins in modifying and protecting cytosine modifications. . . . .	38
<b>Aims</b>	<b>46</b>
4.1 Microscopic image assembly summarizing the method established in chapter I. . . . .	46
4.2 Scheme summarizing the aims of chapters II and III. . . . .	47
4.3 Scheme summarizing the aims of chapters III and IV. . . . .	48
<b>Chapters</b>	<b>50</b>
5.1 Overview of the localization and sequence of mouse and human repetitive elements. . . . .	54
5.2 Representative timeline and setup of FISH experiments. . . . .	56
5.3 Assessment of FISH probe quality. . . . .	60
5.4 Validation of FISH probe specificity by hybridization to metaphase chromosome preparation. . . . .	68
5.5 Overview showing the localization of Alu and LINE-1 elements in human interphase cells. . . . .	69
5.6 MaSat probe localization in mouse myoblast cells at different S-phase substages. . . . .	71
5.7 Overview showing the localization of Alu and LINE-1 elements in human interphase cells. . . . .	72
5.8 High-content/high-throughput Repli-FISH analysis of asynchronous cell populations. . . . .	73
5.9 Genomic features of repetitive DNA elements. . . . .	88
5.10 Replication timing of repetitive DNA elements analyzed by FISH and S-phase substages classification. . . . .	90
5.11 Genome-wide DNA repair kinetics of non-B and repetitive DNA elements. . . . .	92
5.12 DNA repair kinetics of repetitive DNA elements assessed by FISH. . . . .	94
5.13 Distribution replication and repair kinetics of human repetitive elements. . . . .	96
5.14 DNA replication dynamics in mouse embryonic stem cells. . . . .	127
5.15 DNA replication dynamics in differentiated mouse embryonic stem cells. . . . .	130
5.16 HDAC1 deacetylase targeting to chromocenters in mouse embryonic stem cells. . . . .	132
5.17 Replication timing of (sub)-chromosomal structures in mouse embryonic stem cells. . . . .	134
5.18 Cell cycle characteristics of mouse embryonic stem cells. . . . .	136
5.19 3D quantification and analysis of replication foci throughout S-phase in mouse embryonic stem cells. . . . .	137
5.20 DNA replication fiber and genome size analysis in mouse embryonic stem cells. . . . .	139

5.21 Graphical summary of the cellular and molecular DNA replication characteristics in mouse embryonic stem cells. . . . .	141
5.22 Effect of cytosine modifications on DNA double helix stability. . . . .	202
5.23 Manipulation of modified cytosine levels in cells. . . . .	203
5.24 Effect of cytosine modifications on DNA double helix stability <i>in vivo</i> . . . . .	205
5.25 Effect of cytosine modifications on RNA polymerase activity. . . . .	206
5.26 Effect of cytosine modifications on DNA polymerase activities. . . . .	208
5.27 Effect of cytosine modifications on DNA helicase activity <i>in vivo</i> . . . . .	210
5.28 Effect of loss of cytosine modifications in mouse embryonic stem cells on S-phase progression and replication fork speed. . . . .	213
5.29 Graphical summary of the effect of cytosine modifications on DNA double helix stability and metabolism. . . . .	215
<b>Conclusions</b>	<b>260</b>
6.1 Scheme summarizing the main conclusions from chapters I-IV. . . . .	262

---

## List of Main Tables

<b>Chapters</b>	<b>50</b>
1 Primers used for FISH probe generation. . . . .	58
2 Reaction setup for Alu-specific probe PCR. . . . .	59
3 PCR cycling conditions for Alu-specific probe amplification. . . . .	59
4 Reaction mixture for major and minor satellite-specific probe PCR amplification. . . . .	61
5 PCR cycling conditions for major and minor satellite-specific probe amplification. . . . .	61
6 Reaction mixture for telomere repeat-specific probe PCR amplification. . . . .	62
7 PCR cycling conditions for telomere-specific probe amplification. . . . .	62
8 Overview of human repetitive and non-B DNA elements. . . . .	86
9 ENCODE datasets used for histone modifications and replication timing. . . . .	97
10 PCR conditions for the generation of FISH probes. . . . .	98
11 PCR cycling conditions for the generation of FISH probes. . . . .	98
12 mES cell replicon characteristics. . . . .	140



---

# 1 Summary

While the basic composition of the DNA double helix is known for decades, there is increasing evidence for the stable presence of modified DNA bases in the genome. It is also well known that the majority of our genetic material consists of non-coding and repeat DNA which have important functions in genome regulation, maintenance and organization. Importantly, to avert genome instability through the process of (epi)genome duplication, the underlying processes are tightly regulated in space and time. Whereas many studies correlate replication timing and chromatin landscapes of non-repeat DNA, repetitive regions are difficult to map and, thus, are commonly excluded in genome-wide assays. To overcome this drawback, we developed a protocol to simultaneously visualize DNA replication (or DNA repair) and specific DNA repeat elements in individual and structurally intact 3-dimensional cell nuclei (Repli-FISH). Making use of this single cell microscopy approach in human somatic cells, we find that the euchromatic Alu element (short interspersed nuclear element (SINE)) is replicated early during the synthesis (S-) phase, while the pericentromeric satellite III is exclusively replicated in late S-phase. Long interspersed nuclear element 1 (LINE-1) replication, though, is performed throughout S-phase. Together with available population genomics analyses, we demonstrate that the majority of LINE-1 elements is duplicated according to their distinct histone modifications.

In contrast, pericentromeric heterochromatin (chromocenters) in mouse embryonic stem (mES) cells is replicated during early/mid S-phase (stage II). Similarly, minor satellite and p-arm telomeres are also duplicated during this timeframe, as a consequence of a domino-like replication propagation. We find that this shifted replication timing depends on an increased level of histone acetylation and a more open chromocenter structure. Likewise, we unveil a synchronous replication of the Y chromosome at the end of S-phase, independently of the pluripotency state of the cell. We additionally investigated the properties of the mES cell replicon, i.e. the DNA replicated from one origin, and the number of active replication sites. By combining genome-wide origin mapping, single-molecule analysis and super-resolution microscopy (3D-SIM), we demonstrate that each replication focus visualized by 3D-SIM corresponds to an individual replicon. Moreover, up to one quarter of the foci represent unidirectional forks. Furthermore, mES cell replicons are smaller than in somatic cells, resulting from the activation of twice as many origins spaced at half the distance, highlighting fundamental developmental differences in genome replication organization in pluripotent cells.

In the light of the dramatic effects of histone acetylation level changes on DNA replication, we investigated the influence of cytosine modifications on nuclear processes. We, therefore, established cellular systems to modulate the extent of cytosine variants. Those were based on changing the expression levels of modification writer enzymes, as well as on transfection of modified nucleotides into cells. We find highest DNA duplex stabilities for methylated DNA (5mC) *in vitro* and *in vivo*. This effect is reverted by all oxidized cytosine variants. Accordingly, we find enhanced transcription, replication and DNA unwinding rates in the presence of the destabilizing 5-hydroxymethylcytosine (5hmC) and the absence of the stabilizing 5mC. We show that mES cells deficient of 5mC pass more rapidly through S-phase stage II and, consistently, have faster replication forks. These observations are not the result of altered chromatin condensation, structure, accessibility or histone marks and we conclude that the absence of 5mC enhances DNA unwinding and consequently DNA replication. Thereby, modified cytosines may constitute a mechanism for local fine-tuning of DNA processes.

In summary, our data contribute to a better understanding of different levels of epigenetic regulation involved in nuclear metabolic processes.



# 1 Zusammenfassung

Während die grundlegende Zusammensetzung der DNA-Doppelhelix seit Jahrzehnten bekannt ist, mehrten sich die Hinweise auf das stabile Vorhandensein von modifizierten DNA Basen im Genom. Ebenso ist bekannt, dass der Großteil unseres genetischen Materials aus nicht-kodierender und sich wiederholender DNA besteht, die wichtige Funktionen in der Regulierung, Erhaltung und Organisation des Genoms haben. Zur Vermeidung von Genominstabilität durch den Prozess der (Epi-)Genomverdopplung werden die zugrunde liegenden Prozesse räumlich und zeitlich eng reguliert. Während viele Studien den Replikationszeitpunkt und die Chromatinlandschaften von nicht-repetitiver DNA analysieren, sind 'repeat' Regionen schwer zu sequenzieren und zuzuordnen, und werden daher in genomweiten Studien häufig ausgeschlossen. Um diesen Nachteil zu überwinden, haben wir ein Protokoll zur gleichzeitigen Visualisierung von spezifischen DNA Repeat-Elementen und DNA Replikation (oder Reparatur) in einzelnen und strukturell intakten 3-dimensionalen Zellkernen (Repli-FISH), entwickelt. Unter Verwendung dieses einzelzellmikroskopischen Ansatzes in menschlichen somatischen Zellen zeigen wir, dass das euchromatische Alu-Element (short interspersed nuclear element (SINE)) in der frühen Synthese (S-) Phase repliziert wird, während der pericentromerische Satellit III ausschließlich in der späten S-Phase repliziert wird. Die Replikation des 'long interspersed nuclear element' 1 (LINE-1) wird jedoch während der gesamten S-Phase durchgeführt. Der Zusammenschluss von Genomdaten und unserer Einzelzellmikroskopie weist darauf hin, dass die meisten LINE-1 Elemente entsprechend ihrer jeweiligen Histonmodifikationen dupliziert werden.

Im Gegensatz dazu wird das perizentromere Heterochromatin in embryonalen Stammzellen (mES) der Maus während der frühen/mittleren S-Phase (Stadium II) repliziert. In ähnlicher Weise werden in diesem Zeitraum auch Mikrosatelliten DNA und Telomere auf dem chromosomalen p-Arm dupliziert, was wir auf eine domino-ähnliche Aktivierung von Replikationsursprüngen zurückführen. Wir zeigen, dass dieser versetzte Replikationszeitpunkt von einem erhöhten Histonacetylierungsniveau und einer offeneren Chromocenterstruktur abhängt. Ebenso beobachten wir eine synchrone Replikation des Y Chromosoms, die unabhängig vom Pluripotenzzustand der Zelle das Ende der S-Phase markiert. Wir untersuchten zusätzlich die Eigenschaften des mES-Zellreplikons, d.h. der DNA, die von einem Ursprung ausgehend repliziert wird, und die Anzahl der aktiven Replikationsorte. Durch eine Kombination aus genomweiter Kartierung von Replikationsursprüngen, Einzelmolekül- und hochauflösender Mikroskopie (3D-SIM) zeigen wir, dass jeder Replikationsfokus, der durch 3D-SIM visualisiert wird, einem einzelnen Replikon entspricht. Darüber hinaus stellen bis zu einem Viertel der Foci unidirektionale Replikationsgabeln dar. Des Weiteren sind die Replikons von Stammzellen kleiner als die in somatischen Zellen, was auf die Aktivierung von doppelt so vielen Replikationsursprüngen die in halber Entfernung voneinander angeordnet sind, zurückzuführen ist.

Angesichts der dramatischen Auswirkungen von Veränderungen der Histonacetylierunglevel auf die DNA-Replikation untersuchten wir zusätzlich den Einfluss von Cytosin-Modifikationen auf metabolische Kernprozesse. Wir etablierten daher zelluläre Systeme zur Modulation der Menge an modifizierten Cytosinen. Diese basieren auf der Veränderung der Expression von Enzymen, die die Modifikationen anlegen, sowie auf der Transfektion von modifizierten Nukleotiden in Zellen. Wir messen die höchsten Duplex-Stabilitäten für methylierte DNA (5mC), *in vitro* und *in vivo*. Dieser Effekt wird von allen oxidierten Cytosin-Varianten aufgehoben. Dementsprechend finden wir höhere Transkriptions-, Replikations- und DNA-Abwicklungsraten in Gegenwart des destabilisierenden 5-Hydroxymethylcytosins (5hmC) und in Abwesenheit des stabilisierenden 5mC. Wir

---

zeigen, dass mES-Zellen, denen 5mC fehlt, das S-Phasen-Stadium II schneller durchlaufen und folgerichtig schnellere Replikationsgabeln aufweisen. Diese Beobachtungen sind nicht das Ergebnis einer veränderten Chromatinkondensation, -struktur, -zugänglichkeit oder -histonlandschaft. Alles in allem kommen wir zu dem Schluss, dass das Fehlen von 5mC die DNA Entwindung und folglich die DNA Replikation und Transkription beschleunigt. Daher könnten modifizierte Cytosine einen Mechanismus für die lokale Feinabstimmung von DNA-Prozessen darstellen.

Zusammenfassend tragen unsere Daten zu einem besseren Verständnis der verschiedenen Ebenen der epigenetischen Regulation bei, die an Kern-assoziierten Stoffwechselprozessen beteiligt sind.

## 2 Preface

This dissertation is a compendium of one review article, one method paper and three research manuscripts published in peer-reviewed journals:

**Introduction:** Rausch, C., Hastert, F. D. and Cardoso, M. C. (2019). *DNA Modification Readers, Writers and Their Interplay*. J. Mol. Biol. 432: 1731-1746.

**Chapter I:** Weber, P.\*, Rausch, C.\*, Scholl, A. and Cardoso, M. C. (2019). *Repli-FISH (Fluorescence in Situ Hybridization): Application of 3D-(Immuno)-FISH for the Study of DNA Replication Timing of Genetic Repeat Elements*. OBM Genetics 3: 31.

**Chapter II:** Natale, F.\*, Scholl, A.\*, Rapp, A.\*, Yu, W., Rausch, C. and Cardoso, M. C. (2018). *DNA Replication and Repair Kinetics of Alu, LINE-1 and Satellite III Genomic Repetitive Elements*. Epigenetics & Chromatin 11: 61.

**Chapter III:** Rausch, C.\*, Weber, P.\*, Prorok, P., Hörl, D., Maiser, A., Lehmkuhl, A., Chagin, V. O., Casas-Delucchi, C. S., Leonhardt, H. and Cardoso, M. C. (2020). *Developmental Differences in Genome Replication Program and Origin Activation*. NAR 48(22): 12751-12777.

**Chapter IV:** Rausch, C., Zhang, P., Casas-Delucchi, C. S.\*, Daiß, J. L.\*, Engel, C., Coster, G., Hastert, F. D., Weber, P. and Cardoso, M. C. (2021). *Cytosine Base Modifications Regulate DNA Duplex Stability and Metabolism*. NAR, gkab509.

\* equally contributing authors

To relate all chapters, this thesis starts with a general introduction with key aspects of the mammalian genome and DNA replication, followed by a published review article on *DNA Modification Readers, Writers and Their Interplay*. The results are presented as four individual chapters (chapter I-IV) composed of a method paper (chapter I) and three original research papers (chapters II-IV). At the end, conclusions of all results and further perspectives are discussed. Individual author contributions for each publication are detailed in the Appendix (8.2 List of Contributions) of this dissertation.



## 3 General Introduction

---

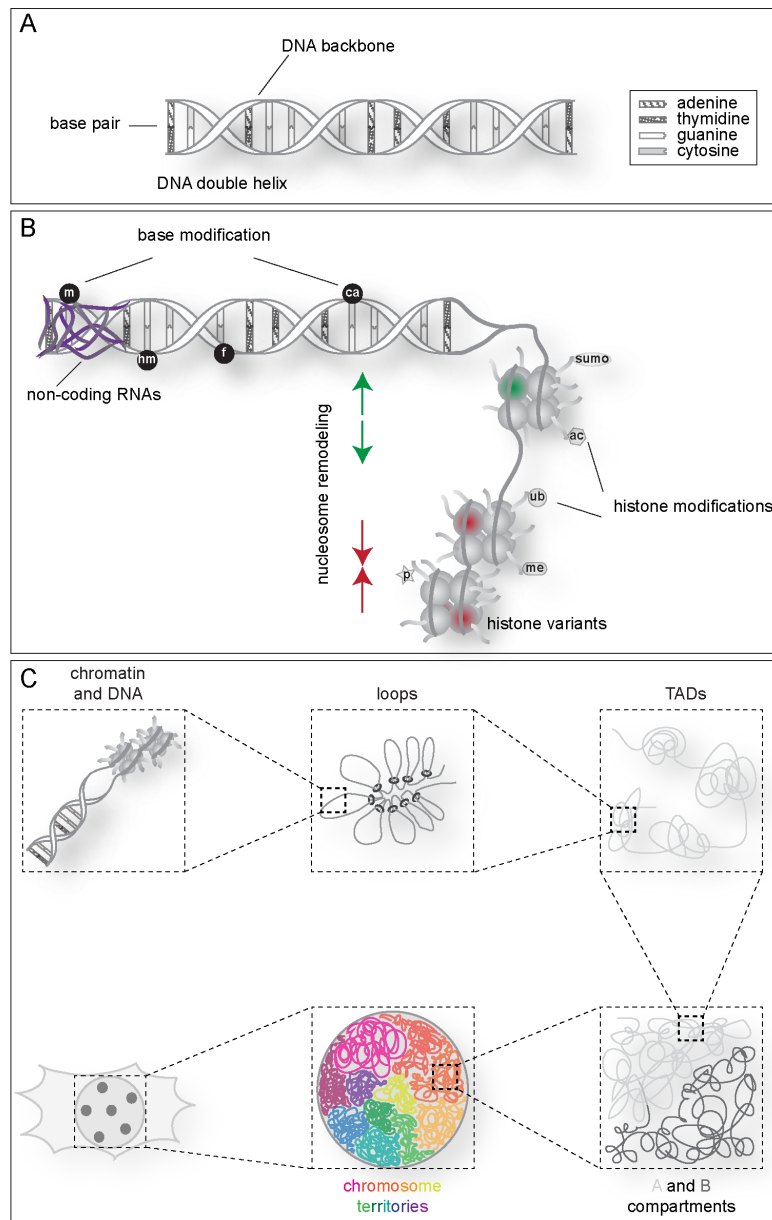
### 3.1 DNA as an Information Storage – The Mammalian Genome

---

Since the elucidation of the DNA double helix structure in the 1950s by Watson and Crick [1], which was considerably promoted by evidence from Rosalind Franklin's DNA X-ray diffraction analyses, the structure, composition and organization of DNA within the cell nucleus were further resolved, resulting in a hierarchical folding model, from the nucleosome complex [2] up to 3-dimensional intra-nuclear chromatin folding and mitotic chromosomes [3, 4].

Already in their 1953 breakthrough publication [1], Watson and Crick highlighted two fundamental characteristics of the DNA molecule - the base sequence complementarity of the two strands and the double helix conformation (Figure 3.1A). Additionally, their model further confirmed the central discovery of the correlation between amounts of adenine and thymine, and cytosine and guanine by Erwin Chargaff [5]. Importantly, their proposed structure allows for the DNA molecule to be replicated into two identical copies (see section 3.2 *DNA Replication – Where, When and How*). While the mere DNA base concatenation represents the linear code defining RNA and protein sequences, the DNA molecule also harbors a more continuous structural code based on DNA configuration, conformation, folding, flexibility and dynamics (see section 3.3 *DNA Modification Readers, Writers and their Interplay*).

To allow packaging of the entire mammalian genome into the limited space of the nucleus, the genetic material is not present as naked DNA in the nucleus but is thoroughly associated with a wide variety of DNA binding proteins and RNA molecules. This DNA-RNA-protein complex is referred to as chromatin and adds a second layer of heritable information to the DNA molecule. The most common DNA-associated proteins found in the chromatin context are histones. Together with defined segments of DNA wrapped around histone octamers, they form nucleosomes. The latter can be dynamically remodeled by either the incorporation of histone variants or specific histone tail post-translational modification signatures. However, the DNA bases themselves can also carry chemical modifications, thereby enlarging the genetic code (Figure 3.1B). It has been long known that besides the four canonical DNA bases adenine (A), thymine (T), cytosine (C) and guanine (G), chemically modified bases are found within the genome of almost all phyla ([6] and discussed in more detail in section 3.3 *DNA Modification Readers, Writers and their Interplay*). Consequently, DNA can store information at different levels and time scales. The DNA sequence itself permits long-term storage, which is stable for many generations ('genetics'). Intermediate-term storage is achieved by specific chromatin compositions (DNA bases, histone variants, histone modifications) and structures ('epigenetics'). This system can be stably propagated to daughter cells, but simultaneously allows erasure or changes in short periods of time. Lastly, short-termed information storage includes highly dynamic and specific DNA-RNA-protein interactions, allowing fast adaptations to changing environmental conditions. All in all, the assembly of DNA into distinct chromatin conformations allows a continuous, dynamic and tight regulation of DNA metabolic processes.

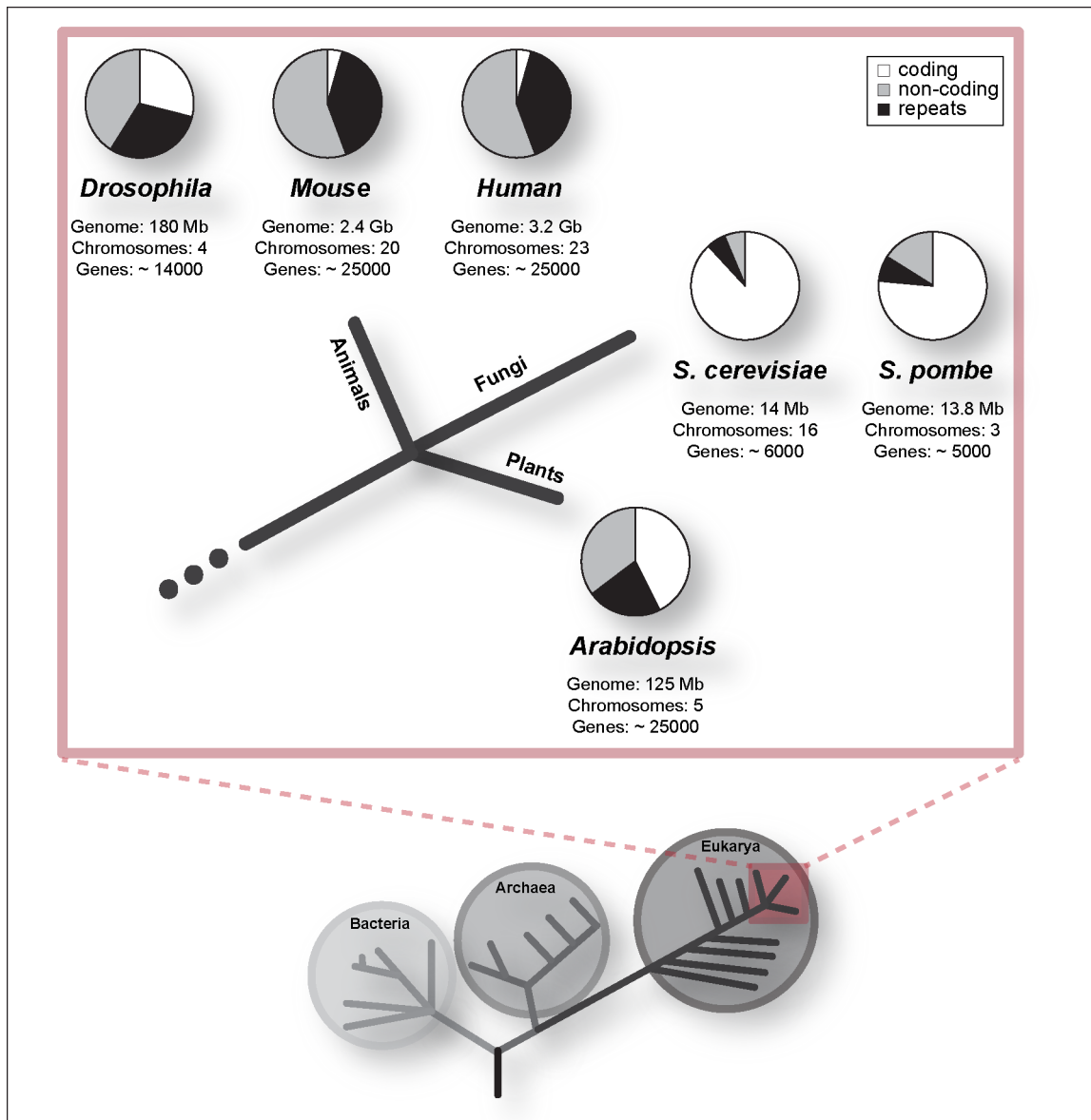


**Figure 3.1: From DNA bases to a hierarchical organization of chromatin. (A)** The DNA polymer is composed of the four canonical DNA bases adenine (A), thymine (T), guanine (G) and cytosine (C) forming A-T and C-G base pairs. These base pairs are organized in a double helix structure. **(B)** To allow packaging of the complete genome into the defined space of the nucleus, the DNA fiber needs to be compacted and is, therefore, tightly associated with DNA binding proteins, resulting in a nucleobase-protein complex, the chromatin. In addition to this linear DNA code, several features are known to add further information and regulatory components to DNA and chromatin. These features include chemical modifications on the DNA bases, non-coding RNAs and specific chromatin marks, such as specific histone variants and histone tail post-translational modifications. All together, the synergy of these components allows nucleosome remodeling, resulting in a tight regulation of DNA accessibility and gene expression. **(C)** At the basic folding degree, the DNA double helix associates with histone octamers, forming nucleosomes. At the next level, the 10-nm chromatin fiber is organized into loop structures (70 - 90 kb) formed by architectural proteins, such as CTCF and cohesin. The loop structure is further compacted to topologically associated domains (TADs, 100 kb - 5 Mb) marked by high DNA contact frequencies within the domain and sharp boundaries. The next level is composed of A/B compartments (500 kb - 7 Mb), which correlate with active and inactive chromatin, respectively. Finally, chromatin forms entire chromosomes occupying distinct spaces in the interphase nucleus termed chromosome territories. m: methyl, hm: hydroxymethyl, f: formyl, ca: carboxyl, sumo: sumoylation, ac: acetylation, ub: ubiquitination, me: methylation, p: phosphorylation.

Compaction and compartmentalization of eukaryotic DNA within the nucleus is characterized by different hierarchical levels (Figure 3.1C). The first level of compaction of the naked DNA double helix is the wrapping around histone octamers, resulting in the formation of nucleosomes (Figure 3.1B). Chromatin is further compacted via the establishment of loop structures, which arrange the DNA into functional units. *In vivo* loop formation is mediated by several architectural (DNA binding) proteins, including the ring-shaped multi-protein complex and molecular motor cohesin and the insulator protein CTCF [7–9]. Cohesin-mediated loop extrusion can be blocked by chromatin-bound CTCF and, consequently, interphase loops are not consecutive but separated by parts of non-looped chromatin. This organization compartmentalizes the genome into insulated domains, enabling distant regulatory elements to come into contact. Interestingly, the ChAHP complex (CHD4, ADNP, HP1) was found to compete with a subset of CTCF binding sites, thereby preventing CTCF looping and locally modulating spatial chromatin organization [10]. Further organization of the genome results in the formation of regulatory landscapes, known as topologically associated domains (TADs), in which three-dimensional chromatin interactions are more frequent than with chromatin outside the given TAD [11]. Additionally, TADs are postulated to represent stable units involved in replication timing definition [12]. Higher-order arrangements of the genome result in the formation of chromatin compartments classified as A compartment, mainly containing open and transcriptionally active chromatin, and B compartments with compact and silenced chromatin [13]. Lastly, the individual eukaryotic chromosomes occupy distinct and restricted locations in the nucleus, the so-called chromosome territories, which establish a basic component of the nuclear architecture.

While the genetic material of all known organisms is composed of nucleic acids undergoing some sort of compaction and organization, genome sizes vary remarkably among the different domains of life and even within the domain of eukarya (Figure 3.2). This variation, however, is not related to organism complexity [14] but rather to the content of non-protein coding DNA in the genomes [15]. Similarly, the complexity of an organism and the number of protein-coding genes cannot be linked. Alike, correlations between genome sizes and numbers of genes are difficult. *Arabidopsis*, for example, has one of the smallest genomes in the plant world (~125 Mb), however, it encloses as many genes as the human genome, i.e. ~25000. At the genomic level, mobile transposable elements (discussed below) are potent mutagenic agents that can induce large chromosomal rearrangements. Additionally, high mobility and the concomitant increase in their copy number can induce genome expansion.

In contrast to bacteria, archaea and some eukaryotes, only a minority of the genomes of mice and humans consists of protein-coding sequences (1.4% and 1.2%, respectively), distributed on ~25000 genes. The vast majority of their genomes, however, consists of non-coding and repetitive sequences (Figure 3.2). Yet, also genomic regions classified as non-coding were shown to undergo pervasive transcription [16]. Additionally, the preservation of the large non-coding fractions of the human genome during evolution, as well as the high phylogenetic conservation of many of the non-coding DNA sequences between evolutionarily distant vertebrates [17], emphasize functional importance within genome organization and likely also regulation. This is in sharp contrast to the view of non-coding sequences as accumulation of residuals of dispensable genes or selfishly replicating parasitic DNA [18, 19]. Although repeat sequences represent around half of the mammalian genomes, the function of most of them, except for centromeres and telomeres, is not yet fully understood. While the unique genomic coding sequences are undeniably of utmost importance, it becomes increasingly clearer that the repetitive segments of the DNA are at the basis of regulatory networking and



**Figure 3.2: Evolution of genomes.** Genome size and gene numbers vary tremendously within the different kingdoms of life. However, no direct correlation between the two numbers can be derived. Genome size increases are mainly ascribed to the evolutionary accumulation of non-protein coding and repeated DNA.

prolific genome formatting and, thus, essential for its function. Already the first drafts of the operon [20] and replicon [21] models postulated protein-genome interactions depending on specific DNA recognition motifs not found in coding sequences. Consequently, repeat sequences are no longer seen as 'junk' or parasitic DNA, but as an indispensable genomic component with functional significance. Repetitive DNA is by now considered an important determinant of chromatin organization, nuclear architecture and epigenetic signatures. Repeat elements are, for instance, involved in shaping distinct eu- and heterochromatin compartments [22–26], determining nucleosome positioning [27], defining insulator and boundary elements [28] and DNA methylation patterns (reviewed in [29, 30]). Transposable elements were recently even postulated to play a potential role in early mammalian development (reviewed in [31]). Moreover, repetitive arrays are important factors for DNA replication initiation, localization and movement. In bacteria and the yeast *Saccharomyces cerevisiae*, origins of DNA replication (discussed in *Where To Start DNA Replication? – Origin Definition*) are repeat arrays of defined DNA sequences [32, 33]. Furthermore, forming telomeres and centromeres is crucial

for genome maintenance and correct transmission of the (epi)genetic information to new cell generations. Given the vital importance of these nuclear structures for cell persistence, a robust system relying on redundant repetitive components allows accurate performances in the absence of some individual units. However, the stoichiometry between repeat DNA recognition motifs and their respective binding proteins is not to be neglected. Similar to protein-protein interactions, protein-DNA interactions also underlie specific ratios and perturbations thereof influence chromatin organization and may result in phenotypic changes [34, 35]. As a consequence, the plethora of repeat elements' abundance is flexible, but changes in array sizes and their spatial distribution within the nucleus are not necessarily phenotypically neutral. Already in 1988, Kornberg and Rykowski could observe the non-random distribution of repeat DNA along human chromosomes and the concomitant chromosomal compartmentalization [36]. Additionally, chromosomes were shown to be organized in a radial manner in the nucleus, with gene-rich chromatin in the center and gene-poor chromatin localized towards the nuclear periphery [37]. In line, nuclear matrix preparations were found to be enriched for repeat DNA [38], suggesting that scaffold/matrix-associated regions (S/MAR) are predominantly composed of repetitive DNA [39]. Furthermore, transposable repeat elements were found to be involved in the establishment of chromatin domain boundaries, thereby contributing to the three-dimensional genome conformation and organization [40, 41]. Repeat elements, hence, influence and define nuclear localization of the DNA fiber and, concurrently, chromatin organization and architecture. Moreover, (repetitive) transposable elements can affect DNA transcription, gene expression regulation, mRNA processing and even translation. Indeed, transposable element exaptation events are commonly observed at new integration loci, resulting in novel regulatory networks orchestrated by the promoter region of a transposable element, thereby also possibly promoting cell adaptability. Likewise, enhancer and silencer abundance can probably be traced back to these integration events. As genome parasites, mobile repeat DNA sequences rely on the host cell for gene expression and, thus, have evolved *cis*-regulatory elements resembling endogenous host promoters. Integration and recombination events result in the frequent deletion of the element's coding sequence whereas the promoter region often persists in the host cell genome (reviewed in [42]). Mobile elements, therefore, represent a rich source for regulatory systems to the host cell, and indeed, chromatin immunoprecipitation (ChIP) assays revealed a multitude of transcription factors binding within the sequences of transposable elements [43]. Additionally, transposable elements are at the basis of a multitude of non-coding regulatory RNA transcripts (microRNAs (miRNAs) and long non-coding RNAs (lncRNAs)), involved in *cis* and *trans* gene expression regulation [44, 45]. Likewise, eukaryotic polyadenylation (poly(A)) transcription termination signals are often derived from transposable elements (reviewed in [46]). Interestingly, some of these elements were even found to stimulate the translational expression of mRNA [47]. All in all, repetitive DNA elements accomplish numerous and essential roles in genome functions and regulation, ranging from basic DNA transcription and its regulation, over DNA restructuring and, concomitantly, chromatin and nuclear organization, to genome maintenance and transmission to following generations. In this context, the view of 'junk' or parasite DNA becomes more and more outdated and (transposable) repetitive elements are by now seen as integral components of the genome and its complex and multi-layered regulatory networks.

Several types of repetitive elements can be distinguished in the mammalian genome. On the one hand, moderately repeated and mobile structures (transposable elements (TE)), such as long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs), long terminal repeat (LTR) elements, endogenous retroviruses (ERV), or DNA transposons are found, on the other hand, highly repetitive elements like

satellite or telomeric repeats exist.

LINEs, SINEs and LTR/ERV elements are retrotransposed via an RNA intermediate. While LINEs are non-LTR containing autonomous mobile genetic elements, SINE retrotransposition is non-autonomous and relies on proteins from autonomous elements. While autonomous transposable elements possess genes for a host independent DNA propagation, they rely on the host cell machinery to express said genes. In eukaryotes, the only active member of the LINE family is LINE-1. In humans, the majority of LINE-1 elements are, however, retrotransposition deficient, due to mutations, 5' truncations, or internal rearrangements [48]. LINE-1 elements have a total length of 6 kb and comprise a 5' untranslated region (5' UTR) with promoter activity in sense (DNA dependent RNA polymerase II) and antisense direction, three open reading frames (ORFs) and a 3' untranslated region (3' UTR) ending in a polyadenylation signal. While the first ORF (ORF1) encodes for a RNA binding protein with chaperone activity (ORF1p) [49], ORF2 codes for a protein with endonuclease and reverse transcriptase activities (ORF2p) [50]. The retrotransposition of LINE-1 elements depends on the cytosolic binding of ORF1p and ORF2p to the LINE-1 RNA transcript and the formation of a ribonucleoprotein particle (RNP). This RNP facilitates re-import of the LINE-1 RNA into the nucleus. Genomic LINE-1 insertion relies on the ORF2p endonuclease and reverse transcriptase activity and is referred to as 'target primed reverse transcription' (TPRT) [51].

One of the most common primate SINE retroelements are the human *Alu* and the rodent B/ID elements, which belong to the 7SL RNA derived SINE1 family [52, 53]. Contrary to LINEs, *Alu* elements are composed of a 280 bp consensus sequence and do not possess a 3' transcription terminator sequence. Instead, they utilize genomic poly-T stretches at various distances downstream of the *Alu* element to terminate transcription. This mechanism implicates high 3' sequence variations between individual elements. The *Alu* sequences do not encode for any proteins, nonetheless, they contain an internal RNA polymerase III promoter that initiates their transcription. Since SINEs are non-autonomous, they are believed to use LINE-1 derived proteins to perform retrotransposition [54, 55].

Opposed to LINEs and SINEs, the LTR retrotransposon structure and replication mechanism resemble exogenous viruses and are characterized by the flanking LTR sequences which are involved in transcriptional regulation. Accumulation of various mutations, however, abolished the production of infectious particles from the proviral DNA (reviewed in [56]). Retrotransposition and integration are catalyzed by an integrase enzyme, similarly to what is known for retroviruses [57]. Contrary, DNA transposons are commonly mobilized by a 'cut-and-paste' mechanism. Transposons are flanked by terminal inverted repeats that are recognized and cleaved by the encoded transposase. The transposon-transposase complex translocates to a recognition site and integration is mediated by the transposase. Interestingly, the genomic distribution of transposable elements is mostly not random, and they exhibit integration preferences for specific genome and chromatin features. Importantly, a balance between the ability for further propagation and non-deleterious effects for the cell is indispensable to ensure 'host' persistence [58].

Considering that these mobile genetic elements have the capability to seriously compromise the genome organization and, thereby, function and integrity, multi-layered mechanisms of protection have evolved in mammals. These include retrotransposition silencing via DNA methylation and histone modifications [59] and (reverse) transcript editing by cytidine deaminase proteins (APOBEC) [60]. Additionally, a complex PIWI (P-element induced wimpy testis) protein and piRNA dependent silencing mechanism was described [61]. The majority of transposable elements is, thus, assumed to be transcriptionally silenced. Since transposable

elements constantly evolve to escape repression, the genomic safety mechanisms are also highly adaptive and redundant. Nonetheless, a variety of examples of transposable elements' promoters driving the expression of host genes are documented ([62] and discussed above), suggesting that transposable (regulatory) elements are not as strongly and consistently silenced as assumed. Indeed, genome-wide mapping assays have found a myriad of active and predominantly lineage-specific transposable elements derived transcription factor binding sites.

In human and mouse genomes, highly repetitive satellite DNA is mainly, but not exclusively, found in heterochromatic regions. Prominent members of the satellite family are human satellite III and mouse major satellite DNA that are found in pericentromeric regions, while human  $\alpha$ -satellite and mouse minor satellites compose the centromeres [48, 63–68]. Mouse centromeric and pericentromeric DNA consists of highly conserved tandem repeats with multiple copies of the repeat unit being organized in a head-to-tail manner. Mouse minor satellite (MiSat) DNA is composed of 120 bp AT-rich repeats spanning approximately 600 kb of the centromeres. In contrast, mouse major satellite (MaSat) DNA consists of 234 bp long repeats that span up to 6 Mb and that are situated adjacent to the MiSat sequences [65]. In mouse cells, (peri)centromeric repeats of the telocentric/acrocentric chromosomes form large clusters of DNA that are visible as so-called 'chromocenters' in DNA counterstained cells. Both satellite repeat categories share several structural and epigenetic signatures, like high levels of cytosine methylation and low histone acetylation levels, however, they also present distinct individual features. While MiSat repeats are characterized by the centromere-specific histone variant CENP-A [69], MaSat regions are bound by heterochromatin protein 1 isoforms (HP1) [70]. Taken together, satellite DNA is proposed to play important roles in centromere function, chromosome pairing and heterochromatin formation, organization and maintenance.

Telomeres are the eukaryotic linear chromosome ends which form large nucleoprotein complexes to protect them from DNA damage responses and chromosome end processing via repair pathways. Telomeres are composed of non-coding, highly conserved, mostly double-stranded and repetitive TTAGGG hexanucleotide tandem repeats that span 10-15 kb of chromosome ends in humans, and that are bound by the shelterin complex [71, 72]. Within this complex, the nucleic acid forms G-quadruplex and T-loop structures [73, 74]. Telomeric DNA contains a single-stranded, 3' G-rich strand that serves as a template for telomerase enzymes. Telomerase is only active in some tissues, including embryonic stem cell populations and many cancer cell lines [75, 76]. Together with the shelterin complex, the telomerase regulates telomere length [72].

In conclusion, DNA and the adopted chromatin structure and conformation represent different levels of a complex data storage system. Higher eukaryotic genomes comprise surprisingly large amounts of repetitive DNA with considerable structural diversity and astonishing functional importance, including nuclear metabolic reactions, chromatin organization and nuclear architecture.



---

## 3.2 DNA Replication – Where, When and How

---

**O***mnis cellula e cellula* - with his postulate from 1855, Rudolf Virchow introduced the fundamental biological principle that every cell derives from a pre-existing cell. This essential formulation of the cell theory implies the need for cell divisions to allow cell persistence. Simple cell division, however, would lead to the continuous dilution of the genetic material. Therefore, complete and error-free duplication of all (epi)genetic information prior to the equal distribution to the daughter cells during cell division, is essential. Incomplete or erroneous genome duplication, or over-replication can cause karyotype changes and mutations, leading to genomic instability and, ultimately, to a variety of diseases and, eventually, to cell death. DNA replication is, thus, a fundamental nuclear metabolic process to ensure cellular and organismic development and needs intensive and vigorous regulation of each individual step.

Hence, DNA replication is highly conserved across all kingdoms of life. Together with their model of the DNA double helix, Watson and Crick postulated already in 1953 a semiconservative DNA replication mechanism [1, 77]. The problem of the duplication of an antiparallel DNA double helix was solved by similar mechanisms across phyla. The strand that is synthesized in the same direction as the replication machinery is moving (leading strand), is replicated continuously, while the second strand (lagging strand) is replicated in a discontinuous manner. The following sections will present an overview of the molecular mechanism of mammalian DNA replication and its regulation in space, time and development.

### 3.2.1 Where To Start DNA Replication? – Origin Definition

**S**ince the earliest studies on DNA replication, it was of great interest to define where DNA replication starts within the genome and how the exact duplication only once per cell cycle is achieved. In 1963, the replicon model by Jacob *et al.* proposed a mechanism for the control of DNA replication initiation in bacteria [21]. The authors suggested that replication initiation relies on two major factors, a replicator and an initiator. While the former is defined as (a) DNA sequence(s) that is/are necessary and sufficient to allow DNA replication initiation, the latter represents the protein recognizing, binding and distorting/unwinding the replicator sequence and recruiting further proteins involved in DNA replication. At the most basic definition, the replicator is or comprises the origin of replication (ORI). The ORI is the DNA site where template DNA unwinding allows the initiation of DNA synthesis (Figure 3.3A). The model additionally introduced that all DNA replicated from one origin forms a replicon [21].

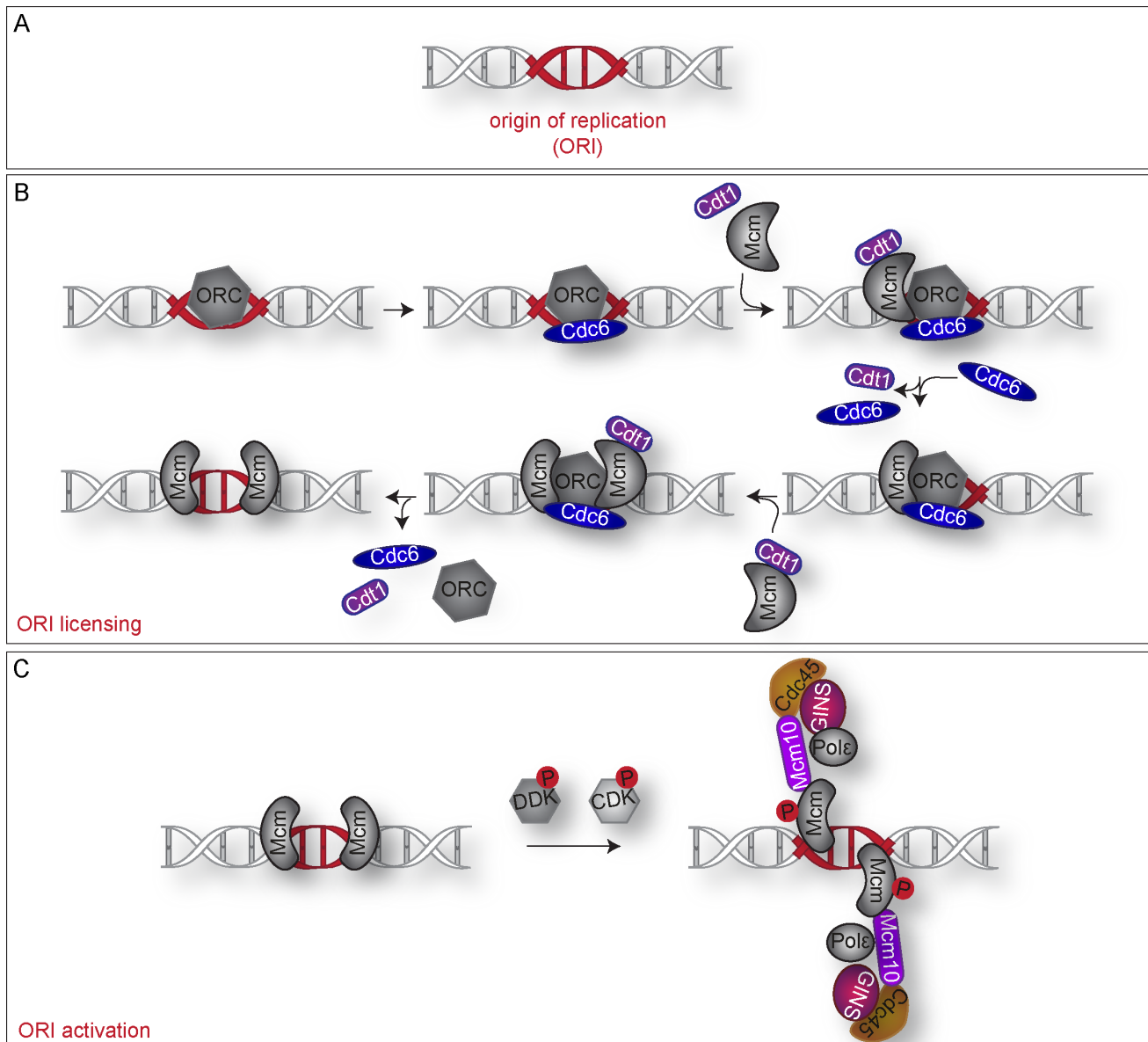
In a very simplistic view, DNA replication can be subdivided into two major steps. First, origins of replication throughout the genome are recognized and licensed by specific protein complexes and second, the activation of a subset of these licensed origins results in the duplication of all genetic and accompanying epigenetic information by the multi-protein DNA synthesis complex (replisome, reviewed in [78–80]).

Although several factors and mechanisms involved in DNA replication initiation are conserved among

bacteria, archaea and eukaryotes, other aspects are highly divergent. One prominent characteristic of mammalian origins of replication is the lack of a clear sequence and signature defining the initiation sites. In most bacteria and lower eukaryotes, replication origins are determined at the DNA sequence level. Replication of the ~4.6 Mb circular chromosome of *Escherichia coli*, for example, starts at a unique origin of replication of 245 bp [32], termed *oriC*. The latter comprises multiple binding sites for the DNA replication initiation factor DnaA, which binds a 9-mer consensus sequence. Upon assembly of the initiation complex at the *oriC* and a local bend in the DNA, *oriC* is unwound. Finally, this initial DNA unwinding allows DnaB helicase binding and formation of a pair of replisomes moving bidirectionally to replicate the entire chromosome [81–85]. Studies in the yeast *Saccharomyces cerevisiae* revealed similar DNA elements responsible for genome duplication. Autonomously replicating sequences (ARS) comprise one or several repeats of an 11-mer consensus sequence with a high AT content and confer the ability for *in vitro* self-propagation to plasmids [33, 86]. While budding yeast is the only eukaryotic organism with defined origins of replication, recent studies have shown that the whole chromatin composition, rather than the mere DNA sequence, promotes replication origin dependence [87, 88]. Origin of replication sites in mammalian cells are also considered to have a biased distribution throughout the genome, at least in differentiated somatic cells. Characterization of ectopic replication from mammalian replicators revealed several ORI signatures, including G-quadruplexes, AT richness, dinucleotide repeats, nucleotide distribution/asymmetry or matrix attachment regions. However, the few studied mammalian replication origins all possess different features [89–96]. The first human origin of replication described in detail was the Lamin B2 origin, located upstream of the Lamin B2 gene [97]. Subsequent work revealed the exact start site of DNA replication in this locus down to the nucleotide level [98]. Alike, a defined replicator with sequence similarity to the yeast ARS was found for the c-myc origin [99–101]. The dihydrofolate reductase (DHFR) gene in Chinese hamster cells, however, is characterized by a 55 kb initiation zone where DNA replication can start [102–104]. In view of the more complex genome architecture of higher eukaryotes, e.g. several large chromosomes with long non-coding segments and highly condensed (peri)centromeric DNA regions, these metazoans may use different classes of origins of replication to ensure proper genome duplication. A general and uniform consensus sequence at the basis for DNA replication initiation might not be compatible with the highly sophisticated genome structure and plasticity of multicellular organisms. However, this sequence independence comes along with a certain randomness of origin placement throughout the genome. Together with the requirement of a conglomerate of replication origins to achieve replication of the large genomes, this implies the risk of unreplicated areas if two neighboring origins of replication are too distant to allow replication of the inter-origin DNA stretch within the restricted time-frame of S-phase. To counteract the deleterious effects of incomplete genome replication emanating from large inter-origin distances, eukaryotes have evolved a sophisticated system of origin redundancy as well as a complex and tight spatio-temporal regulation of genome replication discussed in detail below.

### **3.2.2 How To Start DNA Replication? – Origin Licensing and Activation**

While it is still rather unclear where DNA replication starts within mammalian genomes, the molecular mechanisms of origin licensing and activation are far better understood. DNA replication initiation is a highly spatio-temporally regulated multi-step process. Interestingly, the decisions where DNA replication potentially starts in the next cell cycle already take place far ahead of the actual synthesis phase (S-phase),



**Figure 3.3: How to initiate genome duplication: overview of origin licensing and activation. (A)** Contrary to bacteria and lower eukaryotic organisms, mammalian origins of replication (ORI) are not defined by specific DNA sequences or epigenetic signatures. **(B)** Origin licensing in mammalian cells includes a multitude of different factors and a tight temporal regulation. After ORC binding to the origin of replication, recruitment of CDC6 results in binding of the first CDT1/MCM2-7 complex. Upon CDC6 and CDT1 release, a new CDC6 protein binds to the ORC/MCM assembly, initiating the recruitment and binding of a second CDT1/MCM2-7 complex. ORC, CDT1 and CDC6 become dispensable for subsequent origin activation. The two head-to-head organized MCM complexes represent the pre-replication complex, and together with the encircled chromatin section, define a licensed origin of replication. **(C)** Origin activation is dependent on phosphorylation events of the CDK and DDK kinases. Phosphorylated MCMs recruit essential replication factors, such as CDC45, MCM10 and GINS (CMG complex). Conformational changes in the MCM rings impose a switch from dsDNA binding to ssDNA binding. The MCM helicase complex starts dsDNA unwinding at the origin of replication, thereby enabling DNA synthesis.

namely at the mitotic exit. As a consequence, metazoan pre-replication complex formation (pre-RC) occurs in the absence of an intact nuclear membrane and under low CDK (S-phase specific cyclin-dependent kinase) conditions. At the molecular level, pre-RC formation consists of the recognition of origins of replication

throughout the genome which are subsequently bound by the origin recognition complex (ORC) at the end of S-phase/beginning of G1 (Figure 3.3B). Characterization of the *S. cerevisiae* ORC revealed a hexameric protein complex (ORC1-6) and homologs were identified in several species, including mammals [105–107]. After ORC binding, the AAA<sup>+</sup>-ATPase CDC6 (cell division control protein 6) is recruited by the complex during mitosis to G1 transition. ATP binding by several ORC subunits is essential for ORC/CDC6/DNA complex formation, ultimately culminating in replication initiation (Figure 3.3B, [108–111]). Additionally, CDC6 was proposed to modulate ORC chromatin binding and to promote origin DNA sequence specificity [111, 112]. Interestingly, recruitment of ORC or CDC6 to DNA is sufficient to create an artificial origin of replication, recruiting pre-replication complex proteins and initiating plasmid DNA replication in mammalian cells [113].

Next, the ORC/CDC6 recruits a complex comprising CDC10-dependent transcript 1 protein (CDT1) and the mammalian minichromosome maintenance 2-7 (MCM2-7) core DNA helicase heterohexamer in an open-ring structure [114–118]. Subsequently, CDC6 and CDT1 are sequentially released from the ORC/CDC6/CDT1/MCM2-7 complex via ATP hydrolysis, resulting in an intermediate ORC/MCM2-7 complex [119–122]. CDT1 release from the protein network promotes conformational changes in the MCM hexamer, inducing the closure of the helicase ring around the DNA double helix [118]. Binding of a new CDC6 to ORC and ORC/CDC6/MCM2-7 complex formation rapidly recruits a second MCM2-7 hexamer in a CDT1 dependent manner, resulting in the formation of a symmetrical head-to-head double MCM hexamer encircling the DNA [119, 123–125]. Formation of the double hexamer induces the release of CDC6 followed by simultaneous CDT1 and ORC disassembly from the complex and the closure of the second MCM2-7 ring around DNA [118, 119]. This stable head-to-head MCM2-7 heteromeric complex with DNA allows DNA replication with bidirectional forks emanating from one origin. Additionally, it represents the pre-replication complex and marks origin licensing in eukaryotic cells (Figure 3.3B). Interestingly, it was shown that the loading of two MCM rings necessitates the binding of two ORCs to two recognition sites [126]. A more recent study, however, revealed that loading of an MCM double hexamer is dependent on the binding of ORC molecules to the opposite termini of a single initial helicase ring, thereby initiating the loading of a second single MCM hexamer in the appropriate orientation for the formation of the MCM head-to-head conformation [127]. Generally, MCM loading to replication start sites occurs in excess and is commonly referred to as the 'MCM paradox' (reviewed in [128]). This 'over licensing' serves as a safety mechanism to counteract replicative stress. In case of replication fork stalling, neighboring dormant origins can be activated to complete DNA replication, thereby avoiding genome instability in rapidly changing environmental conditions [129–131]. These 'backup' replication origins need to be licensed before the beginning of S-phase, since pre-replication complex assembly during S-phase bears the risk of re-replication, perturbing genome integrity.

Upon S-phase entry, a subset of the licensed origins of replication throughout the genome becomes activated. Origin activation is a highly manifold process involving the recruitment of a multitude of additional factors to the inactive MCM complex and depends on DBF4-dependent CDC7 kinase (DDK) and S-phase specific cyclin-dependent kinase (CDK) mediated phosphorylation events [123, 132]. In metazoans, the CDK phosphorylation-dependent origin firing is mediated by a complex composed of Treslin/MTBP/TopBP1 (MDM2 binding protein and Topoisomerase II binding protein 1, respectively) [133]. In line, MTBP was recently proposed to be an essential platform for origin firing control [134]. During DDK and CDK kinase activity driven replisome assembly, the heterotetrameric GINS complex (SLD5, PSF1, PSF2 and PSF3), CDC45, MCM10 and polymerases  $\alpha$  and  $\epsilon$  are loaded to the MCM2-7 helicase on the origin DNA (Figure 3.3C) [135–137]. CDC45,

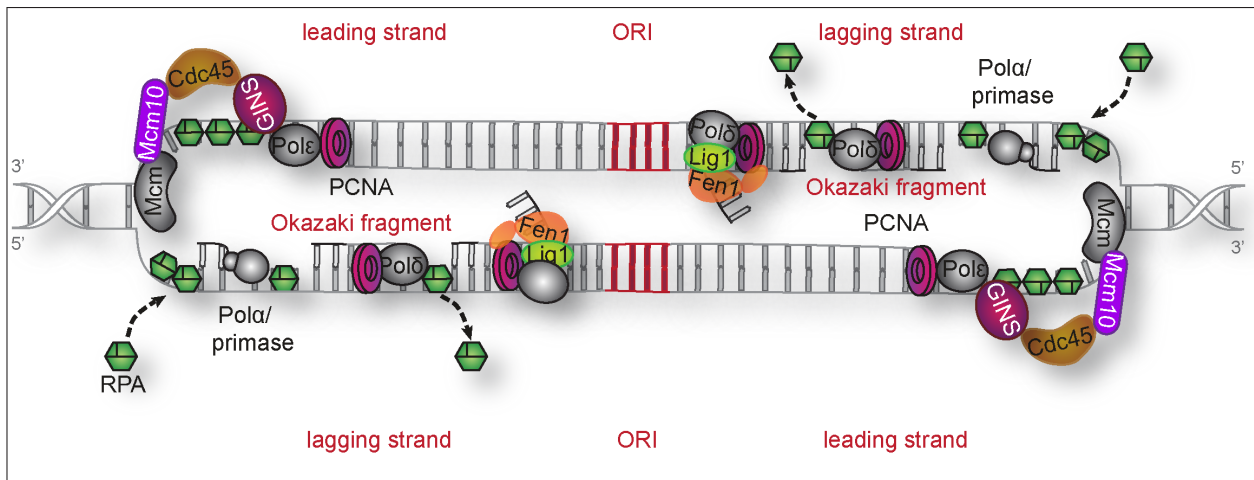
MCM2-7 and GINS form the CMG complex [136–140]. MCM10 binding induces structural changes of the complex, eventually leading to origin firing [136, 137, 141–143]. During helicase activation, crucial MCM10 dependent reorganization of the MCM2-7 double hexamers results in the splitting into two independent MCM complexes [144]. Subsequently, one DNA strand becomes extruded from each hexamer, resulting in two CMG complexes surrounding single-stranded DNA (ssDNA) [145–147], allowing bidirectional DNA replication (Figure 3.3C). The active CMG complex is highly effective in ATP-hydrolysis mediated 3' to 5' DNA unwinding [138, 148]. The above-described mechanisms for origin firing are well known and characterized in budding yeast. Interestingly, the minimal set of purified *S. cerevisiae* replisome components and co-factors allowing *in vitro* DNA replication initiation was recently resolved [137].

Similar to the broad variety of origin types (delimited as for Lamin B2 or broad initiation zone as for DHFR), also the firing efficiency of origins can differ widely. Three different classes of DNA replication origins can be distinguished: flexible, dormant/inactive, or constitutive origins. The latter, the minority of eukaryotic origins, are constantly used in any cell cycle and cell type. Dormant/inactive origins are used as backup origins in case of replication stress, as discussed above. According to the Jesuit model, flexible origins occur in large clusters and can be used stochastically in every cell. This concept also explains the low origin use/firing efficiency observed in eukaryotes ('For many are called, but few are chosen', Matthew 22:14, The Bible, [149]). A recent study identified a class of origins, termed 'core origins', that are at the basis of more than three-quarters of all initiation events in different cell types. Although no specific consensus sequence could be deduced, a general G-richness was characteristic for most of these initiation sites. Additionally, computational algorithms were able to predict the position of the majority of these core origins in human and mouse genomes [150].

### 3.2.3 How To Replicate the DNA? – The Replisome And Leading versus Lagging Strand

As mentioned above, organisms with an antiparallel double-stranded genome replicate their genetic information preferentially by bidirectional replication forks containing a continuously replicated leading strand and a discontinuously replicated lagging strand. While essential replisome proteins are present on both strands, others differ and have specific functions in leading and lagging strand synthesis (Figure 3.4).

Once the CMG helicase complex unwinds the dsDNA template, the resulting ssDNA strands are covered by single-stranded DNA binding proteins (SSB) to avoid DNA reannealing and damage. In eukaryotes, ssDNA is stabilized by the heterotrimeric replication-associated protein A (RPA), consisting of the three subunits RPA1, RPA2 and RPA3 [135]. The actual eukaryotic DNA synthesis machinery comprises three polymerases that propagate fork movement and DNA synthesis, including polymerase/primase alpha (pol  $\alpha$ ), polymerase delta (pol  $\delta$ ) and polymerase epsilon (pol  $\epsilon$ ). Polymerase/primase  $\alpha$  is composed of four subunits and, contrary to the replicative polymerases  $\delta$  and  $\epsilon$ , it is deficient in 3' to 5' proofreading. Instead, it has additional RNA primase activity [151], which confers the unique feature of starting *de novo* DNA synthesis [152]. The polymerase  $\alpha$  enzyme generates small hybrid RNA/DNA primers. In detail, after several ribonucleotides, the RNA primer switches from the RNA primase site into the DNA polymerase site and the short RNA primer is extended with deoxynucleotides [153, 154]. After the primer synthesis, polymerase  $\alpha$  is exchanged to polymerases  $\delta$  and  $\epsilon$ . These replicative polymerases are composed of several subunits, the biggest being the actual DNA polymerase part. It is still under debate if the two polymerases divide labor at the replication fork. While many studies in yeast show leading strand replication mainly by polymerase  $\epsilon$  and lagging strand DNA synthesis



**Figure 3.4: How to duplicate the genome: overview of the individual steps in DNA replication.** Given the antiparallel double-stranded DNA helix, a replication bubble contains two replication forks moving in opposite directions. On both strands, DNA synthesis is initiated by the polymerase/primase  $\alpha$  generating a short RNA/DNA hybrid primer. Unwound ssDNA is immediately covered by single-stranded binding proteins (RPA). Upon PCNA loading, the primase is replaced by polymerase  $\epsilon$  on the leading strand and by polymerase  $\delta$  on the lagging strand. While the leading strand is replicated continuously, the lagging strand is replicated discontinuously. Once the polymerase  $\delta$  reaches the preceding Okazaki fragment, strand displacement takes place and a short flap is generated. The latter is removed by FEN1 and the remaining nick in the DNA is ligated by ligase 1.

mainly by polymerase  $\delta$  [155–158], an important role may also be assigned to polymerase  $\delta$  DNA synthesis on leading strand synthesis [159, 160]. Nonetheless, and contrary to polymerase  $\alpha$ , both polymerases work in tight conjunction with the proliferating cell nuclear antigen (PCNA) to ensure high processivity (reviewed in [161, 162]). Importantly, PCNA loading to the DNA segment primed by polymerase  $\alpha$  effects the switch to polymerase  $\delta$  or  $\epsilon$ . PCNA is a homotrimeric ring protein loaded onto chromatin by the clamp loader replication factor C (RFC) and is generally seen as the loading platform for a multitude of replication-associated factors [163–166]. Hence, PCNA is also a key protein in DNA repair processes (Figure 3.4) [167]. Similar to the key replication initiation factors, the core budding yeast replisome components enabling *in vitro* leading and lagging strand synthesis with purified proteins at *in vivo* replication rates were defined in a recent study [160].

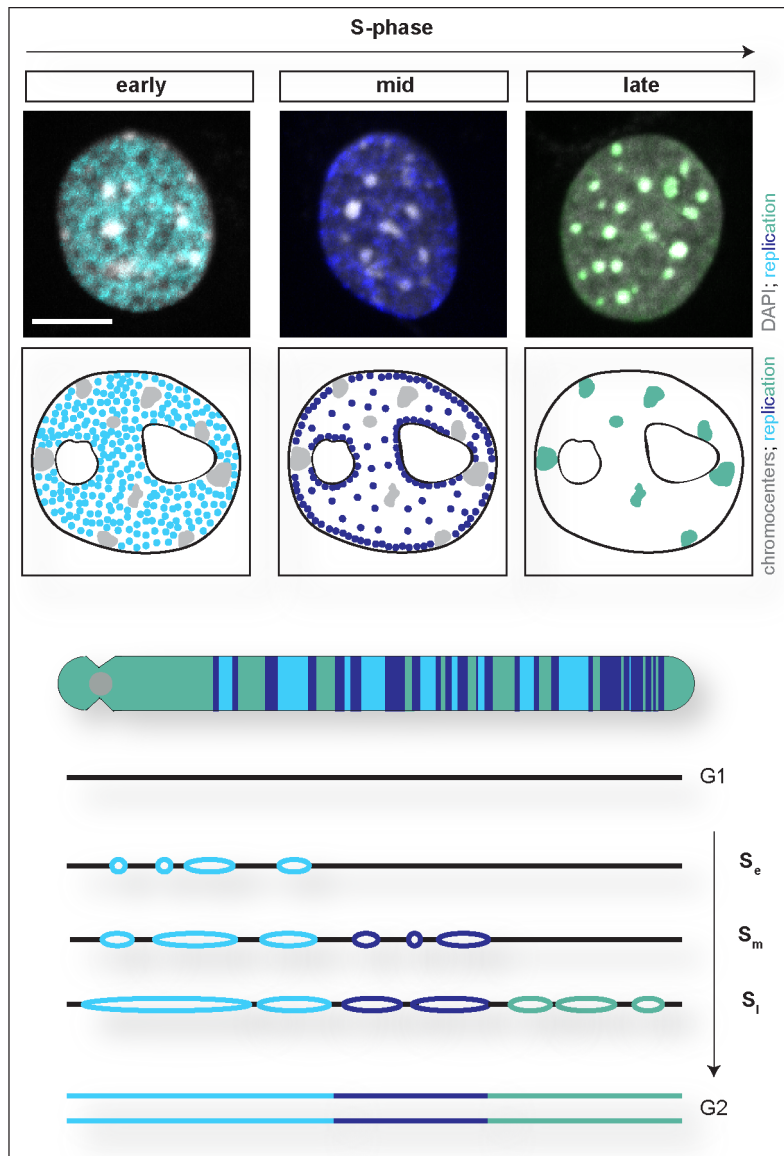
DNA polymerases possess 5' to 3' synthesis activity. While this is perfectly applicable in a continuous manner for replication of the leading strand by polymerase  $\epsilon$ , lagging strand duplication needs discontinuous DNA synthesis. Already in 1968, the Okazakis proposed a model for the latter [168]. Similar to leading strand replication, Okazaki fragments are primed by polymerase  $\alpha$ . After PCNA loading, primer segments are elongated by polymerase  $\delta$ . Once polymerase  $\delta$  encounters the preceding Okazaki fragment, a short strand displacement is initiated and the resulting 5'-flap is cut by the flap endonuclease 1 (FEN1). DNA-DNA nicks are repaired by DNA ligase 1 [169]. Polymerase  $\delta$ , FEN1 and ligase 1 are the core enzymes for Okazaki fragment maturation (Figure 3.4). In view of the differences in leading and lagging strand replication machineries, additional mechanisms to regulate and coordinate efficient replication of both strands are needed. Studies in the T4 bacteriophage led to the proposal of the trombone model, where lagging strand DNA is bent back on itself, in a way that the two polymerases can perform DNA synthesis coordinately [170]. DNA polymerase alpha-binding protein CTF4 was proposed to be the eukaryotic replisome coordinator in such a trombone model [171, 172]. However, yeast CTF4 knock-outs are viable, despite showing increased genome instability,

suggesting the presence of more or other factors coordinating the replisome [173, 174]. Additionally, recent cryo-EM studies of the *S. cerevisiae* CMG helicase bound to DNA revealed that the polymerase/primase  $\alpha$  and the polymerase  $\epsilon$  are located at opposite sides of the GINS complex. These structures suggest that the specific polymerase  $\alpha$  positioning would allow immediate priming on the lagging strand that gets excluded from the helicase core. On the other hand, polymerase  $\epsilon$  arrangement would enable direct DNA synthesis on the leading strand extruding from the CMG central channel [146].

### 3.2.4 Where and When To Replicate DNA? – DNA Replication in Time and Space

**M**any DNA replication factors, features and mechanisms are conserved across numerous organisms, including yeast, flies, frogs and humans (reviewed in [80]). Although the initial origin licensing occurs at hundreds and thousands of genetic loci, only a subset of these marked origins are sequentially activated during S-phase. Early DNA replication studies showed that large chromosomal segments replicate at specific times during S-phase [175]. Together, these findings reveal the existence of a distinct and conserved DNA replication timing program that, in humans, correlates positively with transcriptional activity and gene density [176]. Generally, gene-rich euchromatic genome regions replicate early in S-phase (early S-phase,  $S_e$ ) and AT-rich, gene-poor heterochromatic regions are duplicated late during S-phase (late S-phase,  $S_l$ ). Altogether, this highly regulated spatio-temporal order of origin activation suggests an important biological function of the replication timing program. Indeed, the tight regulation of DNA replication over time ensures that the number of active replication forks does not exceed the available rate-limiting factors, such as replisome components and nucleotides [177]. Although this explains why not all origins fire at the same time, it does not elucidate why the distinct order of origin firing is of importance. On the one hand, replication timing control was proposed to regulate gene dosage regulation [178]. However, most genes underlie mechanisms reducing gene expression two-fold after replication [179, 180]. On the other hand, the replication program may result in a concentration of genomic variation in late replicating and generally gene-poor DNA. These genomic regions were previously shown to display higher frequencies of point mutations, probably due to a downregulation of DNA mismatch repair proteins during mid S-phase [181, 182]. Lastly, during DNA replication not only the mere DNA sequence is replicated, but also all epigenetic features and chromatin structures need to be copied to the daughter chromatin. Therefore, the highly regulated spatio-temporal replication program may be at the basis for a robust maintenance of cell type-specific chromatin features and for allowing efficient change of the latter during cell fate transition.

DNA replication timing profile studies (RT profiles) from large cell populations revealed distinct replication domains of about 1.5 - 2.5 Mb [176, 183, 184]. They exhibit sharp boundaries between neighboring domains with different replication timing and alter along individual chromosomes [184]. While RT-profiles can directly be connected to the underlying DNA sequences, they have poor temporal resolution and do not provide information about the spatial DNA organization within the nucleus. Microscopic analyses, however, permit detailed insights into the spatial (and temporal) organization of DNA replication. Incorporation and detection of modified nucleotides or components of the replisome, allow *in situ* visualization of newly synthesized DNA and sites of ongoing DNA replication in a three-dimensional cell nucleus. These sites are referred to as 'replication foci' (RFi). Observation of S-phase progression revealed the existence of specific replication patterns (Figure 3.5 up, [185–188]), representing the dynamic distribution of replication foci during genome duplica-



**Figure 3.5: Replication pattern in mouse somatic cells: visualization of the spatio-temporal organization and progression of DNA replication.** Schematic representation of the three main replication patterns observed in mouse somatic cells. At the beginning of S-phase (early S-phase) replication signals are distributed throughout the nuclear interior. During mid S-phase, replication foci are mainly seen at the nuclear and nucleolar periphery. The end of S-phase (late S-phase) is characterized by highly condensed pericentromeric heterochromatin replication. Here, the replication signals increase in size and colocalize with DAPI intense chromocenters. This spatio-temporal S-phase progression is also addressed in the domino-like DNA replication model. According to this model, DNA replication is initiated by a few stochastically activated origins in early S-phase (cyan). Conformingly to the next-in-line activation process, the ongoing DNA synthesis triggers the activation of proximal origins (blue). As S-phase progresses, also origins located in late replicating DNA get activated (cyan-green) resulting in the completion of genome duplication.

tion. In somatic cells, DNA replication follows the compaction state of the underlying chromatin and three main replication patterns are generally distinguished. During early S-phase ( $S_e$ ), transcriptionally active and gene-rich euchromatin is replicated, mid S-phase ( $S_m$ ) is characterized by facultative heterochromatin replication, and during late S-phase ( $S_l$ ) inactive and condensed constitutive heterochromatin is duplicated. *In situ*, these three patterns exhibit specific three-dimensional distributions. In mouse somatic cells, early S-phase is characterized by the distribution of replication foci throughout the nuclear interior while omitting

the nucle(ol)ar periphery and condensed pericentromeric heterochromatin (seen as regions with intense DNA counterstaining, (chromocenters)). Next, replication foci are seen at the nuclear and nucleolar border and in female cells, a prominent accumulation of RFI determines replication of the inactive X (Xi). Lastly, replication foci decrease in number while increasing in size and colocalize with DAPI intense constitutive heterochromatin, i.e. chromocenters (Figure 3.5 up).

Interestingly, pulse-chase studies revealed that replication foci distribution remains stable over several cell cycles, hinting towards a conserved chromatin organization regulating important nuclear processes, such as DNA replication. Furthermore, this organization was proposed to rely on clusters of replicons, i.e. the segment of DNA replicated by a pair of bidirectional replication forks emanating from one single origin of replication [185]. Genome-wide replication timing analyses further confirmed the coordinated activation of replication origin clusters over large chromosomal segments, as already proposed by single-molecule studies [184, 185, 189]. Subsequently, two models explaining the organization of replication machineries and their interaction with chromatin emerged. The first model involves replisomes being attached to a diffuse nucleoskeleton during S-phase. Electron microscopic studies of embedded HeLa cells revealed fixed and granulated structures, the replication factories. Achieving DNA duplication by these static replication machineries involves continuous DNA reeling through the factories [190]. However, the existence of the underlying matrix or scaffold is still debated. Additionally, fluorescence recovery after photobleaching (FRAP) experiments unveiled the preferential *de novo* assembly of replisomes adjacent to pre-existing ones [163], suggesting that replisomes do not persist as fixed and permanent machineries throughout S-phase. This contiguous activation is at the basis for the second model, proposing that DNA is replicated in a domino-like manner which determines the temporal order of origin firing. This initiation mode implies a self-propagating origin activation process using the next-in-line principle which induces ORI activation spreading over a complete chromosome until it becomes fully duplicated (Figure 3.5 down, [191–193]). In detail, the initial stochastic firing of a subset of origin clusters [194], preferentially within an open chromatin conformation, leads to a chain reaction, propagating origin firing to later origin clusters in relative spatial proximity. This origin initiation spreading is consistent with the increase in origin activation efficiency as S-phase progresses [195, 196]. Indeed, replication initiation was shown to be inefficient in early S-phase, while relatively late origin firing efficiency was significantly higher. This increased ability to initiate DNA replication correlates with, and probably depends on, the higher probability of any chromosome segment to be in close proximity to actively replicating DNA sites at later time points of S-phase. Consequently, origins located in highly condensed and late replicating constitutive heterochromatin become accessible since more of the adjacent chromatin undergoes local decondensation and gets replicated. Accordingly, the next-in-line activation of replication sites correlates well with their one-dimensional linear organization on chromosomes [197, 198].

Several regulators for DNA replication timing were identified, studying the defined functions of these regulators is, however, challenging, due to robust maintenance systems for replication timing preservation. Besides the well-known correlation of DNA transcription and early DNA replication [176, 184, 199], several proteins like Rif1, G9a and Treslin were shown to regulate the replication timing in cells [200–202]. Considering the replication timing program correlation with chromatin compaction, additional prominent candidates for the regulation of DNA replication are epigenetic marks. In humans, ORC binding sites were found to correlate well with DNaseI hypersensitive DNA segments. In accordance with the notion of DNA replication initiation within chromatin with an open conformation, histone modifications typically found in active chro-

matin, namely H3K27ac and H3K4m2, were found to be predictive ORC binding factors [203]. Albeit, H4K20 methylated histones and the involved writer enzymes were also suggested to play important roles in DNA replication regulation [204–207]. Generally, the reversible addition of acetyl groups to histone tail residues changes chromatin conformation to a less compact and more open structure. Histone acetylation is an important regulator of DNA replication and is currently the best predictor for early replicating chromatin. Genome-wide studies as well as single-cell microscopy analysis demonstrated that genomic loci with increased histone acetylation levels and high accessibility tend to replicate early during S-phase and loss of histone acetylation leads to replication timing switch to late S-phase [208–214]. Yeast strains deficient for the histone deacetylase Rpd3p showed advanced replication timing of large parts of their genome and the increased levels of histone acetylation resulted in an earlier firing of ORIs, especially of normally late initiating ORIs [214]. Similarly, Casas-Delucchi *et al.* showed by a microscopic approach that global histone hyperacetylation leads to the activation of a subset of ORIs within normally late replicating constitutive heterochromatin during early S-phase in somatic mouse cells. Consequently, histone hypoacetylation is required to maintain late DNA replication timing [? ]. Moreover, studies on the  $\beta$ -globin genomic domain related histone acetylation and DNA replication timing control in human cells [215]. Deacetylation inhibition of its genomic locus via trichostatin A (TSA) treatment in HeLa cells, advanced its replication timing [216], while targeting a histone deacetylase to a  $\beta$ -globin transgene in erythroid cells, delayed DNA replication onset of the latter [217]. Likewise, TSA treatment of human cells resulted in earlier replication of imprinted genes [209]. Additionally, the transcriptionally silenced locus of monoallelically expressed genes replicated later than the active counterpart, and the latter exhibited enrichment of H3 and H4 acetylation marks [212]. Generally, all epigenetic marks representing distinct features of eu- or heterochromatin, including DNA and histone modifications as well as higher chromatin structures, are potential DNA replication timing regulators.

### **3.2.5 How To Organize DNA Replication? – DNA Replication and Chromatin (Loop) Organization**

**C**hromosomes are organized in loop structures within the nucleus throughout the cell cycle. Loops are known to compact the DNA polymer, to organize it in time and space and to be important for essential nuclear processes such as DNA transcription [218], mitotic/meiotic chromosome compaction and functional organization during the interphase (reviewed in [219]).

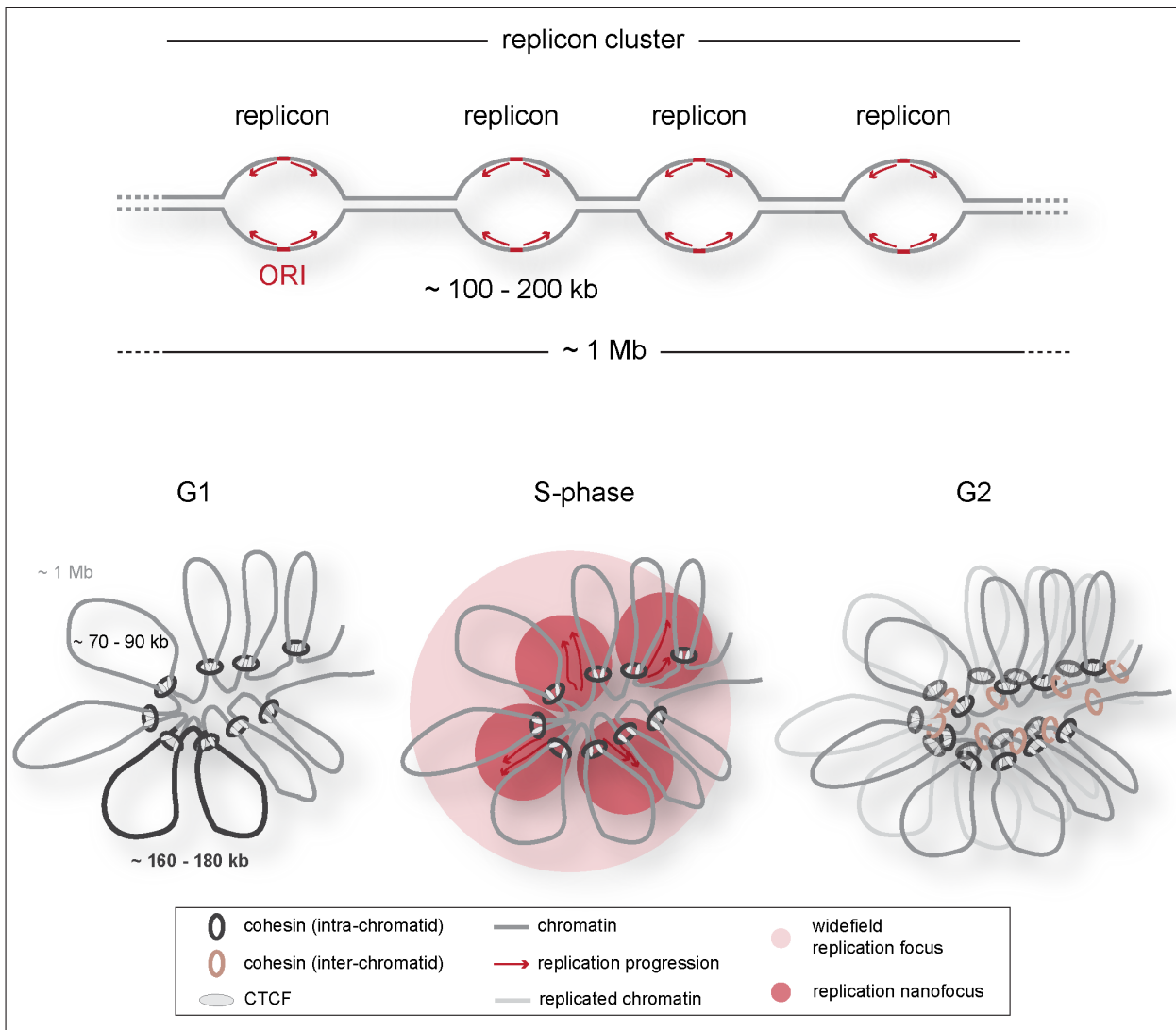
Early autoradiographic fiber studies showed that replicons are organized and activated in clusters. These clusters are on average 1 Mb in size and consist of 2 to 9 individual replicons of 100 - 200 kb [185, 220]. DNA halo experiments revealed a striking coincidence between replicon and chromatin loop sizes [221, 222]. Early nucleotide labeling experiments revealed that these replicons become visible as replication foci in three-dimensional interphase cell nuclei and contain at least 1 Mb of DNA [223]. Consequently, replication foci represent the activated replicon clusters observed in single-molecule DNA fiber analyses (Figure 3.6). Hand in hand with the development of microscopes with higher optical resolutions, the number of identified replication foci also increased dramatically (reviewed in [224]). Recent 3D structured illumination microscopy (3D-SIM) and single-molecule DNA fiber analyses shed further light on the organization of replication foci in human and mouse somatic cells [225]. 3D-SIM unveiled an average of 5000 replication nanofoci at any given time during S-phase. This result further enhanced the fact that replication foci seen with conventional resolution microscopy (average of 1000 RFi) do not represent individual replicons, but clusters of 4-5 of these basic

units of DNA replication (Figure 3.6). Further comparison between the numbers of nanofoci visualized by 3D-SIM and the theoretical number of active replication sites derived from genome size, replication fork speed and total S-phase duration, demonstrated that replication nanofoci represent individual replicons. Specific chromatin clustering circumstances even allow the visualization of individual replication forks. Interestingly, the replication foci size distribution was shown to be very similar, even after different nucleotide labeling times [226]. This observation, together with the persistence of replication foci over several cell cycles, highlights an important connection between DNA replication and the underlying chromatin structure. Altogether, the relation between loop and replicon sizes and the fixed size of replication foci suggest a crucial function of loop structures in delimiting the DNA segment contained in a replicon, ultimately defining them as functional units in the context of DNA replication (reviewed in [227]). Hence, DNA replication relies on a structural DNA unit of about 70-90 kb, representing one chromatin loop, and the bidirectional forks of one replicon label a DNA segment with the length of a pair of loops, i.e. 160-180 kb. The latter values are equivalent to the average inter-origin distances measured in mouse and human somatic cells, respectively [225]. Thus, chromatin (loop) organization is an important regulator of DNA-dependent processes. Besides allowing the tight packaging of chromatin within the nucleus, chromatin loops enable long-range interactions of specific DNA segments distant from each other along the linear DNA fiber, e.g. gene enhancer and promoter regions as seen for the  $\beta$ -globin gene [228]. Cohesin was additionally found to be present at replication origins and to even physically interact with the MCM2-7 complex. Coherently, cohesin knock-down perturbed chromatin organization and, in turn, origin activation, replicon size and S-phase length [229]. Recent studies in cohesin knock-down cells revealed bigger replication nanofoci, i.e. replicons, suggesting a disturbed and more relaxed chromatin conformation [230].

### 3.2.6 How to Duplicate More Than DNA? – DNA Replication and Epigenetics

**H**igher organisms do not only face the challenge of duplicating their genetic sequence, but the additional information arising from DNA modifications, histone composition and modifications and general chromatin higher-order structure also needs to be replicated accurately. In view of this, DNA replication is a double-edged sword, being, on the one hand, productive by synthesizing new DNA strands to enable cell duplication, but, on the other hand, also disruptive, since it destroys the established chromatin composition and structure to enable fork progression. To oppose deleterious effects emanating from this local chromatin disassembly, epigenome replication is tightly coupled to DNA replication. Consequently, the replication machinery contains, in addition to DNA replication factors, proteins involved in chromatin duplication.

In comparison to other epigenetic marks, maintenance of the DNA methylome (5-methylcytosine, 5mC) is quite well studied and understood. The cell cycle-dependent mechanisms of 5mC inheritance are relying on the key factors DNA methyltransferase 1 (DNMT1) and its facilitating factor UHRF1 (ubiquitin-like with plant homeodomain (PHD) and ring finger domains 1). DNMT1 methylates hemi-methylated DNA generated upon DNA replication via a DNA base flipping mechanism [231, 232]. In a first step, hemi-methylated DNA is recognized and bound by the UHRF1 SRA (SET and ring finger associated) domain, resulting in a conformational change that, in turn, allows tandem tudor domain (TTD) mediated H3K9m3 binding in close proximity of the hemi-methylated DNA [233, 234]. Subsequent activation of the UHRF1 E3 ubiquitin ligase leads to mono-ubiquitylation of two lysines in the H3 histone tail [235–237]. These marks are in turn bound by the TS domain



**Figure 3.6: Replicon and chromatin loop organization.** Replicons are organized in clusters of  $\sim 1$  Mb DNA with 2-9 replicons of 100 - 200 kb. To achieve chromatin packaging in the limited space of the cell nucleus, DNA needs to fold into a compact conformation, including wrapping around histone octamers and organization into loop structures mediated by cohesin (intra-chromatid) and CTCF. During S-phase, replicon clusters can be visualized as replication foci with standard optical resolution microscopy. Using 3D-SIM, replication nanofoci can be attributed to individual replicons. DNA replication relies on a structural unit representing a DNA loop of circa 70-90 kb, and two forks of a bidirectional replicon label a length of DNA that corresponds to a pair of loops, i.e. 160-190 kb. In G2, the two sister chromatids are held together by cohesin molecules (inter-chromatid).

of DNMT1 and found to be crucial for methyltransferase activity [235]. DNMT1 nuclear localization is strongly cell cycle-dependent [238]. While in G1 the methylase is homogeneously distributed throughout the nucleoplasm, association with replication foci is observed during S-phase. In early S-phase, DNMT1 is targeted to sites of active replication sites via the UHRF1 dependent dual mono-ubiquitylation of the PCNA-associated factor 15 (PAF15) and localization at pericentromeric heterochromatin during late S-phase is more dependent on the UHRF1 mediated dual mono-ubiquitylation of histone H3 interacting with the DNMT1 targeting sequence (TS) [235, 239-242]. During G2, DNMT1 persists and is reloaded at pericentromeric heterochromatin DNA regions, possibly ensuring complete DNA re-methylation of these highly modified DNA segments in a DNA replication-independent manner [243].

Contrary to the maintenance of 5mC, little is known about how the oxidized cytosine modifications 5-

hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine (5hmC, 5fC and 5caC, respectively) are preserved over DNA replication. Although primordially seen as pure transient DNA demethylation intermediates, there is growing evidence that 5hmC and 5fC can also act as stable epigenetic marks [244, 245]. Similar to 5mC, 5hmC levels were found to be stable during the cell cycle and 5mC oxidation occurs on the newly synthesized strands with a time delay. This is in agreement with the fact that 5hmC can only originate from an existing 5mC base. However, the large time gap between 5mC generation and oxidation thereof to 5hmC suggests a different 5hmC maintenance mechanism in comparison to DNA methylation, and 5hmC generation probably does not immediately take place on the nascent DNA strand [244]. Yet, the exact mechanism by which these oxidized bases are maintained after DNA replication remains elusive.

To allow compaction of DNA into highly condensed structures, DNA is associated with a core histone octamer consisting of two H2A-H2B and two H3-H4 histone dimers. Each histone core particle is wrapped with 147 bp of DNA, further stabilized by the linker histone H1 and individual nucleosomes are spaced by linker DNA. To allow fork progression during DNA replication, nucleosomes are disassembled ahead of the replication fork and chromatin composition and structure need to be reassembled once the DNA backbone is doubled. Nascent chromatin contains parental histones that are recycled from dismantled nucleosomes and newly compiled, naïve nucleosomes. Globally, HeLa cells were shown to contain a 1:1 ratio of new versus recycled histones, but these values may vary cell type and locus-specifically [246]. Importantly, new histones are uniformly acetylated at lysines 5 and 12 of histone 4 (H4K5ac and H4K12ac, discussed in [247]). The distribution of parental modified histones can be achieved in two ways. First, it was supposed that parental H3-H4 tetramers split and mix with new histones to guarantee post-translational modification (PTM) inheritance on the template and new strands [248]. However, more recent work revealed the almost complete absence of parental and new histone mixing in HeLa H3-H4 tetramers, arguing for the reuse of intact and modified tetramers. PTMs would consequently be copied from neighboring modified histones [249]. In contrast, H3-H4 tetramers can assemble in any combination with new or recycled H2A-H2B dimers [250]. A recent study using the reconstituted histone complex incorporation into chromatin of permeabilized cell method (RhIP) revealed a striking correlation between histone incorporation and the chromatin state. While histone deposition in active chromatin can be achieved in a replication-coupled as well as in a replication-independent way, heterochromatic histone deposition mainly relies on DNA replication [251].

### **3.2.7 When and How To Reorganize DNA Replication? – DNA Replication and Development**

**A**s discussed before, single-origin initiation is a very flexible process allowing the adaptation to dynamically changing environmental conditions and the preservation of genome integrity. The non-static nature of the complex process of DNA replication is best visualized during *Xenopus* and *Drosophila* embryonic development. Here, cell divisions occur within the range of minutes, requiring genome duplication in extraordinarily short time frames. In differentiated cells, however, cell cycle and, importantly, also S-phase, take several hours [252–255]. These striking changes in replication timing management come along with massive rearrangements of the origin and replicon organization [256–258]. Early developmental stages of these two organisms exhibit high origin activity, short inter-origin distances and, consequently, small replicons. Of note, these cell cycles are achieved in the complete absence of transcription. The lack of gene expression is compensated by maternally derived protein supplies [253, 259]. Interestingly, the larger inter-

origin distances observed in differentiated cells can be reverted when DNA is conditioned in *Xenopus* egg extracts. Coherently, loop sizes and ORC distribution are remodeled [260, 261]. Differential ORC binding in various *Drosophila* cell types emphasizes the dynamic remodeling of DNA replication during cell lineage commitment [262, 263]. Similar cell type and differentiation state-specific origin usage was also found in mouse and human cells [264, 265]. Additionally, relations between DNA replication and active transcription were postulated, hypothesizing that replication organization is associated with and influenced by the transcription program characterizing cell differentiation. However, analysis of individual genetic loci also revealed that such origin initiation does not necessarily and solely rely on transcription. In the mouse *HoxB* locus, several active origins of replication were mapped in undifferentiated cells. Upon induction of cell differentiation and *HoxB* transcription, all origins within this region get silenced, except for one which turns into the single dominant origin of replication within the whole domain [266]. This observation in differentiated cells is in sharp contrast to the genome-wide correlation of origin initiation in transcriptionally active DNA segments [267] and emphasizes the importance of single-loci studies. Additionally, the early cleavage stages in *Xenopus* and *Drosophila* embryos take place in the complete absence of DNA transcription, still, distinct replication timing programs were identified. These attempts to establish a causality between DNA replication timing and transcription highlight the fact that the two processes correlate probably only indirectly and are regulated by similar, yet, separate mechanisms.

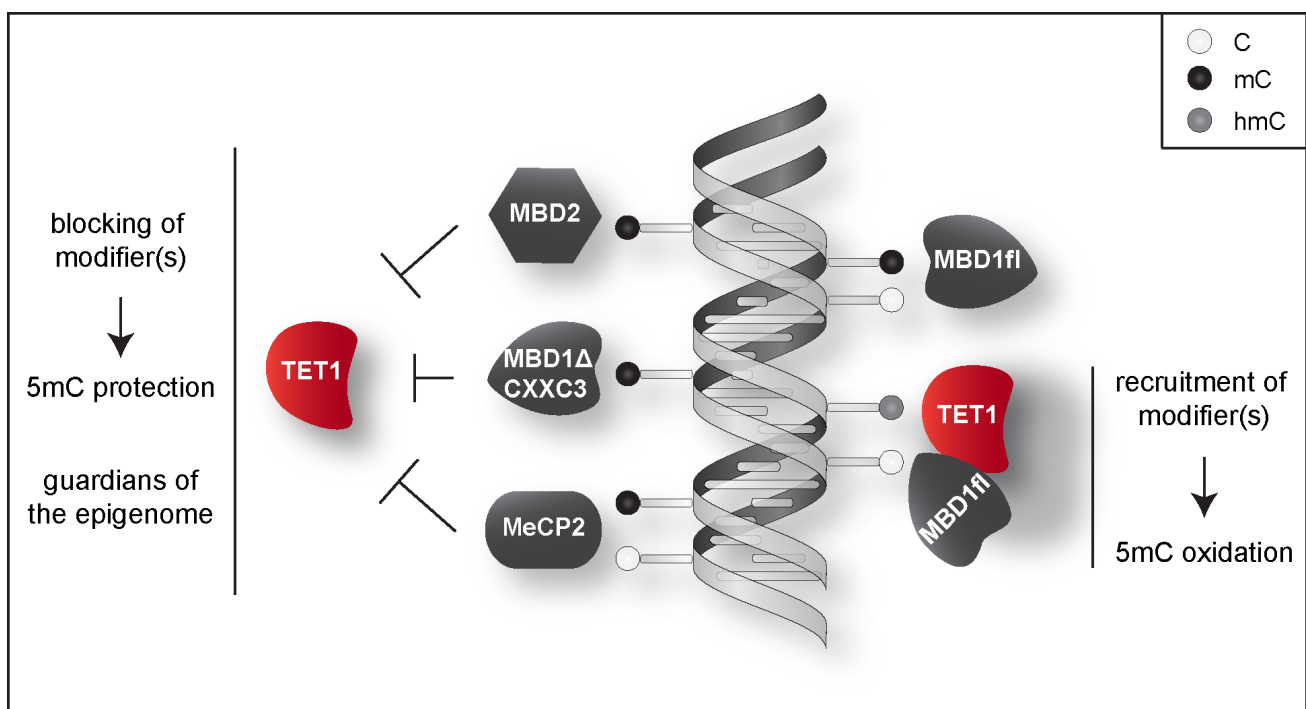
Although most of our knowledge of developmental-associated differences in DNA replication arises from frogs and flies, it was shown that already 1-cell stage mouse embryos have a tight spatio-temporal regulation of DNA replication [268]. Furthermore, during induced differentiation from embryonic stem cells to neural progenitor cells, ~20% of the genome are affected by replication timing reorganization [184]. Thus, multicellular organisms are able to dramatically reorganize their replication program and this flexibility is essential for genome maintenance, proper cell differentiation and organism development. The strong interconnection between DNA replication and DNA transcription may allow easier ORI firing in transcribed DNA regions. Similarly, it may introduce a segmentation of the genome into independent regions 'focussing' mainly on replication or transcription, thereby reducing encounters and collisions of these two essential genome metabolic processes, thereby preserving genome stability. In line, flexible origin firing may allow adaptive transcription crucial for cell differentiation.

The extensive reorganization of DNA replication programs during development comes hand in hand with a broad chromatin restructuring. As described before, chromatin consists of a complex DNA-RNA-protein network which adds a second layer of heritable information to the DNA molecule. While the main protein components of chromatin are histones, many other proteins participate and interact with the DNA-RNA-histone structures to establish and maintain the chromatin conformation. Among these are DNA base modification writer enzymes, such as DNMT (5mC) and Ten-eleven translocation (TET, 5hmC, 5fC and 5caC) enzymes, and DNA modification reader proteins, such as methyl-CpG binding domain (MBD) proteins. Their manifold interactions and tight spatio-temporal regulation are essential for genome composition, integrity, architecture and function. A detailed description of DNMT, TET and MBD proteins, as well as the interplay between DNA base modifications, their readers and writers will be discussed in the following *DNA Modification Readers, Writers and their Interplay* section.

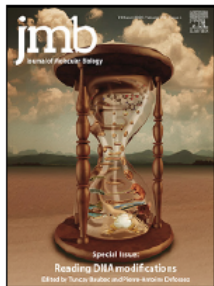
### 3.3 DNA Modification Readers, Writers and Their Interplay

C. Rausch prepared figure 3 and wrote the manuscript (except for sections 3.1 - *DNMT writer enzymes and the 5mC mark* written together with F. D. Hastert and section 4.1.3 - *TET isoforms* written by F. D. Hastert) together with M. C. Cardoso.

#### Graphical Abstract







## DNA Modification Readers and Writers and Their Interplay

Cathia Rausch, Florian D. Hastert and M. Cristina Cardoso

*Cell Biology and Epigenetics*, Department of Biology, Technische Universität Darmstadt, 64287 Darmstadt, Germany

**Correspondence to M. Cristina Cardoso:** Department of Biology, TU Darmstadt, Schnittspahnstr. 10, 64287 Darmstadt, Germany. Fax: +49 6151 16 21882. [cardoso@bio.tu-darmstadt.de](mailto:cardoso@bio.tu-darmstadt.de)

<https://doi.org/10.1016/j.jmb.2019.12.018>

**Edited by Pierre-Antoine Defossez**

### Abstracts

Genomic DNA is modified in a postreplicative manner and several modifications, the enzymes responsible for their deposition as well as proteins that read these modifications, have been described. Here, we focus on the impact of DNA modifications on the DNA helix and review the writers and readers of cytosine modifications and how they interplay to shape genome composition, stability, and function.

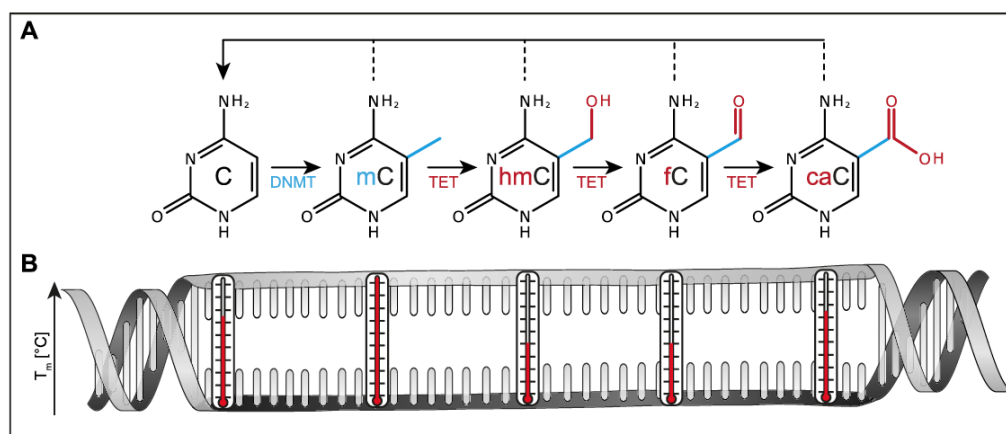
© 2019 Elsevier Ltd. All rights reserved.

### Introduction

Epigenetic modifications of mammalian DNA are crucial regulators of organismic development but also in pathogenesis. The most abundant and most studied mammalian DNA modification was described by Rollin Hotchkiss in 1948, when he identified, in addition to the four canonical DNA bases, a fifth peak on a paper chromatogram from thymus deoxyribonucleic acids. He termed this novel mammalian DNA base “epicytosine” [1], unknowingly anticipating the modern definition of epigenetics and the additional layer of information that epigenetic DNA/chromatin modifications add to the genetic information of every cell. In the same study, Hotchkiss suggested that “epicytosine” corresponded to methylated cytosine. In 1972, a sixth DNA base, named 5-hydroxymethylcytosine (5hmC), was identified in different vertebrate tissues, although this was likely the result of an experimental artifact as in the same study 5-methylcytosine (5mC) was not detected [2]. Hence, 5mC remained the only noncanonical mammalian DNA base that was studied to a larger extent for a long time and helped explaining how different cellular phenotypes can arise from the same nucleotide sequence. The postreplicative covalent addition of a methyl group to the fifth carbon atom of the cytosine pyrimidine ring by the DNA methyltransferase (DNMT) enzymes and oxidation thereof by the Ten-eleven-

translocation (TET) enzymes diversify the genome and shape the chromatin landscape (Fig. 1A). DNA cytosine methylation is predominantly found in the context of symmetric CpG dinucleotides and, in the human genome, 70%–80% of all CpGs harbor methylated cytosines [3,4]. This modification is crucial in numerous cellular processes such as X-chromosome inactivation and imprinting, gene repression, control of cellular development and differentiation, silencing of repetitive elements, and genome stability maintenance [4–7]. Meanwhile, two more cytosine variants, 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), were identified as oxidation products of 5hmC mediated by TET enzymes [8] (Fig. 1A), shedding light into a possible active DNA demethylation pathway [9]. However, 5hmC and 5fC were also identified as stable epigenetic marks [10,11] and all cytosine modifications are recognized by specific reader proteins [12]. The (hydroxy)methylome is specifically read, amongst others, by members of the methyl-CpG binding domain (MBD) protein family (reviewed in Ref. [13]).

Here, we review the effects of DNA base modifications on DNA double helix structure and the interplay between cytosine methylation readers and modifiers. Although methylation of the nitrogen-6 of deoxyadenosine (m6dA) has been recently described in DNA from mammals [14,15] and linked to nucleosomal positioning in *Tetrahymena* [16], it



**Fig. 1. Effects of epigenetic cytosine modifications.** (A) DNA methyltransferases (DNMT) catalyze the addition of a methyl group to the fifth carbon atom of the cytosine pyrimidine ring creating 5-methylcytosine (5mC). The methyl group is oxidized by Ten-eleven-translocation (TET) proteins creating 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) in an iterative manner. (B) Schematic representation of the effect that different epigenetic cytosine modifications exert on the melting temperature ( $T_m$ ; graphically depicted as a thermometer) of the DNA double helix.

will be omitted from this review as its existence is still under debate [17].

### Effect of cytosine modifications on DNA double helix structure and recognition by proteins

The flexibility of the dynamic DNA double helix system is a fundamental condition to allow tight protein-DNA contacts, is sequence dependent, and mainly determined by the stacking energy of base steps, hydrogen bondings between base pairs, and exocyclic functional groups pointing within the major or minor groove [18]. Since DNA replication, transcription, and repair all rely on unwinding of the DNA double helix, the influence of base modifications on helix parameters might directly affect DNA metabolic processes.

A variety of *in vitro* reports showed that, besides the mere DNA base composition that influences the melting temperature ( $T_m$ ) of double-stranded DNA (dsDNA) [19], also the addition of chemical moieties, such as methyl or hydroxymethyl groups to DNA bases can influence the mechano-physical properties of the DNA double helix. C5 cytosine methylation was early on described to increase the melting temperature of an oligonucleotide carrying the modification in comparison to an unmethylated counterpart in a highly sequence-dependent manner [20,21]. In the following years, these findings were confirmed using UV spectroscopy, thermal melting analysis, and mechanical strand separation experiments and were also extended to the oxidative methylcytosine derivatives. Generally, cytosine methylation was reported to increase DNA melting

temperature, indicative of a stabilization effect on the DNA double helix, whereas oxidation of the methyl group decreases the  $T_m$ , reversing the methylation-induced duplex stabilizing impact (Fig. 1B) [22–33]. 5hmC was in some studies found to not only revert the stabilizing influence of cytosine methylation, but to even further destabilize the double helix, resulting in a  $T_m$  lower than that of nonmodified cytosine ( $T_m$ : 5mC > C > 5hmC) [24,25,32].

The influence of cytosine modifications on DNA structure is still highly debated. The general view, however, is that the modified bases do not change the global B-DNA conformation, although local structural changes around the modification site were measured in a sequence-dependent manner [27,32,34–37]. Molecular dynamics and all-atom Monte-Carlo simulations showed that cytosine methylation changes the geometry of base pair steps resulting in an increase of the local curvature of the methylated DNA [34,37–40]. In addition, methylated DNA stretches can show slid structures that lead to local DNA distortions [22]; they are stiffer, more torsionally rigid, and less flexible, since the steric hindrance of the bulky methyl group within the major groove counteracts DNA bending [37,41–43]. Moreover, cytosine methylation might lead to a slight widening of the major groove and in turn a narrower minor groove [44]. On the contrary, 5hmC and 5fC containing DNA exhibit changes in base pair geometry that result in an increased flexibility [32,42], as well as alterations in the minor groove geometry [45]. Wanunu et al. showed computationally and experimentally that the increasing polarity of the cytosine modification leads to a lowering of the rigidity of the DNA stretch. In turn, the change from hydrophobicity (5mC) to hydrophilicity (oxidized

variants) of the DNA base modification can lead to a less stable DNA duplex [32], thereby linking DNA structural changes to DNA flexibility and rigidity, the solvent shell state, and eventually double helix stability.

The addition of a functional group to DNA bases can also influence DNA structure by changing base stacking and hydrogen bonding between base pairs. Cytosine methylation enhances base stacking interactions [22,28,30,34,46], due to a higher molecular polarizability upon methyl substitution [47]. Interestingly, base stacking abilities for 5hmC, 5fC, and 5caC were found to be similar [48]. Hydrogen bondings between base pairs containing 5mC or 5hmC were found not to be influenced by cytosine modifications [39,46,49]; however, there are reports that 5fC and 5caC substantially affect H-bonding interactions [34,50] and form intranucleobase hydrogen bonds between the oxygens of the ketone groups and the exocyclic N4 amino group of the cytosine [45,48].

These base modification-dependent changes in the local structure of DNA were proposed to represent a way to their selective recognition by proteins. High-resolution DNase I cleavage profiles and bisulfite sequencing data showed a significant increase in DNase I cleavage adjacent to methylated CpGs over unmethylated CpGs and this DNase hypersensitivity was attributed to methylation-induced minor groove narrowing [51]. The electron-withdrawing effect of the formyl and carboxyl groups and the concomitant weakening of the 5fC:G and 5caC:G base pairing was postulated to facilitate the flipping out of the double helix of these modified bases for recognition and excision by the thymine-DNA glycosylase (TDG) during the DNA demethylation pathway [50]. Accordingly, minor groove alterations in formylated and carboxylated DNA were proposed to be specifically recognized by TDG, resulting in its substrate specificity [26,45]. The contradictory results found by others, however, provide doubts to the 5fC-induced structural changes of DNA being a weighty DNA recognition platform for proteins, as the changes are rather subtle and very local [35,45]. In general, (modified) DNA readout can depend on base or shape readout. Both can be influenced by modification of cytosines; the addition of a bulky methyl group to the major groove could facilitate or interfere with protein side chain contacts [52], whereas the local structural change in DNA upon base modification may allow or inhibit protein recognition and binding [51]. Hence, changes in DNA shape were also demonstrated to play a role in enhancing DNA binding of the transcription factor p53 [53].

The changes in groove geometry that come along with cytosine methylation were related to alterations in solvation dynamics and water densities compared to unmodified DNA [32]. Conformingly and in view of

DNA shape variations as a mechanism for specific protein recognition, MeCP2 was found to interact and recognize specific water molecules within the hydration shell of the major groove of methylated DNA, rather than the methylated cytosine *per se* [54]. Additionally, the above-mentioned narrowing of the minor groove observed in 5mC containing DNA enhances the negative electrostatic potential. The latter, in turn, can more efficiently attract basic side chains of DNA binding proteins [44,51,55]. Of note, the MBD domains of MeCP2, MBD1, and MBD2 were found to identify the mCpG dinucleotide via two conserved positively charged arginine rich fingers [54,56,57], which might exploit the described changes for binding site identification. Besides the recognition/binding of mCpG dinucleotides via its MBD domain, MeCP2 exhibits two conserved AT-hooks that specifically recognize adjacent AT-runs [58], which were also shown to have a tendency for binding narrow minor grooves [55]. It is therefore tempting to speculate about a possible DNA shape-based recognition mechanism of 5mC that relies on local changes of the DNA double helix structure arising upon base modification.

## DNMT writers, MBD readers, and TET modifiers

### DNMT writer enzymes and the 5mC mark

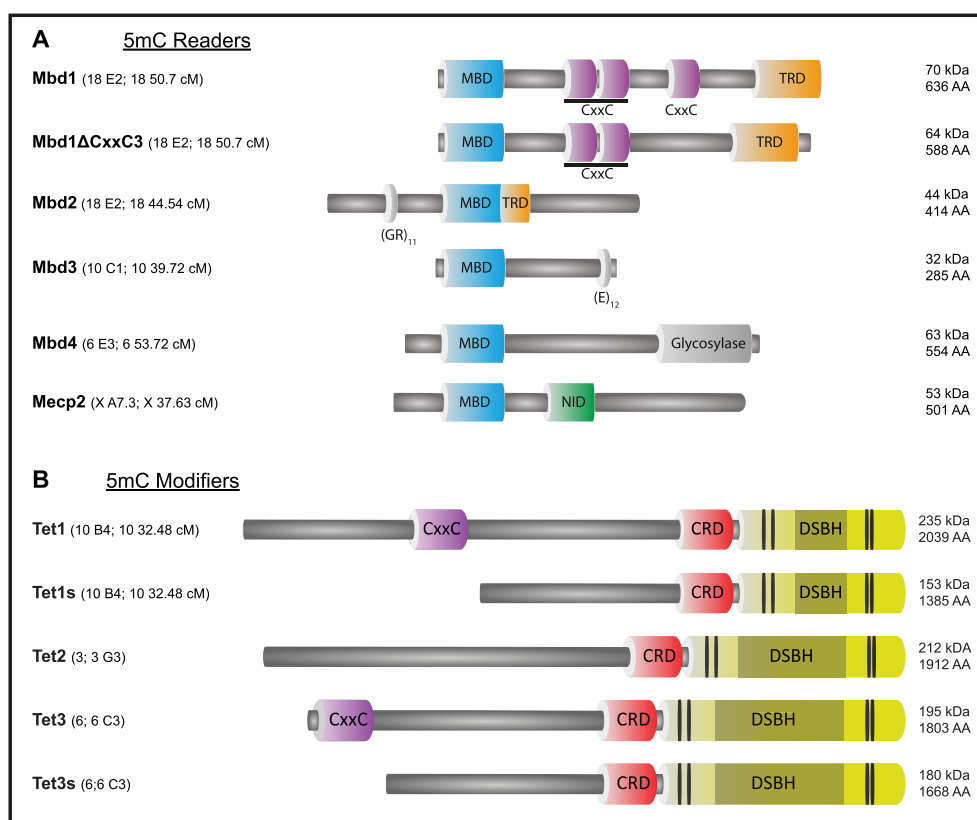
Methylation of cytosine in eukaryotic DNA relies on the action of the members of the S-adenosyl-L-methionine dependent DNA methyltransferase family (DNMT), which are capable of introducing new DNA methylation sites, as well as maintaining the methylation state upon DNA replication during the cell cycle. DNMT1, the first member of this family, can be considered an epigenetic janitor, since it methylates hemimethylated CpG sites that occur in the course of DNA replication and repair [59–61]. DNMT3a and DNMT3b, in contrast, are *de novo* methyltransferases; both show a strong affinity for unmethylated DNA and establish the initial DNA methylation patterns during embryogenesis and gametogenesis. Both proteins are essential for pluripotency repression during cell differentiation and development, are highly expressed in ES and germ cells, but are downregulated upon differentiation and in general in somatic tissues [62–64]. While DNMT3a was shown to play an important role in setting parental imprints [63], DNMT3b was linked to the correct methylation pattern of pericentric DNA regions and, hence, chromosome structure [62]. The physiological importance of DNMT1, DNMT3a, and DNMT3b is underlined by the negative effects their depletion has on organismic development and survival, as well as their link to human diseases.

The majority of cytosines in the human genome that reside within CpG dinucleotides are marked by a methyl group [3,4], and, due to the mutagenic potential of 5mC, a global CpG depletion is observed [65–67]. However, the majority of the CpGs are hypermethylated and found mostly in repetitive and parasitic DNA elements, such as retrotransposons, LINE1 (long interspersed nuclear elements), and Alu elements [68,69], whose silencing depends on CpG methylation. Recent studies showed that transcriptional regulation via DNA methylation is context-dependent, with methylation within promoter regions generally correlated with gene repression, and gene body methylation typically related to gene activation [70–73]. In addition to CpG methylation, recent sequencing-based studies demonstrated the presence of non-CpG methylation in pluripotent cells and neurons [74–78]. Interestingly, the methyl-CpG binding protein MeCP2 was recently shown to

recognize and bind methylated cytosines in non-CpG contexts (mCAC) and to regulate the repression of long genes in the brain [56,79–82].

### MBD reader proteins

DNA methylation was early-on linked to transcriptional repression amongst other mechanisms via the binding of members of the MBD protein family (Fig. 2A). The MBD (methyl-CpG binding domain) protein family is one of the best studied protein families capable of reading the genomic (hydroxy) methylome and thereby mediating between cytosine modifications, DNA and histone modifying enzyme complexes, non-coding RNAs and distinct chromatin states. The protein family comprises 11 members, namely MeCP2, MBD1-6, SETDB1, SETDB2, BAZ2A, and BAZ2B [83–88]. Besides the conserved methyl-CpG binding domain (MBD) that all



**Fig. 2. Schematic representation of selected 5mC reader and modifier proteins. (A)** Domain organization of selected members of the mouse methyl-CpG binding domain proteins including MBD1, MBD2, MBD3, MBD4, MeCP2, and a MBD1 isoform that lacks the third CXXC domain. The domains shown include the MBD (methyl-CpG binding domain), CXXC (CXXC zinc finger domain), TRD (transcriptional repression domain), and NID (NCoR/SMRT interaction domain, GR (glycine/arginine), E (glutamic acid)). **(B)** Domain organizations of the long isoforms of mouse TET1, TET2, and TET3 and the short isoforms of TET1 and TET3. Shown are the N-terminal zinc finger domain (CXXC), the cysteine-rich domain (CRD), and the double-stranded beta helix (DSBH), which is interrupted by a low complexity insert and harbors the Fe(II)- and 2-oxoglutarate binding sites. The respective chromosomal locations are given in brackets next to the names. Protein length in amino acid (AA) and the predicted size in kilo Dalton (kDa) are given next to the structures. Protein domain graphs were generated using DOG 1.0 [213].

family members share, the domain composition of MBD proteins is highly specific for each member, conferring them distinct DNA and protein binding specificities and functions. Hereafter, we will focus on the best studied members of the family, MBD1-4 and MeCP2.

### MeCP2

MeCP2 was the first protein described to selectively bind symmetrically methylated CpG dinucleotides [88], and its MBD domain was later used to identify further MBD domain containing proteins [86]. Alternative splicing leads to two MeCP2 isoforms (MeCP2 e1 and MeCP2 e2) that differ in their N-terminus [89] and are mainly considered to have similar functions although they exhibit different expression patterns in developing and postnatal mice brain [90,91]. Both variants contain an N-terminal MBD domain mediating specific binding of methylated CpGs [45] and a transcriptional repression domain (TRD) involved in transcriptional silencing [88,92]. MeCP2 DNA binding, however, is not merely mediated via its MBD domain. Regions outside this domain, including the TRD, the intervening domain (ID), and the C-terminal domain alpha, also contribute to the overall sequence unspecific binding of DNA [93,94]. Additional to the interactions with DNA, MeCP2 also harbors a chromatin interaction surface that was found to interact with histone H3 [95]. Besides the MeCP2-DNA interactions, various protein-protein interactions (reviewed in Ref. [13]), as for example, complexes with histone deacetylase 2 (HDAC2) and the transcriptional co-repressor SIN3A [96] as well as homo- and hetero-interactions of MeCP2 and MBD2 [97], were demonstrated and shown to establish and keep a repressive heterochromatin state in a dose-dependent manner [98]. MeCP2 is highly expressed in the central nervous system, especially in postmitotic neurons [99,100]. *MeCP2*<sup>-/-</sup> mice show a Rett syndrome like phenotype with reduced brain and neuronal cell size [101,102]. Various deletions and mutations within the *MeCP2* gene were found in Rett syndrome patients (reviewed in Ref. [13]).

### MBD1

MBD1 is the largest MBD protein family member and has, due to its unique structure, particular binding abilities and specific functions in gene regulation. Besides its N-terminal MBD and C-terminal TRD domains, MBD1 contains, isoform-dependent, two or three CXXC zinc finger motifs [85,103,104]. Yeast-2-hybrid and co-immunoprecipitation assays revealed interactions with a variety of chromatin-associated proteins involved in histone modification and replication dependent assembly,

DNA transcription and repair, polycomb repressive complex and sumoylation [105–115]. The isoform containing the first two CXXC motifs (CXXC1 and CXXC2) [103] can bind methylated CpGs via its MBD domain [116]. The presence of the third CXXC domain (CXXC3), which is homologous to, among others, the zinc finger of DNMT1, was shown to bind unmethylated CpGs [103,117], and enables MBD1 to also bind to unmodified DNA [103,118]. This dual-affinity to methylated and nonmodified DNA of the isoform containing three zinc fingers (MBD1a) enables MBD1 to regulate transcriptional repression regardless of the modification state of the DNA, whereas the isoform lacking the CXXC3 (MBD1b) can only repress expression of highly methylated genes [103,119,120]. The third zinc finger could thus facilitate MBD1 binding in genomes with low DNA methylation, as in preimplantation embryos or primordial germ cells (PGCs). Studies in fibroblast cells showed that the CXXC3 targets nonmodified CpGs, that MBD1 can repress the expression of non-methylated genes, that MBD1 harboring all three CXXC fingers can accumulate at hypomethylated constitutive heterochromatin in *Dnmt1*<sup>-/-</sup> fibroblasts, and that the deletion of CXXC3 drastically reduces localization at heterochromatin in these cells [103,121]. This is in contrast to a ChIP-seq study in mouse ES cells that showed no binding of MBD1 to unmethylated DNA, despite the presence of the CXXC3, and neither mutations of the MBD domain nor loss of DNA methylation could cause binding to nonmodified DNA [122]. This discrepancy may be related to the fact that, in the latter study, the long MBD1 isoform was over-expressed ectopically and constitutively as a fusion protein with BirA biotin ligase. Altogether these findings point towards MBD1 acting in a cell type specific mode and likely dependent on the level and/or ratio of its isoforms. Although MBD1 is expressed in a variety of tissues, high expression levels are especially found in neuronal stem cells (NSCs) and neurons. MBD1 plays an important role in the maintenance of NSC integrity and stemness and its deficiency in mice leads to accumulation of undifferentiated NSCs and impairs the transition into the neuronal lineage [123]. *Mbd1*<sup>-/-</sup> mice are viable and develop normally, but have decreased adult neurogenesis and knockout neural stem cells show less neuronal differentiation as well as increased genomic instability [124].

### MBD2

MBD2 and MBD3 are the only MBD protein family members that share high sequence similarity also outside the MBD domain, suggesting an ancestral gene duplication [85,86]. MBD2 exists in three isoforms, MBD2a, MBD2b, and MBD2c (also referred to as MBD2t), which all contain the MBD

domain. MBD2a and MBD2b arise from different translational start sites and vary only in the inclusion of an N-terminal glycine and arginine (GR) repeat in MBD2a [85,125]. Both variants contain the C-terminal TRD domain, which is essential for interactions with corepressor complexes, such as the nucleosome remodeling and histone deacetylase (NuRD) or the SIN3A complex [126–129]. MBD2c includes an alternative exon, which results in a truncated protein lacking the TRD domain due to a premature stop codon [85]. This isoform is solely expressed in testes and ES cells and does not interact with the NuRD complex, due to the lack of the TRD domain [85,122,130]. Arginine methylation within the GR motif by protein arginine methyltransferase 1 and 5 (PRMT1 and PRMT5) [131,132] regulates MBD2-DNA and MBD2-HDAC repressor complex formation and reduces the MBD2 transcriptional repression action [132]. *MBD2*<sup>-/-</sup> mice are viable and fertile and, contrary to *Mecp2*<sup>-/-</sup> and *Mbd1*<sup>-/-</sup> mice, show only a mild phenotype, including pup nurturing deficits and hypoactivity, suggesting that the impact on brain function is smaller than that of MBD1 or MeCP2 [133,134]. Alternative splicing and the resulting switch between MBD2a and MBD2c is essential for self-renewal in human pluripotent stem cells (hPSCs). While overexpression of the long isoform MBD2a in hPSCs interferes with pluripotency and promotes differentiation, potentially through downregulating NANOG and OCT4 expression via the NuRD complex, overexpressing MBD2c, which shows no interaction with the NuRD complex, does not influence hPSCs. Of note, overexpressing MBD2c in combination with reprogramming factors in fibroblasts enhanced reprogramming and the formation of induced pluripotent stem cells [130].

### MBD3

MBD3 is the smallest member of the MBD protein family, yet three isoforms are described. Besides the N-terminal MBD domain, it contains a C-terminal poly-glutamate stretch [85]. The major difference to MBD2 is that MBD3 has two crucial amino acid exchanges in its MBD domain, which impair the binding of methylated DNA. Binding to 5hmC was proposed, is however still debated [12,122,135–138]. Similar to MBD2, MBD3 is also part of the NuRD complex; however, biochemical analyses showed that they are mutually exclusive within the complex [131,139,140]. MBD3 deletion leads to early embryonic lethality in mice [133] and stem cells lacking MBD3 fail to differentiate and show aberrant self-renewal, even in the absence of leukemia inhibitory factor (LIF) [141,142]. MBD3 depletion in somatic cells together with expression of the reprogramming factors OCT4, SOX2, KLF4, and c-MYC enhanced reprogramming efficiency

[143,144], and accordingly MBD3 was proposed to play an important role in pluripotency and lineage commitment [137,145,146]. Nevertheless, opposing results were found when neural and epiblast-derived stem cells were used [147], indicating a context-dependent influence of MBD3 on pluripotency.

### MBD4

MBD4 (also known as MED1) is associated with specific functions due to its unique domain composition. Besides its N-terminal MBD domain, MBD4 contains a C-terminal glycosylase motif, unique among MBD proteins [85]. The MBD enables MBD4 binding to methylated CpG as well as mCAC sites for the human variant [56]; however, it was found to also bind 5hmC in neural progenitor cells [12]. The glycosylase domain of MBD4 is similar to the one found in 3-methyladenine DNA glycosylase II (Alka) [148], recognizes mCG/TG mismatches, and shows enzymatic activity towards T/G or U/G mismatches that occur via deamination of 5mC or cytosine, respectively [125,149]. Base removal via MBD4 relies on several steps, including a base flipping mechanism and, eventually, repair of the abasic site via BER proteins [148,150,151]. MBD4 null mice are viable and fertile, but they show increased number of CpG to TpG transitions [152]. Although MBD4 was shown to excise 5hmU *in vitro* and was therefore implicated in DNA demethylation [153], oxidation-dependent reactivation of methylated reporter genes was only possible in the presence of TDG but not MBD4 [154].

## DNA methylation and binding of MBD readers

In line with *in vitro* binding assays [85,88], genome-wide mapping of MBD protein binding in mouse ES and differentiated cells showed a general linear dependence of binding and methylation density for all MBDs except MBD3, as well as the requirement of a functional MBD domain and methylated CGs, CAs, and CACs [56,155]. For some MBDs, additional sequence preference has been reported, with high affinity binding of MeCP2 relying upon an A/T run adjacent to the mCpG [156] and the MBD1 MBD shown to preferentially bind to methylated CpGs within T<sup>m</sup>CGCA and TG<sup>m</sup>CGCA sequences [157]. These prerequisites target MBD1/2/4 and MeCP2 to methylated, regulatory, and inactive regions of the genome and, conformingly, *de novo* methylated CpG-rich regions gain MBD binding during differentiation [122]. Still, substantial binding to unmethylated CpG islands was found for MBD2 and MBD4 [122] and DNA binding of MeCP2 was previously shown to not only rely on DNA

methylation [93,155,158,159]. Sequence unspecific and unmethylated DNA binding is described to be mediated by the respective additional domains or protein-protein interactions. An example for an additional domain that confers unmethylated DNA binding is the isoform-specific CXXC3 domain of MBD1 as mentioned before [103]. MBD3 distribution is highly controversial, as binding downstream of CpG-rich and hydroxymethylated promoters, as well as CpG density and (hydroxy)methylation independent binding of regulatory regions were reported [122,137]. Some studies also report binding of MeCP2 as well as MBD4 to 5hmC [12,137,160], but this conclusion has not been validated by other studies [94,99,153].

Microscopic analysis in tissues and in cultured wild type and DNMT deficient cells of the subnuclear distribution of MBDs relative to DNA distribution, compaction, and methylation unanimously reported their enrichment at highly methylated and compacted DNA, with the exception of MBD3 [85,92,98,103,121,158]. These types of analyses by their nature do not exclude tandem or other repeat DNA sequences, which are though mostly non- or under-represented in genomics studies.

### TET modifiers and oxidized cytosine variants

#### *Ten-eleven-translocation proteins, oxidative cytosine variants, and DNA demethylation*

The 5-methylcytosine mark was for long regarded to be a relatively stable epigenetic mark, due to the chemical nature of the C-C bond between the pyrimidine C5 and the methyl group carbon. Therefore, DNA demethylation was believed to occur via replication-dependent dilution either upon inhibiting DNA methyltransferases or in their complete absence. This hypothesis, however, had to be reviewed when a replication-independent demethylation wave was found in mouse zygotes, where the paternal genome was shown to experience an almost complete loss of the methyl marks before the first round of DNA replication [161,162]. Several attempts to identify the methyl cytosine demethylase were followed by intense debate and not confirmed in other laboratories. A major breakthrough in the field of active loss of DNA methylation was achieved in 2009, when an additional modified cytosine base, 5hmC, and the corresponding writer enzymes were discovered [163,164]. Besides the sixth base 5hmC, three human homologs of trypanosomal 2-oxoglutarate- and iron(II)-dependent dioxygenases JBP1 and JBP2 were identified, named TET1, TET2, and TET3 (Fig. 2B). The acronym TET stands for “Ten-Eleven Translocation” and derives from a t(10;11)(q22;q23) translocation that generates a MLL-TET1 fusion protein, which is found in acute myeloid leukemia [165,166].

TETs were shown to catalyze the oxidation reaction from 5mC to 5hmC in cultured cells as well as *in vivo* [164]. Subsequently, TET proteins were shown to further oxidize 5hmC to 5fC and 5caC [8,167,168]. After their discovery, TET proteins were positioned in the context of active DNA demethylation, assuming that the demethylation reaction relies on the stepwise oxidation of 5mC to 5caC by TETs, followed by thymine DNA glycosylase (TDG) or Nei-like 1 (NEIL1) glycosylase mediated excision of 5fC or 5caC [9,154,169–171]. The resulting abasic site is subsequently targeted by base-excision repair (BER) proteins, leading to the insertion of a nonmodified cytosine [167,172]. TET enzymes and oxidative cytosine variants are suggested to play an essential role in major demethylation cascades during embryonic development [162,173–179] and accordingly, the loss of functional TET and TDG enzymes impairs mouse ES cell differentiation and is embryonically lethal in mice [101,102]. Recent reports, however, suggest that 5hmC and 5fC can also act as stable epigenetic marks, as opposed to being mere DNA demethylation intermediates [9–11,171]. Quantification of genomic cytosine modifications revealed drastic differences in the abundance of the different derivatives. Whereas methylated cytosines account for ~4.5% of all cytosines, 5hmC levels are ten-fold lower, with 0.45% of all cytosines in average in brain tissue and 0.39% in mouse embryonic stem (ES) cells [8,168,180]. In mouse ES cells, 5fC levels are 10 to 100-fold lower than 5hmC levels, only 0.02% to 0.002% of all cytosines carry the formyl mark [8,168] and 5caC is even less abundant, accounting for only 0.0002% of all cytosines [8].

#### *Genomic distribution of oxidative cytosine variants*

Genome-wide studies in ES cells and brain tissues found 5hmC to be enriched in euchromatic chromatin regions and especially at/around transcriptional start sites, promoters with moderate and low CpG content and gene bodies, where the amount is positively correlated to gene expression [181–185]. In ES cells, the oxidative mark is widespread but predominantly found at developmentally regulated bivalent genes that are polycomb repressive complex 2 (PRC2) targets, near transcription factor binding sites, at gene enhancers and CTCF binding sites as well as DNase I hypersensitive chromatin [186–188]. In the brain, hydroxymethylated cytosines are more enriched at poised enhancers than at active enhancers [189]. Similar to 5hmC, 5fC is found at promoters of actively transcribed genes, H3K4me3 and RNA polymerase II marked transcriptional start sites, exons, and poised enhancers in ES cells [190–192].

### TET isoforms

A structural feature that separates the three TET proteins from one another and finetunes their respective biological function in different developmental stages and tissues is their N-terminal CXXC domain. Although the zinc finger domains of TET1 and TET3 as well as the TET2 associated IDAX/CXXC4 belong to the same subgroup of CXXC-domains [193], they highly differ in their respective binding affinities and substrates. TET2 lost its zinc finger during evolution in a chromosomal inversion, and it is now encoded by the genomically adjacent IDAX/CXXC4, which negatively regulates TET2 expression and modulates its catalytic activity [194]. TET1 and TET3, in contrast, kept their respective zinc finger domains and the CXXC domain of TET3 was shown to preferentially bind 5caC [195]. However, TET1 CXXC domain was found to bind mostly nonmodified DNA [117], which was implicated in the prevention of DNA methylation spreading into unmethylated, euchromatic regions [196]. While three different isoforms of TET2 and also TET3 have been characterized to date [195,197], an N-terminally truncated TET1 isoform, which lacks the CXXC domain, was described recently and attributed a role in mouse development, but also in cancerogenesis [198,199].

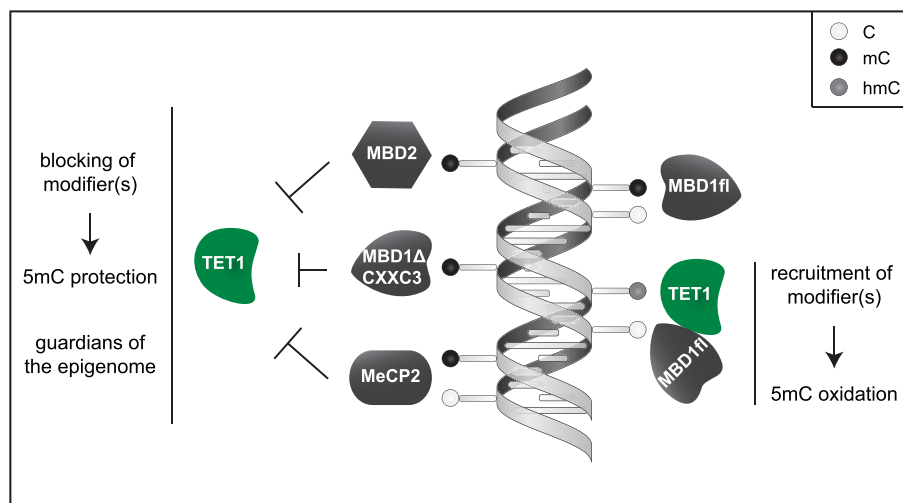
### Interplay of DNA methylation readers and modifiers

Cytosine methylation writers, readers, and modifiers need a tight spatio-temporal regulation to avoid aberrant DNA methylation, a hallmark of numerous human disorders. Deregulation of members of the DNMT, MBD, or TET protein families and their interplay can impair the genomic methylome, as well as chromatin structure and organization, as observed in patients suffering from the neurodegenerative HSNIE disorder (hereditary sensory and autonomic neuropathy, mutations in the TS domain of DNMT1), the Rett syndrome (MeCP2 mutations), or AML (acute myeloid leukemia, TET2 mutations/deletions) [98,200–202]. Shaping of DNA modification patterns can be regulated by the localization, activity, stability, interactome, substrate preference, and cofactor availability of the writer and modifier enzymes [168,203–209]. As previously discussed, the distribution of DNA modifications throughout the genome is not random but tightly associated with highly specific genomic regions and concomitantly presenting distinct protein recognition platforms. Consequently, the binding profile of cytosine modification reader proteins is greatly influenced and defined by the genomic landscape of the modifications and the latter is proposed to shape the genomic location of the

readers. Indeed, the catalytic activity of TET proteins was found to affect the binding of different members of the MBD protein family [12]. However, since MBD readers and TET modifiers share (at least) one common substrate, the opposite role can also be envisaged: recognition and binding of 5mC (and/or 5hmC) by MBDs modulate the activity of TET proteins and in turn the epigenetic makeup of DNA, its physical properties, and hence DNA metabolism. According to this model, the reader proteins shape the DNA modification distribution via regulation of the accessibility of the modified DNA bases to modifier enzymes.

Hashimoto and coworkers found by *in vitro* inhibition experiments that binding of methylated DNA by MeCP2 inhibits TET1 activity, thereby suggesting a potential mechanism for 5hmC level regulation *in vivo* [153]. Another study suggested that MeCP2 binding of 5mC itself was correlated with lack of 5mC oxidation [210]. Recently, a more indepth analysis on the interplay of 5mC, MBD, and TET proteins interrogated whether and how binding of MBD proteins to a methylated substrate influences TET-mediated DNA demethylation. Interestingly and in line with the respective substrate preference, especially the binding of the MBD protein family members MeCP2 and MBD2 was found to protect 5mC from TET1 mediated oxidation. Importantly, quantification of the TET1 level demonstrated that it was similar to the physiological levels and the same was the case for MeCP2 levels that mimicked its concentration in neuronal cells. Hence, the conditions used reflected the physiological situation. This blocking of TET-mediated 5mC oxidation was MBD protein concentration dependent and relied both on MeCP2 and MBD2 binding 5mC directly via their MBD or, in the case of MeCP2, additionally by its binding to nonmodified DNA (Fig. 3). Thus, protection of TET1-dependent 5mC oxidation arose from direct blocking of access to the DNA as well as 5mC.

To ensure genomic stability, minimize transcriptional noise of noncoding genomic repeats (e.g., major satellite repeats, MaSat), and insertions of mobile retrotransposons (e.g., long interspersed nuclear elements, LINES), cells developed a variety of mechanisms to silence the expression of these elements, the most important being cytosine methylation. MBD protein-mediated blocking of 5mC oxidation was shown to prevent the reactivation of silenced repetitive elements such as major satellite repeats or LINE elements, thereby maintaining chromatin composition, as well as genome stability [158,211]. Accordingly, chromocentric 5hmC levels in the pons of a mouse model for Rett syndrome (MeCP2<sup>y/-</sup>) were significantly increased and *in situ* RNA-FISH as well as quantitative PCR analyses revealed higher levels of major satellite expression compared to wildtype. This suggested that



**Fig. 3. Role of MBD proteins in modifying and protecting cytosine modifications.** Graphical summary of the specific DNA binding abilities of MBD proteins and their influence on the DNA base modifier TET1. Whereas MBD1 full-length (fl) recruits TET1 to methylated cytosine thereby enhancing 5-hydroxymethylcytosine generation, MeCP2, MBD2, and MBD1 $\Delta$ CXXC3 bind 5-methylcytosine, blocking TET1 access and protecting 5-methylcytosine from oxidation. C = cytosine, mC = 5-methylcytosine, hmC = 5-hydroxymethylcytosine.

unrestricted TET1 activity might increase repeat element transcription in animal models of Rett syndrome. Thus, these studies disclose the potential of a pathophysiological pathway of TET enzymes in Rett syndrome, as well as a role of MeCP2 and MBD2 as guardians of the epigenome and in genome stability.

In contrast to MeCP2 and MBD2, MBD3 and MBD4 were not found to influence TET1-mediated 5hmC formation. On the other hand, MBD1 was shown to enhance TET1, but not TET2 and TET3, mediated 5mC oxidation. The underlying mechanism relies on an MBD1-dependent increase of TET1 localization at highly methylated heterochromatin, and this recruitment is favored by protein-protein interactions between MBD1 and TET1 [121]. Interestingly, the recruitment is cell-cycle stage independent. Akin to the blocking effect observed upon MeCP2 or MBD2 binding to 5mC, the MBD1-dependent TET1 recruitment effect observed in low TET1 overexpressing cells reflects physiological situation comparable to mouse ES cells. Recruitment and catalytic activity of TET1 at heterochromatin ultimately leads to the displacement of MBD1, as 5hmC is a less favored binding substrate relative to 5mC. Interestingly, MBD1-mediated enhancement of TET1 localization at heterochromatin as well as 5mC oxidation is MBD1 isoform dependent (Fig. 2A). In contrast to the MBD1 isoform containing all three CXXC domains (MBD1 fl or MBD1a), the short isoform lacking the third zinc finger (MBD1 $\Delta$ CXXC3 or MBD1b) blocks TET1-mediated 5hmC formation, suggesting different biological functions of the MBD1 isoforms (Fig. 3). This

CXXC3 dependency might indicate that the latter targets MBD1 to CpG-dense regions via binding to unmethylated CpGs, recruiting TET1 without simultaneously blocking its ability to oxidize methylated DNA substrates [121].

### Concluding remarks/Perspectives

Altogether these studies unveil new mechanisms of a time and dose-dependent cross-regulation of cytosine modifications and their readers and modifiers, relying on inhibiting each other's binding. Until now, DNA modifications are generally seen as shapers of the reader's binding profiles, and, hence, their genomic localization [160,212]. Conversely, cytosine modification readers do not only play important roles in translating the modifications into distinct chromatin states, thereby regulating transcription and conserving genome integrity, but also in shaping the genomic modification landscape by protecting from or enhancing oxidation.

### Acknowledgments

We thank all the past and present members of our laboratory for their many contributions along the years. Last but not least, we thank our collaborators over the years, who have made our work so much more enjoyable.

## Author contributions

Florian D. Hastert, Cathia Rausch: Conceptualization, Writing—Original Draft Preparation, Visualization.

M. Cristina Cardoso: Conceptualization, Writing—Reviewing and Editing, Funding Acquisition.

## Declarations of interest

None.

## Funding

Our research has been supported by grants of the German Research Foundation, Germany (DFG CA 198/12-1 to M.C.C.).

Received 14 September 2019;

Received in revised form 28 November 2019;

Accepted 5 December 2019

Available online 20 December 2019

### Keywords:

chromatin;  
DNA helix stability;  
DNA methylation readers;  
DNA modifications;  
DNA modification writers

### Abbreviations used:

5caC, 5-carboxylcytosine; 5fC, 5-formylcytosine; 5hmC, 5-hydroxymethylcytosine; 5mC, 5-methylcytosine; C, Cytosine; CGI, CpG island; CpG, Cytosine-phosphate-guanine dinucleotide; CXXC, Cysteine/X/X/cysteine zinc finger motif; DNMT, DNA methyltransferase; ES cells, Embryonic stem cells; ID, Intervening domain; MBD, Methyl-CpG-binding domain; TET, Ten-eleven translocation;  $T_m$ , Melting temperature; TRD, Transcriptional repression domain.

## References

- [1] R.D. Hotchkiss, The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography, *J. Biol. Chem.* 175 (1948) 315–332.
- [2] N.W. Penn, R. Suwalski, C. O'Riley, K. Bojanowski, R. Yura, The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid, *Biochem. J.* 126 (1972) 781–790.
- [3] M. Ehrlich, M.A. Gama-Sosa, L.H. Huang, R.M. Midgett, K.C. Kuo, R.A. McCune, et al., Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells, *Nucleic Acids Res.* 10 (1982) 2709–2721.
- [4] A. Bird, DNA methylation patterns and epigenetic memory, *Genes Dev.* 16 (2002) 6–21.
- [5] E. Li, C. Beard, R. Jaenisch, Role for DNA methylation in genomic imprinting, *Nature* 366 (1993) 362–365.
- [6] M.M. Suzuki, A. Bird, DNA methylation landscapes: provocative insights from epigenomics, *Nat. Rev. Genet.* 9 (2008) 465–476.
- [7] C.P. Walsh, J.R. Chaillet, T.H. Bestor, Transcription of IAP endogenous retroviruses is constrained by cytosine methylation, *Nat. Genet.* 20 (1998) 116–117.
- [8] S. Ito, L. Shen, Q. Dai, S.C. Wu, L.B. Collins, J.A. Swenberg, et al., Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine, *Science* 333 (2011) 1300–1303.
- [9] R.M. Kohli, Y. Zhang, TET enzymes, TDG and the dynamics of DNA demethylation, *Nature* 502 (2013) 472–479.
- [10] M. Bachman, S. Uribe-Lewis, X. Yang, H.E. Burgess, M. Iurlaro, W. Reik, et al., 5-Formylcytosine can be a stable DNA modification in mammals, *Nat. Chem. Biol.* 11 (2015) 555–557.
- [11] M. Bachman, S. Uribe-Lewis, X. Yang, M. Williams, A. Murrell, S. Balasubramanian, 5-Hydroxymethylcytosine is a predominantly stable DNA modification, *Nat. Chem.* 6 (2014) 1049–1055.
- [12] C.G. Spruijt, F. Gnerlich, A.H. Smits, T. Pfaffeneder, P.W. Jansen, C. Bauer, et al., Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives, *Cell* 152 (2013) 1146–1159.
- [13] A.K. Ludwig, P. Zhang, M.C. Cardoso, Modifiers and readers of DNA modifications and their impact on genome structure, expression, and stability in disease, *Front. Genet.* 7 (2016) 115.
- [14] M.J. Koziol, C.R. Bradshaw, G.E. Allen, A.S.H. Costa, C. Frezza, J.B. Gurdon, Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications, *Nat. Struct. Mol. Biol.* 23 (2016) 24–30.
- [15] T.P. Wu, T. Wang, M.G. Seetin, Y. Lai, S. Zhu, K. Lin, et al., DNA methylation on N(6)-adenine in mammalian embryonic stem cells, *Nature* 532 (2016) 329–333.
- [16] G.Z. Luo, Z. Hao, L. Luo, M. Shen, D. Sparvoli, Y. Zheng, et al., N(6)-methyldeoxyadenosine directs nucleosome positioning in Tetrahymena DNA, *Genome Biol.* 19 (2018) 200.
- [17] S. Schiffers, C. Ebert, R. Rahimoff, O. Kosmatchev, J. Steinbacher, A.V. Bohne, et al., Quantitative LC-MS provides No evidence for m6 dA or m4 dC in the genome of mouse embryonic stem cells and tissues, *Angew. Chem. Int. Ed.* 56 (2017) 11268–11271.
- [18] A.A. Travers, The structural basis of DNA flexibility, *Philos. Trans. Ser. Math. Phys. Eng. Sci.* 362 (2004) 1423–1438.
- [19] J. Marmur, P. Doty, Heterogeneity in deoxyribonucleic acids. I. Dependence on composition of the configurational stability of deoxyribonucleic acids, *Nature* 183 (1959) 1427–1429.
- [20] J.D. Hoheisel, A.G. Craig, H. Lehrach, Effect of 5-bromo- and 5-methyldeoxycytosine on duplex stability and discrimination of the NotI octadeoxynucleotide. Quantitative measurements using thin-layer chromatography, *J. Biol. Chem.* 265 (1990) 16656–16660.
- [21] A. Lefebvre, O. Mauffret, S. el Antri, M. Monnot, E. Lescot, S. Fermandjian, Sequence dependent effects of CpG cytosine methylation. A joint 1H-NMR and 31P-NMR study, *Eur. J. Biochem.* 229 (1995) 445–454.
- [22] C. Acosta-Silva, V. Branchadell, J. Bertran, A. Oliva, Mutual relationship between stacking and hydrogen bonding in DNA. Theoretical study of guanine-cytosine, guanine-5-

- methylcytosine, and their dimers, *J. Phys. Chem. B* 114 (2010) 10217–10227.
- [23] M. Banyay, A. Graslund, Structural effects of cytosine methylation on DNA sugar pucker studied by FTIR, *J. Mol. Biol.* 324 (2002) 667–676.
- [24] L. Lercher, M.A. McDonough, A.H. El-Sagheer, A. Thalhammer, S. Kriacionis, T. Brown, et al., Structural insights into how 5-hydroxymethylation influences transcription factor binding, *Chem. Commun.* 50 (2014) 1794–1796.
- [25] C.M. Lopez, A.J. Lloyd, K. Leonard, M.J. Wilkinson, Differential effect of three base modifications on DNA thermostability revealed by high resolution melting, *Anal. Chem.* 84 (2012) 7336–7342.
- [26] E.A. Raiber, P. Murat, D.Y. Chirgadze, D. Beraldi, B.F. Luisi, S. Balasubramanian, 5-Formylcytosine alters the structure of the DNA double helix, *Nat. Struct. Mol. Biol.* 22 (2015) 44–49.
- [27] D. Renciuik, O. Blacque, M. Vorlickova, B. Spingler, Crystal structures of B-DNA dodecamer containing the epigenetic modifications 5-hydroxymethylcytosine or 5-methylcytosine, *Nucleic Acids Res.* 41 (2013) 9891–9900.
- [28] P.M. Severin, X. Zou, H.E. Gaub, K. Schulten, Cytosine methylation alters DNA mechanical properties, *Nucleic Acids Res.* 39 (2011) 8740–8751.
- [29] P.M. Severin, X. Zou, K. Schulten, H.E. Gaub, Effects of cytosine hydroxymethylation on DNA strand separation, *Biophys. J.* 104 (2013) 208–215.
- [30] Q. Song, Z. Qiu, H. Wang, Y. Xia, J. Shen, Y. Zhang, The effect of methylation on the hydrogen-bonding and stacking interaction of nucleic acid bases, *Struct. Chem.* 24 (2013) 55–65.
- [31] A. Thalhammer, A.S. Hansen, A.H. El-Sagheer, T. Brown, C.J. Schofield, Hydroxylation of methylated CpG dinucleotides reverses stabilisation of DNA duplexes by cytosine 5-methylation, *Chem. Commun.* 47 (2011) 5325–5327.
- [32] M. Wanunu, D. Cohen-Karni, R.R. Johnson, L. Fields, J. Benner, N. Peterman, et al., Discrimination of methylcytosine from hydroxymethylcytosine in DNA molecules, *J. Am. Chem. Soc.* 133 (2011) 486–492.
- [33] P. Zhang, F.D. Hastert, A.K. Ludwig, K. Breitwieser, M. Hofstätter, M.C. Cardoso, DNA base flipping analytical pipeline, *Biol. Methods Protoc.* 2 (2017) 1–12.
- [34] A.T.P. Carvalho, M.L. Gouveia, C. Raju Kanna, S. Warmlander, J. Platts, S.C.L. Kamerlin, Theoretical modelling of epigenetically modified DNA sequences, *F1000Research* 4 (2015) 52.
- [35] J.S. Hardwick, D. Ptchelkine, A.H. El-Sagheer, I. Tear, D. Singleton, S.E.V. Phillips, et al., 5-Formylcytosine does not change the global structure of DNA, *Nat. Struct. Mol. Biol.* 24 (2017) 544–552.
- [36] A. Moser, R. Guza, N. Tretyakova, D.M. York, Density functional study of the influence of C5 cytosine substitution in base pairs with guanine, *Theor. Chem. Acc.* 122 (2009) 179–188.
- [37] A. Perez, C.L. Castellazzi, F. Battistini, K. Collinet, O. Flores, O. Deniz, et al., Impact of methylation on the physical properties of DNA, *Biophys. J.* 102 (2012) 2140–2148.
- [38] A.T. Carvalho, L. Gouveia, C.R. Kanna, S.K. Warmlander, J.A. Platts, S.C. Kamerlin, Understanding the structural and dynamic consequences of DNA epigenetic modifications: computational insights into cytosine methylation and hydroxymethylation, *Epigenetics* 9 (2014) 1604–1612.
- [39] J. Horberg, A. Reymer, A sequence environment modulates the impact of methylation on the torsional rigidity of DNA, *Chem. Commun.* 54 (2018) 11885–11888.
- [40] S. Rao, T.P. Chiu, J.F. Kribelbauer, R.S. Mann, H.J. Bussemaker, R. Rohs, Systematic prediction of DNA shape changes due to CpG methylation explains epigenetic effects on protein-DNA binding, *Epigenet. Chromatin* 11 (2018) 6.
- [41] A. Mendonca, E.H. Chang, W. Liu, C. Yuan, Hydroxymethylation of DNA influences nucleosomal conformation and stability in vitro, *Biochim. Biophys. Acta* 1839 (2014) 1323–1329.
- [42] T.T. Ngo, J. Yoo, Q. Dai, Q. Zhang, C. He, A. Aksimentiev, et al., Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability, *Nat. Commun.* 7 (2016) 10813.
- [43] A. Jeltsch, Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases, *Chem-biochem : Euro. J. Chem. Biol.* 3 (2002) 274–293.
- [44] A.C. Dantas Machado, T. Zhou, S. Rao, P. Goel, C. Rastogi, A. Lazarovici, et al., Evolving insights on how cytosine methylation affects protein-DNA binding, *Briefings Funct. Genom.* 14 (2015) 61–73.
- [45] T. Fu, L. Liu, Q.L. Yang, Y. Wang, P. Xu, L. Zhang, et al., Thymine DNA glycosylase recognizes the geometry alteration of minor grooves induced by 5-formylcytosine and 5-carboxylcytosine, *Chem. Sci.* 10 (2019) 7407–7417.
- [46] J. Norberg, M. Vihinen, Molecular dynamics simulation of the effects of cytosine methylation on structure of oligonucleotides, *J. Mol. Struct. THEOCHEM* 546 (2001) 51–62.
- [47] L.C. Sowers, B.R. Shaw, W.D. Sedwick, Base stacking and molecular polarizability: effect of a methyl group in the 5-position of pyrimidines, *Biochem. Biophys. Res. Commun.* 148 (1987) 790–794.
- [48] M.W. Szulik, P.S. Pallan, B. Nocek, M. Voehler, S. Banerjee, S. Brooks, et al., Differential stabilities and sequence-dependent base pair opening dynamics of Watson-Crick base pairs with 5-hydroxymethylcytosine, 5-formylcytosine, or 5-carboxylcytosine, *Biochemistry* 54 (2015) 1294–1305.
- [49] M.B. Helabad, N. Kanaan, P. Imhof, Base flip in DNA studied by molecular dynamics simulations of differently-oxidized forms of methyl-Cytosine, *Int. J. Mol. Sci.* 15 (2014) 11799–11816.
- [50] Q. Dai, P.J. Sanstead, C.S. Peng, D. Han, C. He, A. Tokmakoff, Weakened N3 hydrogen bonding by 5-formylcytosine and 5-carboxylcytosine reduces their base-pairing stability, *ACS Chem. Biol.* 11 (2016) 470–477.
- [51] A. Lazarovici, T. Zhou, A. Shafer, A.C. Dantas Machado, T.R. Riley, R. Sandstrom, et al., Probing DNA shape and methylation state on a genomic scale with DNase I, *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) 6376–6381.
- [52] Y. Liu, X. Zhang, R.M. Blumenthal, X. Cheng, A common mode of recognition for methylated CpG, *Trends Biochem. Sci.* 38 (2013) 177–183.
- [53] M. Kitayner, H. Rozenberg, R. Rohs, O. Suad, D. Rabinovich, B. Honig, et al., Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs, *Nat. Struct. Mol. Biol.* 17 (2010) 423–429.
- [54] K.L. Ho, I.W. McNaie, L. Schmiedeburg, R.J. Klose, A.P. Bird, M.D. Walkinshaw, MeCP2 binding to DNA depends upon hydration at methyl-CpG, *Mol. Cell* 29 (2008) 525–531.

- [55] R. Rohs, S.M. West, A. Sosinsky, P. Liu, R.S. Mann, B. Honig, The role of DNA shape in protein-DNA recognition, *Nature* 461 (2009) 1248–1253.
- [56] K. Liu, C. Xu, M. Lei, A. Yang, P. Loppnau, T.R. Hughes, et al., Structural basis for the ability of MBD domains to bind methyl-CG and TG sites in DNA, *J. Biol. Chem.* 293 (2018) 7344–7354.
- [57] I. Ohki, N. Shimotake, N. Fujita, J. Jee, T. Ikegami, M. Nakao, et al., Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA, *Cell* 105 (2001) 487–497.
- [58] M.J. Lyst, J. Connelly, C. Merusi, A. Bird, Sequence-specific DNA binding by AT-hook motifs in MeCP2, *FEBS Lett.* 590 (2016) 2927–2933.
- [59] L.S. Chuang, H.I. Ian, T.W. Koh, H.H. Ng, G. Xu, B.F. Li, Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1, *Science* 277 (1997) 1996–2000.
- [60] H. Leonhardt, A.W. Page, H.U. Weier, T.H. Bestor, A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei, *Cell* 71 (1992) 865–873.
- [61] O. Mortusewicz, L. Schermelleh, J. Walter, M.C. Cardoso, H. Leonhardt, Recruitment of DNA methyltransferase I to DNA repair sites, *Proc. Natl. Acad. Sci. U. S. A.* 102 (2005) 8905–8909.
- [62] M. Okano, D.W. Bell, D.A. Haber, E. Li, DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development, *Cell* 99 (1999) 247–257.
- [63] M. Kaneda, M. Okano, K. Hata, T. Sado, N. Tsujimoto, E. Li, et al., Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting, *Nature* 429 (2004) 900–903.
- [64] Z.D. Smith, A. Meissner, DNA methylation: roles in mammalian development, *Nat. Rev. Genet.* 14 (2013) 204–220.
- [65] A.P. Bird, DNA methylation and the frequency of CpG in animal DNA, *Nucleic Acids Res.* 8 (1980) 1499–1504.
- [66] D.N. Cooper, M. Krawczak, Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes, *Hum. Genet.* 83 (1989) 181–188.
- [67] C. Coulondre, J.H. Miller, P.J. Farabaugh, W. Gilbert, Molecular basis of base substitution hotspots in *Escherichia coli*, *Nature* 274 (1978) 775–780.
- [68] K.D. Robertson, A.P. Wolffe, DNA methylation in health and disease, *Nat. Rev. Genet.* 1 (2000) 11–19.
- [69] Y. Zheng, B.T. Joyce, L. Liu, Z. Zhang, W.A. Kibbe, W. Zhang, et al., Prediction of genome-wide DNA methylation in repetitive elements, *Nucleic Acids Res.* 45 (2017) 8697–8711.
- [70] M.P. Ball, J.B. Li, Y. Gao, J.H. Lee, E.M. LeProust, I.H. Park, et al., Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells, *Nat. Biotechnol.* 27 (2009) 361–368.
- [71] P.A. Jones, Functions of DNA methylation: islands, start sites, gene bodies and beyond, *Nat. Rev. Genet.* 13 (2012) 484–492.
- [72] D. Schubeler, Function and information content of DNA methylation, *Nature* 517 (2015) 321–326.
- [73] X. Yang, H. Han, D.D. De Carvalho, F.D. Lay, P.A. Jones, G. Liang, Gene body methylation can alter gene expression and is a therapeutic target in cancer, *Cancer Cell* 26 (2014) 577–590.
- [74] J.H. Lee, S.J. Park, K. Nakai, Differential landscape of non-CpG methylation in embryonic stem cells and neurons caused by DNMT3s, *Sci. Rep.* 7 (2017) 11295.
- [75] R. Lister, E.A. Mukamel, J.R. Nery, M. Urich, C.A. Puddifoot, N.D. Johnson, et al., Global epigenomic reconfiguration during mammalian brain development, *Science* 341 (2013) 1237905.
- [76] R. Lister, M. Pelizzola, R.H. Dowen, R.D. Hawkins, G. Hon, J. Tonti-Filippini, et al., Human DNA methylomes at base resolution show widespread epigenomic differences, *Nature* 462 (2009) 315–322.
- [77] M.D. Schultz, Y. He, J.W. Whitaker, M. Hariharan, E.A. Mukamel, D. Leung, et al., Human body epigenome maps reveal noncanonical DNA methylation variation, *Nature* 523 (2015) 212–216.
- [78] M.J. Ziller, F. Muller, J. Liao, Y. Zhang, H. Gu, C. Bock, et al., Genomic distribution and inter-sample variation of non-CpG methylation across human cell types, *PLoS Genet.* 7 (2011) e1002389.
- [79] H.W. Gabel, B. Kinde, H. Stroud, C.S. Gilbert, D.A. Harmin, N.R. Kastan, et al., Disruption of DNA-methylation-dependent long gene repression in Rett syndrome, *Nature* 522 (2015) 89–93.
- [80] J.U. Guo, Y. Su, J.H. Shin, J. Shin, H. Li, B. Xie, et al., Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain, *Nat. Neurosci.* 17 (2014) 215–222.
- [81] S. Lagger, J.C. Connelly, G. Schweikert, S. Webb, J. Selfridge, B.H. Ramsahoye, et al., MeCP2 recognizes cytosine methylated tri-nucleotide and di-nucleotide sequences to tune transcription in the mammalian brain, *PLoS Genet.* 13 (2017), e1006793.
- [82] M.J. Sperlazza, S.M. Bilinovich, L.M. Sinanan, F.R. Javier, D.C. Williams Jr., Structural basis of MeCP2 distribution on non-CpG methylated and hydroxymethylated DNA, *J. Mol. Biol.* 429 (2017) 1581–1594.
- [83] H.I. Baymaz, A. Fournier, S. Laget, Z. Ji, P.W. Jansen, A.H. Smits, et al., MBD5 and MBD6 interact with the human PR-DUB complex through their methyl-CpG-binding domain, *Proteomics* 14 (2014) 2179–2189.
- [84] S.H. Cross, R.R. Meehan, X. Nan, A. Bird, A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins, *Nat. Genet.* 16 (1997) 256–259.
- [85] B. Hendrich, A. Bird, Identification and characterization of a family of mammalian methyl-CpG binding proteins, *Mol. Cell. Biol.* 18 (1998) 6538–6547.
- [86] B. Hendrich, S. Tweedie, The methyl-CpG binding domain and the evolving role of DNA methylation in animals, *Trends Genet. : TIG (Trends Genet.)* 19 (2003) 269–277.
- [87] S. Laget, M. Joulie, F. Le Masson, N. Sasai, E. Christians, S. Pradhan, et al., The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA, *PLoS One* 5 (2010) e11982.
- [88] J.D. Lewis, R.R. Meehan, W.J. Henzel, I. Maurer-Fogy, P. Jeppesen, F. Klein, et al., Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA, *Cell* 69 (1992) 905–914.
- [89] S. Kriaucionis, A. Bird, The major form of MeCP2 has a novel N-terminus generated by alternative splicing, *Nucleic Acids Res.* 32 (2004) 1818–1823.
- [90] A. Krithivas, M. Fujimuro, M. Weidner, D.B. Young, S.D. Hayward, Protein interactions targeting the latency-associated nuclear antigen of Kaposi's sarcoma-associated

- herpesvirus to cell chromosomes, *J. Virol.* 76 (2002) 11596–11604.
- [91] Y. Li, H. Wang, J. Muffat, A.W. Cheng, D.A. Orlando, J. Lovén, et al., Global transcriptional and translational repression in human-embryonic-stem-cell-derived Rett syndrome neurons, *Cell Stem Cell* 13 (2013) 446–458.
- [92] X. Nan, F.J. Campoy, A. Bird, MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin, *Cell* 88 (1997) 471–481.
- [93] R.P. Ghosh, T. Nikitina, R.A. Horowitz-Scherer, L.M. Gierasch, V.N. Uversky, K. Hite, et al., Unique physical properties and interactions of the domains of methylated DNA binding protein 2, *Biochemistry* 49 (2010) 4395–4410.
- [94] V. Valinluck, H.H. Tsai, D.K. Rogstad, A. Burdzy, A. Bird, L.C. Sowers, Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2), *Nucleic Acids Res.* 32 (2004) 4100–4108.
- [95] T. Nikitina, X. Shi, R.P. Ghosh, R.A. Horowitz-Scherer, J.C. Hansen, C.L. Woodcock, Multiple modes of interaction between the methylated DNA binding protein MeCP2 and chromatin, *Mol. Cell. Biol.* 27 (2007) 864–877.
- [96] P.L. Jones, G.J. Veenstra, P.A. Wade, D. Vermaak, S.U. Kass, N. Landsberger, et al., Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription, *Nat. Genet.* 19 (1998) 187–191.
- [97] A. Becker, L. Allmann, M. Hofstätter, V. Casa, P. Weber, A. Lehmkühl, et al., Direct homo- and hetero-interactions of MeCP2 and MBD2, *PLoS One* 8 (2013) e53730.
- [98] A. Brero, H.P. Easwaran, D. Nowak, I. Grunewald, T. Cremer, H. Leonhardt, et al., Methyl CpG-binding proteins induce large-scale chromatin reorganization during terminal differentiation, *J. Cell Biol.* 169 (2005) 733–743.
- [99] M. Iurlaro, G. Ficiz, D. Oxley, E.A. Raiber, M. Bachman, M.J. Booth, et al., A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation, *Genome Biol.* 14 (2013) R119.
- [100] J. Barau, A. Teissandier, N. Zamudio, S. Roy, V. Nalesso, Y. Herault, et al., The DNA methyltransferase DNMT3C protects male germ cells from transposon activity, *Science* 354 (2016) 909–912.
- [101] D. Cortazar, C. Kunz, J. Selfridge, T. Lettieri, Y. Saito, E. MacDougall, et al., Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability, *Nature* 470 (2011) 419–423.
- [102] M.M. Dawlaty, A. Breiling, T. Le, M.I. Barrasa, G. Raddatz, Q. Gao, et al., Loss of Tet enzymes compromises proper differentiation of embryonic stem cells, *Dev. Cell* 29 (2014) 102–111.
- [103] H.F. Jorgensen, I. Ben-Porath, A.P. Bird, Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains, *Mol. Cell. Biol.* 24 (2004) 3387–3395.
- [104] H.H. Ng, P. Jeppesen, A. Bird, Active repression of methylated genes by the chromosomal protein MBD1, *Mol. Cell. Biol.* 20 (2000) 1394–1406.
- [105] N. Fujita, S. Watanabe, T. Ichimura, Y. Ohkuma, T. Chiba, H. Saya, et al., MCAF mediates MBD1-dependent transcriptional repression, *Mol. Cell. Biol.* 23 (2003) 2834–2843.
- [106] N. Fujita, S. Watanabe, T. Ichimura, S. Tsuruzoe, Y. Shinkai, M. Tachibana, et al., Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA methylation-based transcriptional repression, *J. Biol. Chem.* 278 (2003) 24132–24138.
- [107] U.F. Hameed, J. Lim, Q. Zhang, M.A. Wasik, D. Yang, K. Swaminathan, Transcriptional repressor domain of MBD1 is intrinsically disordered and interacts with its binding partners in a selective manner, *Sci. Rep.* 4 (2014) 4896.
- [108] T. Ichimura, S. Watanabe, Y. Sakamoto, T. Aoto, N. Fujita, M. Nakao, Transcriptional repression and heterochromatin formation by MBD1 and MCAF/AM family proteins, *J. Biol. Chem.* 280 (2005) 13928–13935.
- [109] M.J. Lyst, X. Nan, I. Stancheva, Regulation of MBD1-mediated transcriptional repression by SUMO and PIAS proteins, *EMBO J.* 25 (2006) 5317–5328.
- [110] B.E. Reese, K.E. Bachman, S.B. Baylin, M.R. Rountree, The methyl-CpG binding protein MBD1 interacts with the p150 subunit of chromatin assembly factor 1, *Mol. Cell. Biol.* 23 (2003) 3226–3236.
- [111] Y. Sakamoto, S. Watanabe, T. Ichimura, M. Kawasuji, H. Koseki, H. Baba, et al., Overlapping roles of the methylated DNA-binding protein MBD1 and polycomb group proteins in transcriptional repression of HOXA genes and heterochromatin foci formation, *J. Biol. Chem.* 282 (2007) 16391–16400.
- [112] Y. Uchimura, T. Ichimura, J. Uwada, T. Tachibana, S. Sugahara, M. Nakao, et al., Involvement of SUMO modification in MBD1- and MCAF1-mediated heterochromatin formation, *J. Biol. Chem.* 281 (2006) 23180–23190.
- [113] R. Villa, L. Morey, V.A. Raker, M. Buschbeck, A. Gutierrez, F. De Santis, et al., The methyl-CpG binding protein MBD1 is required for PML-RAR $\alpha$  function, *Proc. Natl. Acad. Sci. U. S. A.* 103 (2006) 1400–1405.
- [114] S. Watanabe, T. Ichimura, N. Fujita, S. Tsuruzoe, I. Ohki, M. Shirakawa, et al., Methylated DNA-binding domain 1 and methylpurine-DNA glycosylase link transcriptional repression and DNA repair in chromatin, *Proc. Natl. Acad. Sci. U. S. A.* 100 (2003) 12859–12864.
- [115] J. Xu, W. Zhu, W. Xu, X. Cui, L. Chen, S. Ji, et al., Silencing of MBD1 reverses pancreatic cancer therapy resistance through inhibition of DNA damage repair, *Int. J. Oncol.* 42 (2013) 2046–2052.
- [116] C. Bianchi, R. Zangi, Molecular dynamics study of the recognition of dimethylated CpG sites by MBD1 protein, *J. Chem. Inf. Model.* 55 (2015) 636–644.
- [117] C. Frauer, A. Rottach, D. Meilinger, S. Bultmann, K. Fellingner, S. Hasenoder, et al., Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1, *PLoS One* 6 (2011) e16627.
- [118] M. Nakao, S. Matsui, S. Yamamoto, K. Okumura, M. Shirakawa, N. Fujita, Regulation of transcription and chromatin by methyl-CpG binding protein MBD1, *Brain Dev.* 23 (Suppl 1) (2001) S174–S176.
- [119] N. Fujita, N. Shimotake, I. Ohki, T. Chiba, H. Saya, M. Shirakawa, et al., Mechanism of transcriptional regulation by methyl-CpG binding protein MBD1, *Mol. Cell. Biol.* 20 (2000) 5107–5118.
- [120] N. Fujita, S. Takebayashi, K. Okumura, S. Kudo, T. Chiba, H. Saya, et al., Methylation-mediated transcriptional silencing in euchromatin by methyl-CpG binding protein MBD1 isoforms, *Mol. Cell. Biol.* 19 (1999) 6415–6426.
- [121] P. Zhang, C. Rausch, F.D. Hastert, B. Boneva, A. Filatova, S.J. Patil, et al., Methyl-CpG binding domain protein 1 regulates localization and activity of Tet1 in a CXXC3

- domain-dependent manner, *Nucleic Acids Res.* 45 (2017) 7118–7136.
- [122] T. Baubec, R. Ivanek, F. Lienert, D. Schubeler, Methylation-dependent and -independent genomic targeting principles of the MBD protein family, *Cell* 153 (2013) 480–492.
- [123] E.M. Jobe, Y. Gao, B.E. Eisinger, J.K. Mladucky, C.C. Giuliani, L.E. Kelnhofer, et al., Methyl-CpG-binding protein MBD1 regulates neuronal lineage commitment through maintaining adult neural stem cell identity, *J. Neurosci.* 37 (2017) 523–536.
- [124] X. Zhao, T. Ueba, B.R. Christie, B. Barkho, M.J. McConnell, K. Nakashima, et al., Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal function, *Proc. Natl. Acad. Sci. U. S. A.* 100 (2003) 6777–6782.
- [125] B. Hendrich, C. Abbott, H. McQueen, D. Chambers, S. Cross, A. Bird, Genomic structure and chromosomal mapping of the murine and human Mbd1, Mbd2, Mbd3, and Mbd4 genes, *Mamm. Genome: Off. J. Int. Mamm. Genome Soc.* 10 (1999) 906–912.
- [126] J. Boeke, O. Ammerpohl, S. Kegel, U. Moehren, R. Renkawitz, The minimal repression domain of MBD2b overlaps with the methyl-CpG-binding domain and binds directly to Sin3A, *J. Biol. Chem.* 275 (2000) 34963–34967.
- [127] M.A. Desai, H.D. Webb, L.M. Sinanan, J.N. Scarsdale, N.M. Walavalkar, G.D. Ginder, et al., An intrinsically disordered region of methyl-CpG binding domain protein 2 (MBD2) recruits the histone deacetylase core of the NuRD complex, *Nucleic Acids Res.* 43 (2015) 3100–3113.
- [128] Q. Feng, Y. Zhang, The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes, *Genes Dev.* 15 (2001) 827–832.
- [129] M.N. Gnanapragasam, J.N. Scarsdale, M.L. Amaya, H.D. Webb, M.A. Desai, N.M. Walavalkar, et al., p66Alfa-MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2-NuRD complex, *Proc. Natl. Acad. Sci. U. S. A.* 108 (2011) 7487–7492.
- [130] Y. Lu, Y.H. Loh, H. Li, M. Cesana, S.B. Ficarro, J.R. Parikh, et al., Alternative splicing of MBD2 supports self-renewal in human pluripotent stem cells, *Cell Stem Cell* 15 (2014) 92–101.
- [131] X. Le Guezennec, M. Vermeulen, A.B. Brinkman, W.A. Hoeijmakers, A. Cohen, E. Lasonder, et al., MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties, *Mol. Cell Biol.* 26 (2006) 843–851.
- [132] C.P. Tan, S. Nakielny, Control of the DNA methylation system component MBD2 by protein arginine methylation, *Mol. Cell Biol.* 26 (2006) 7224–7235.
- [133] B. Hendrich, J. Guy, B. Ramsahoye, V.A. Wilson, A. Bird, Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development, *Genes Dev.* 15 (2001) 710–723.
- [134] K.H. Wood, B.S. Johnson, S.A. Welsh, J.Y. Lee, Y. Cui, E. Krizman, et al., Tagging methyl-CpG-binding domain proteins reveals different spatiotemporal expression and supports distinct functions, *Epigenomics* 8 (2016) 455–473.
- [135] M.F. Fraga, E. Ballestar, G. Montoya, P. Taysavang, P.A. Wade, M. Esteller, The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties, *Nucleic Acids Res.* 31 (2003) 1765–1774.
- [136] M. Saito, F. Ishikawa, The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2, *J. Biol. Chem.* 277 (2002) 35434–35439.
- [137] O. Yildirim, R. Li, J.H. Hung, P.B. Chen, X. Dong, L.S. Ee, et al., Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells, *Cell* 147 (2011) 1498–1510.
- [138] J.M. Cramer, J.N. Scarsdale, N.M. Walavalkar, W.A. Buchwald, G.D. Ginder, D.C. Williams Jr., Probing the dynamic distribution of bound states for methylcytosine-binding domains on DNA, *J. Biol. Chem.* 289 (2014) 1294–1302.
- [139] K. Gunther, M. Rust, J. Leers, T. Boettger, M. Scharfe, M. Jarek, et al., Differential roles for MBD2 and MBD3 at methylated CpG islands, active promoters and binding to exon sequences, *Nucleic Acids Res.* 41 (2013) 3010–3021.
- [140] Y. Zhang, H.H. Ng, H. Erdjument-Bromage, P. Tempst, A. Bird, D. Reinberg, Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation, *Genes Dev.* 13 (1999) 1924–1935.
- [141] K. Kaji, I.M. Caballero, R. MacLeod, J. Nichols, V.A. Wilson, B. Hendrich, The NuRD component Mbd3 is required for pluripotency of embryonic stem cells, *Nat. Cell Biol.* 8 (2006) 285–292.
- [142] K. Kaji, J. Nichols, B. Hendrich, Mbd3, a component of the NuRD co-repressor complex, is required for development of pluripotent cells, *Development* 134 (2007) 1123–1132.
- [143] M. Luo, T. Ling, W. Xie, H. Sun, Y. Zhou, Q. Zhu, et al., NuRD blocks reprogramming of mouse somatic cells into pluripotent stem cells, *Stem Cells* 31 (2013) 1278–1286.
- [144] Y. Rais, A. Zviran, S. Geula, O. Gafni, E. Chomsky, S. Viukov, et al., Deterministic direct reprogramming of somatic cells to pluripotency, *Nature* 502 (2013) 65–70.
- [145] M. Hirasaki, A. Ueda, M.N. Asaka, K. Uranishi, A. Suzuki, M. Kohda, et al., Identification of the coiled-coil domain as an essential methyl-CpG-binding domain protein 3 element for preserving lineage commitment potential of embryonic stem cells, *Stem Cells* 36 (2018) 1355–1367.
- [146] N. Reynolds, P. Latos, A. Hynes-Allen, R. Loos, D. Leaford, A. O'Shaughnessy, et al., NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment, *Cell Stem Cell* 10 (2012) 583–594.
- [147] R.L. Dos Santos, L. Tosti, A. Radzisheuskaya, I.M. Caballero, K. Kaji, B. Hendrich, et al., MBD3/NuRD facilitates induction of pluripotency in a context-dependent manner, *Cell Stem Cell* 15 (2014) 392.
- [148] P. Wu, C. Qiu, A. Sohail, X. Zhang, A.S. Bhagwat, X. Cheng, Mismatch repair in methylated DNA. Structure and activity of the mismatch-specific thymine glycosylase domain of methyl-CpG-binding protein MBD4, *J. Biol. Chem.* 278 (2003) 5285–5291.
- [149] J. Otani, K. Arita, T. Kato, M. Kinoshita, H. Kimura, I. Suetake, et al., Structural basis of the versatile DNA recognition ability of the methyl-CpG binding domain of methyl-CpG binding domain protein 4, *J. Biol. Chem.* 288 (2013) 6351–6362.

- [150] H.E. Krokan, H. Nilsen, F. Skorpen, M. Otterlei, G. Slupphaug, Base excision repair of DNA in mammalian cells, *FEBS Lett.* 476 (2000) 73–77.
- [151] B.A. Manvilla, A. Maiti, M.C. Begley, E.A. Toth, A.C. Drohat, Crystal structure of human methyl-binding domain IV glycosylase bound to abasic DNA, *J. Mol. Biol.* 420 (2012) 164–175.
- [152] C.B. Millar, J. Guy, O.J. Sansom, J. Selfridge, E. MacDougall, B. Hendrich, et al., Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice, *Science* 297 (2002) 403–405.
- [153] H. Hashimoto, Y. Liu, A.K. Upadhyay, Y. Chang, S.B. Howerton, P.M. Vertino, et al., Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation, *Nucleic Acids Res.* 40 (2012) 4841–4849.
- [154] U. Muller, C. Bauer, M. Siegl, A. Rottach, H. Leonhardt, TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation, *Nucleic Acids Res.* 42 (2014) 8592–8604.
- [155] H. Stroud, S.C. Su, S. Hrvatin, A.W. Greben, W. Renthal, L.D. Boxer, et al., Early-life gene expression in neurons modulates lasting epigenetic states, *Cell* 171 (2017) 1151–1164 e16.
- [156] R.J. Klose, S.A. Sarraf, L. Schmiedeberg, S.M. McDermott, I. Stancheva, A.P. Bird, DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG, *Mol. Cell* 19 (2005) 667–678.
- [157] T. Clouaire, J.I. de Las Heras, C. Merusi, I. Stancheva, Recruitment of MBD1 to target genes requires sequence-specific interaction of the MBD domain with methylated DNA, *Nucleic Acids Res.* 38 (2010) 4620–4634.
- [158] A.K. Ludwig, P. Zhang, F.D. Hastert, S. Meyer, C. Rausch, H.D. Herce, et al., Binding of MBD proteins to DNA blocks Tet1 function thereby modulating transcriptional noise, *Nucleic Acids Res.* 45 (2017) 2438–2457.
- [159] P.J. Skene, R.S. Illingworth, S. Webb, A.R. Kerr, K.D. James, D.J. Turner, et al., Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state, *Mol. Cell* 37 (2010) 457–468.
- [160] M. Mellén, P. Ayata, S. Dewell, S. Kriaucionis, N. Heintz, MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system, *Cell* 151 (2012) 1417–1430.
- [161] W. Mayer, A. Niveleau, J. Walter, R. Fundele, T. Haaf, Demethylation of the zygotic paternal genome, *Nature* 403 (2000) 501–502.
- [162] J. Oswald, S. Engemann, N. Lane, W. Mayer, A. Olek, R. Fundele, et al., Active demethylation of the paternal genome in the mouse zygote, *Curr. Biol. : CB (Curr. Biol.)* 10 (2000) 475–478.
- [163] S. Kriaucionis, N. Heintz, The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain, *Science* 324 (2009) 929–930.
- [164] M. Tahiliani, K.P. Koh, Y. Shen, W.A. Pastor, H. Bandukwala, Y. Brudno, et al., Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1, *Science* 324 (2009) 930–935.
- [165] R.B. Lorsbach, J. Moore, S. Mathew, S.C. Raimondi, S.T. Mukatira, J.R. Downing, TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23), *Leukemia* 17 (2003) 637–641.
- [166] R. Ono, T. Taki, T. Taketani, M. Taniwaki, H. Kobayashi, Y. Hayashi, LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23), *Cancer Res.* 62 (2002) 4075–4080.
- [167] Y.F. He, B.Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, et al., Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA, *Science* 333 (2011) 1303–1307.
- [168] T. Pfaffeneder, B. Hackner, M. Truss, M. Munzel, M. Muller, C.A. Deiml, et al., The discovery of 5-formylcytosine in embryonic stem cell DNA, *Angew. Chem.* 50 (2011) 7008–7012.
- [169] S.R. Dalton, A. Bellacosa, DNA demethylation by TDG, *Epigenomics* 4 (2012) 459–467.
- [170] A. Slyvka, K. Mierzejewska, M. Bochtler, Nei-like 1 (NEIL1) excises 5-carboxylcytosine directly and stimulates TDG-mediated 5-formyl and 5-carboxylcytosine excision, *Sci. Rep.* 7 (2017) 9001.
- [171] H. Wu, Y. Zhang, Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation, *Genes Dev.* 25 (2011) 2436–2452.
- [172] A. Maiti, A.C. Drohat, Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites, *J. Biol. Chem.* 286 (2011) 35334–35338.
- [173] J.A. Hackett, J.J. Zyllicz, M.A. Surani, Parallel mechanisms of epigenetic reprogramming in the germline, *Trends Genet. : TIG (Trends Genet.)* 28 (2012) 164–174.
- [174] P. Hajkova, S. Erhardt, N. Lane, T. Haaf, O. El-Maarri, W. Reik, et al., Epigenetic reprogramming in mouse primordial germ cells, *Mech. Dev.* 117 (2002) 15–23.
- [175] P.W.S. Hill, H.G. Leitch, C.E. Requena, Z. Sun, R. Amouroux, M. Roman-Trufero, et al., Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte, *Nature* 555 (2018) 392–396.
- [176] K. Iqbal, S.G. Jin, G.P. Pfeifer, P.E. Szabo, Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine, *Proc. Natl. Acad. Sci. U. S. A.* 108 (2011) 3642–3647.
- [177] D.M. Messerschmidt, B.B. Knowles, D. Solter, DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos, *Genes Dev.* 28 (2014) 812–828.
- [178] Y. Okamoto, N. Yoshida, T. Suzuki, N. Shimosawa, M. Asami, T. Matsuda, et al., DNA methylation dynamics in mouse preimplantation embryos revealed by mass spectrometry, *Sci. Rep.* 6 (2016) 19134.
- [179] H. Sasaki, Y. Matsui, Epigenetic events in mammalian germ-cell development: reprogramming and beyond, *Nat. Rev. Genet.* 9 (2008) 129–140.
- [180] M. Munzel, D. Globisch, T. Bruckl, M. Wagner, V. Welzmler, S. Michalakos, et al., Quantification of the sixth DNA base hydroxymethylcytosine in the brain, *Angew. Chem.* 49 (2010) 5375–5377.
- [181] M.J. Booth, M.R. Branco, G. Ficiz, D. Oxley, F. Krueger, W. Reik, et al., Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution, *Science* 336 (2012) 934–937.
- [182] G. Ficiz, M.R. Branco, S. Seisenberger, F. Santos, F. Krueger, T.A. Hore, et al., Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation, *Nature* 473 (2011) 398–402.

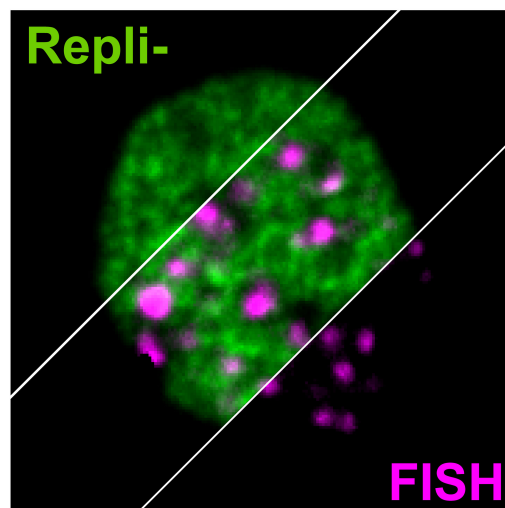
- [183] S.G. Jin, X. Wu, A.X. Li, G.P. Pfeifer, Genomic mapping of 5-hydroxymethylcytosine in the human brain, *Nucleic Acids Res.* 39 (2011) 5015–5024.
- [184] W.A. Pastor, U.J. Pape, Y. Huang, H.R. Henderson, R. Lister, M. Ko, et al., Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells, *Nature* 473 (2011) 394–397.
- [185] Y. Xu, F. Wu, L. Tan, L. Kong, L. Xiong, J. Deng, et al., Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells, *Mol. Cell* 42 (2011) 451–464.
- [186] F. Neri, D. Incarnato, A. Krepelova, S. Rapelli, A. Pagnani, R. Zecchina, et al., Genome-wide analysis identifies a functional association of Tet1 and Polycomb repressive complex 2 in mouse embryonic stem cells, *Genome Biol.* 14 (2013) R91.
- [187] A.B. Robertson, J.A. Dahl, C.B. Vagbo, P. Tripathi, H.E. Krokan, A. Klungland, A novel method for the efficient and selective identification of 5-hydroxymethylcytosine in genomic DNA, *Nucleic Acids Res.* 39 (2011) e55.
- [188] M. Yu, G.C. Hon, K.E. Szulwach, C.X. Song, L. Zhang, A. Kim, et al., Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome, *Cell* 149 (2012) 1368–1380.
- [189] L. Wen, X. Li, L. Yan, Y. Tan, R. Li, Y. Zhao, et al., Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain, *Genome Biol.* 15 (2014) R49.
- [190] F. Neri, D. Incarnato, A. Krepelova, S. Rapelli, F. Anselmi, C. Parlato, et al., Single-base resolution analysis of 5-formyl and 5-carboxyl cytosine reveals promoter DNA methylation dynamics, *Cell Rep.* 10 (2015) 674–683.
- [191] E.A. Raiber, D. Beraldi, G. Ficz, H.E. Burgess, M.R. Branco, P. Murat, et al., Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase, *Genome Biol.* 13 (2012) R69.
- [192] C.X. Song, K.E. Szulwach, Q. Dai, Y. Fu, S.Q. Mao, L. Lin, et al., Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming, *Cell* 153 (2013) 678–691.
- [193] C. Xu, K. Liu, M. Lei, A. Yang, Y. Li, T.R. Hughes, et al., DNA sequence recognition of human CXXC domains and their structural determinants, *Structure* 26 (2018) 85–95 e3.
- [194] M. Ko, J. An, H.S. Bandukwala, L. Chavez, T. Aijo, W.A. Pastor, et al., Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX, *Nature* 497 (2013) 122–126.
- [195] S.G. Jin, Z.M. Zhang, T.L. Dunwell, M.R. Harter, X. Wu, J. Johnson, et al., Tet3 reads 5-carboxylcytosine through its CXXC domain and is a potential guardian against neurodegeneration, *Cell Rep.* 14 (2016) 493–505.
- [196] C. Jin, Y. Lu, J. Jelinek, S. Liang, M.R. Estecio, M.C. Barton, et al., TET1 is a maintenance DNA demethylase that prevents methylation spreading in differentiated cells, *Nucleic Acids Res.* 42 (2014) 6956–6971.
- [197] S.M. Langemeijer, R.P. Kuiper, M. Berends, R. Knops, M.G. Aslanyan, M. Massop, et al., Acquired mutations in TET2 are common in myelodysplastic syndromes, *Nat. Genet.* 41 (2009) 838–842.
- [198] W. Zhang, W. Xia, Q. Wang, A.J. Towers, J. Chen, R. Gao, et al., Isoform switch of TET1 regulates DNA demethylation and mouse development, *Mol. Cell* 64 (2016) 1062–1073.
- [199] C.R. Good, J. Madzo, B. Patel, S. Maegawa, N. Engel, J. Jelinek, et al., A novel isoform of TET1 that lacks a CXXC domain is overexpressed in cancer, *Nucleic Acids Res.* 45 (2017) 8269–8281.
- [200] N. Agarwal, A. Becker, K.L. Jost, S. Haase, B.K. Thakur, A. Brero, et al., MeCP2 Rett mutations affect large scale chromatin organization, *Hum. Mol. Genet.* 20 (2011) 4187–4195.
- [201] M. Ko, Y. Huang, A.M. Jankowska, U.J. Pape, M. Tahiliani, H.S. Bandukwala, et al., Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2, *Nature* 468 (2010) 839–843.
- [202] M. Smets, S. Link, P. Wolf, K. Schneider, V. Solis, J. Ryan, et al., DNMT1 mutations found in HSNIE patients affect interaction with UHRF1 and neuronal differentiation, *Hum. Mol. Genet.* 26 (2017) 1522–1534.
- [203] C. Bauer, K. Gobel, N. Nagaraj, C. Colantuoni, M. Wang, U. Muller, et al., Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the O-linked N-acetylglucosamine transferase (OGT), *J. Biol. Chem.* 290 (2015) 4801–4812.
- [204] K. Blaschke, K.T. Ebata, M.M. Karimi, J.A. Zepeda-Martinez, P. Goyal, S. Mahapatra, et al., Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells, *Nature* 500 (2013) 222–226.
- [205] M.C. Cardoso, H. Leonhardt, DNA methyltransferase is actively retained in the cytoplasm during early development, *J. Cell Biol.* 147 (1999) 25–32.
- [206] S.R. Kinney, S. Pradhan, Regulation of expression and activity of DNA (cytosine-5) methyltransferases in mammalian cells, *Prog. Mol. Biol. Transl. Sci.* 101 (2011) 311–333.
- [207] W. Qin, H. Leonhardt, G. Pichler, Regulation of DNA methyltransferase 1 by interactions and modifications, *Nucleus* 2 (2011) 392–402.
- [208] F. Spada, A. Haemmer, D. Kuch, U. Rothbauer, L. Schermelleh, E. Kremmer, et al., DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells, *J. Cell Biol.* 176 (2007) 565–571.
- [209] P. Vella, A. Scelfo, S. Jammula, F. Chiacchiera, K. Williams, A. Cuomo, et al., Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells, *Mol. Cell* 49 (2013) 645–656.
- [210] K.E. Szulwach, X. Li, Y. Li, C.X. Song, H. Wu, Q. Dai, et al., 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging, *Nat. Neurosci.* 14 (2011) 1607–1616.
- [211] P. Zhang, A.K. Ludwig, F.D. Hastert, C. Rausch, A. Lehmkuhl, I. Hellmann, et al., L1 retrotransposition is activated by Ten-eleven-translocation protein 1 and repressed by methyl-CpG binding proteins, *Nucleus* 8 (2017) 548–562.
- [212] M. Mellen, P. Ayata, N. Heintz, 5-hydroxymethylcytosine accumulation in postmitotic neurons results in functional demethylation of expressed genes, *Proc. Natl. Acad. Sci. U. S. A.* 114 (2017) E7812–E7821.
- [213] J. Ren, L. Wen, X. Gao, C. Jin, Y. Xue, X. Yao, DOG 1.0: illustrator of protein domain structures, *Cell Res.* 19 (2009) 271–273.

## 4 Aims of the Study

**E**rror-free replication of the genetic and epigenetic information once every cell cycle is indispensable for genome integrity and, eventually, cell survival and persistence. Although it is well known that human and mouse genomes are, by far, not only composed of protein-coding DNA (1-2%), not much is known about the non-coding (54-59%) and repetitive (40-44%) parts of the genome. While genome- (and population) wide sequencing techniques exist, due to their nature, spatio-temporal resolution is reduced and/or indirectly obtained and repeat sequences are dramatically underrepresented. Yet, remarkable genomic reorganizations and dynamic replication timing program changes that are also affecting these neglected genomic parts, are observed during development of several species. We, therefore, developed a single cell microscopic approach, allowing detailed analysis and correlation of DNA replication and repeat elements in 4D. We investigated replication timing of repeat elements such as major and minor satellites, telomeres, satellite III, Alu and LINE-1 in mouse and human cells, respectively. To further understand how DNA replication is organized and regulated in mice, we analyzed DNA replication in pluripotent and differentiated mouse embryonic stem (mES) cells. Additionally, epigenetic DNA modifications are known to not only be mere DNA demethylation intermediates, but to also act as stable epigenetic marks in the genome. We, therefore, investigated the effect of cytosine base modifications on DNA metabolic processes, and, in particular, DNA replication. This section highlights the main (overall) aims of this thesis.

In **chapter I** (Repli-FISH (Fluorescence in Situ Hybridization): Application of 3D- (Immuno)-FISH for the Study of DNA Replication Timing of Genetic Repeat Elements) we:

- established a single cell microscopic approach to analyze the replication timing of specific DNA elements/sequences in structurally intact cell nuclei,
- developed a protocol consisting of the combined visualization of repetitive DNA sequences based on fluorescence *in situ* hybridization and the (immuno)fluorescent detection of DNA replication (Repli-FISH),
- validated this method to analyze the replication timing of repeat elements in mouse and human cells.



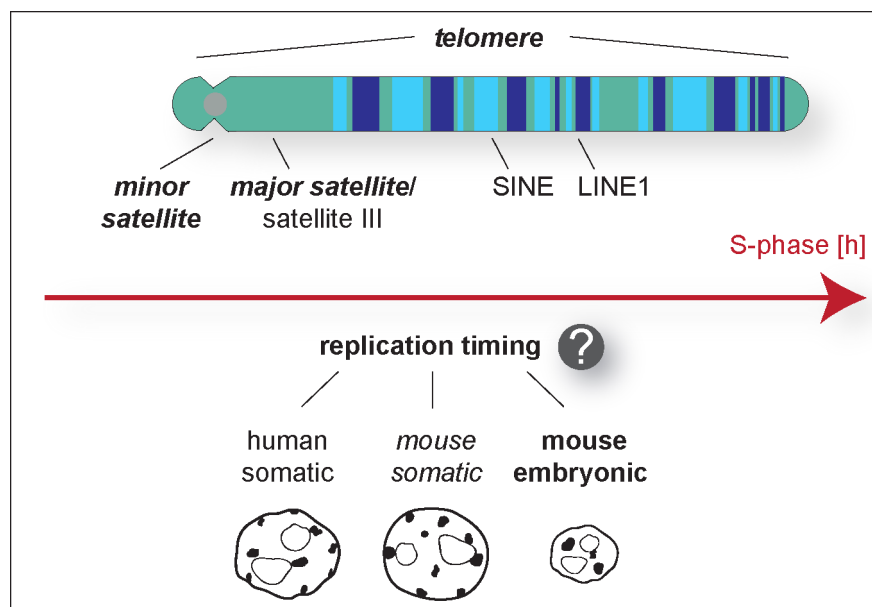
**Figure 4.1:** Microscopic image assembly summarizing the method established in chapter I.

In **chapter II** (DNA Replication and Repair Kinetics of Alu, LINE-1 and Satellite III Genomic Repetitive Elements) we investigated:

- the correlation between genome-wide replication timing of repetitive DNA elements with GC content, gene density and chromatin state,
- the replication timing of interspersed and tandem repeats by a FISH based single cell microscopic analysis (Repli-FISH),
- the genome-wide and single cell repair kinetics of repetitive DNA elements and the correlation with chromatin states.

In **chapter III** (Developmental Differences in Genome Replication Program and Origin Activation) we studied:

- the spatio-temporal S-phase progression in undifferentiated and pluripotent mES cells and compared it to the replication program in differentiated cells,
- the replication timing of specific (sub)chromosomal elements in mES cells (Repli-FISH),
- the correlation of replication timing and changes in histone modifications and chromatin state,
- origin activation and the molecular properties of the replicon in mES cells.

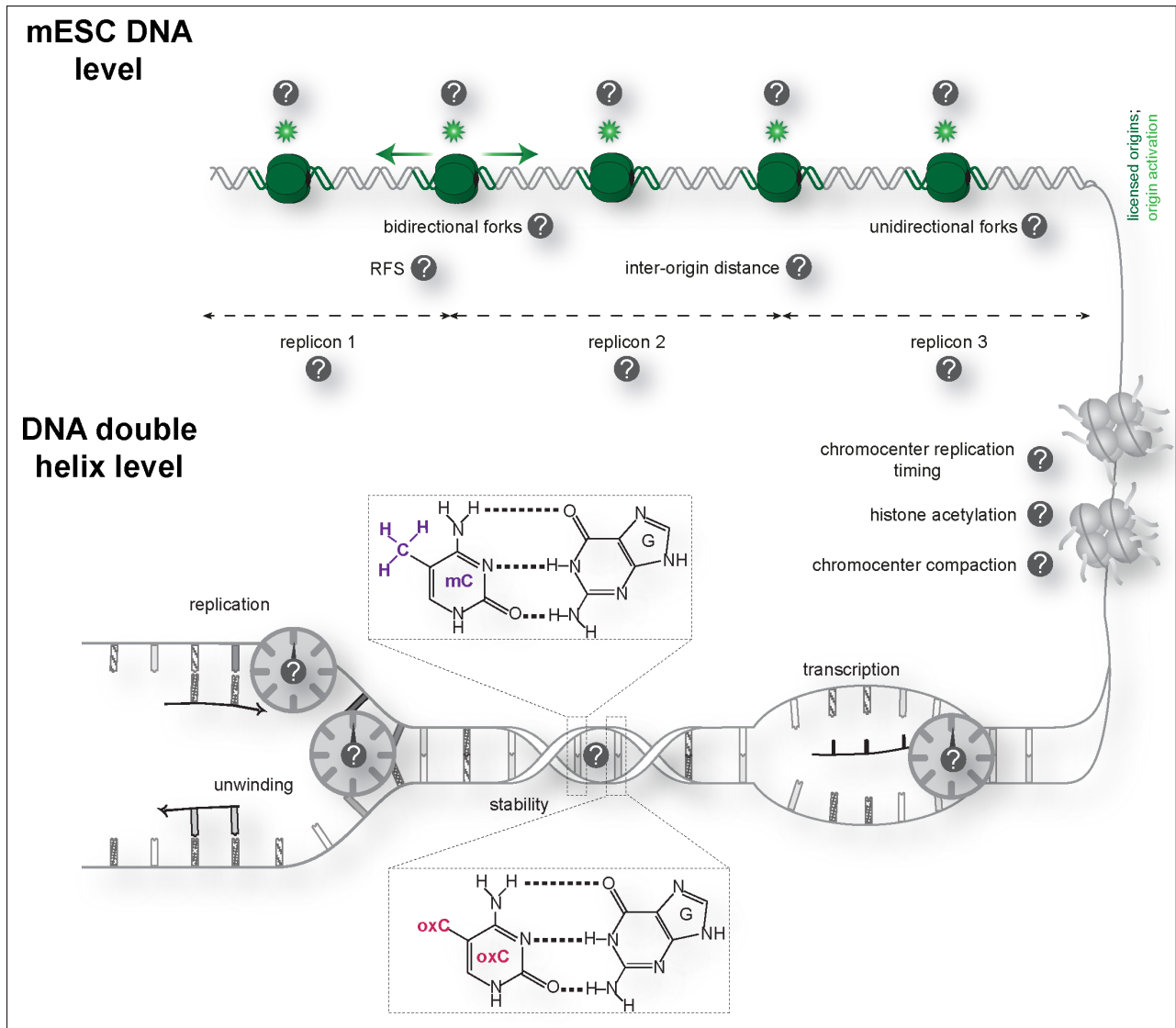


**Figure 4.2:** Scheme summarizing the aims of chapters II and III.

In **chapter IV** (Cytosine Base Modifications Regulate DNA Duplex Stability and Metabolism) we analyzed:

- how cytosine base modification levels can be manipulated in cells,
- DNA double helix stability in the presence of cytosine base modifications,

- DNA transcription, replication and unwinding processivity rates in the presence of cytosine modifications *in vitro* and *in vivo*,
- cell-cycle and S-phase distributions, spatio-temporal S-phase progression, replication fork speed and histone modifications in mES cells depleted of all cytosine modifications.



**Figure 4.3:** Scheme summarizing the aims of chapters III and IV.

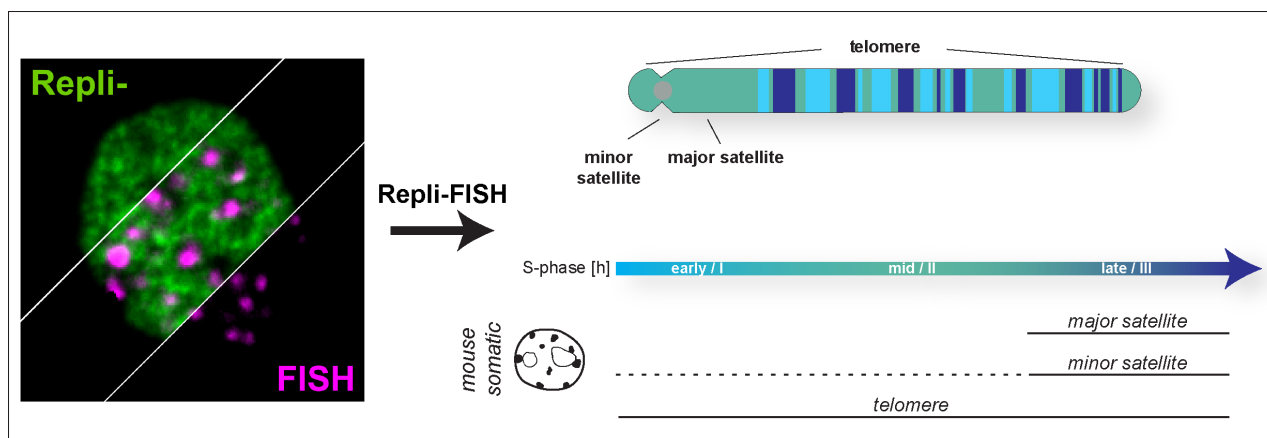


## 5 Result Chapters

### 5.1 CHAPTER I: Repli-FISH – Application of 3D-(Immuno)-FISH for the Study of DNA Replication Timing of Genetic Repeat Elements

C. Rausch prepared all figures and tables and wrote the manuscript together with P. Weber and M. C. Cardoso.  
C. Rausch performed all experiments except for figure 4A, 4B, 5A left, 5B left, 6, 8 and S4 and analyzed all experiments together with P. Weber.

#### Graphical Abstract





Open Access

OBM Genetics



Technical Note

## Repli-FISH (Fluorescence *in Situ* Hybridization): Application of 3D-(Immuno)-FISH for the Study of DNA Replication Timing of Genetic Repeat Elements

Patrick Weber <sup>†</sup>, Cathia Rausch <sup>†</sup>, Annina Scholl, M. Cristina Cardoso <sup>\*</sup>

Cell Biology and Epigenetics, Department of Biology, Technische Universität Darmstadt, 64287 Darmstadt, Germany; E-Mails: mail.pweber@gmail.com, cathia.rausch@gmail.com, annina.scholl@gmail.com, cardoso@bio.tu-darmstadt.de

<sup>†</sup> These authors contributed equally to this work.

<sup>\*</sup> **Correspondence:** M. Cristina Cardoso; E-mail: cardoso@bio.tu-darmstadt.de

**Academic Editor:** Thomas Liehr

**Special Issue:** [Applications of Fluorescence \*in Situ\* Hybridization](#)

*OBM Genetics*

2019, volume 3, issue 1

doi:10.21926/obm.genet.1901062

**Received:** July 22, 2018

**Accepted:** January 8, 2019

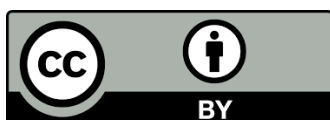
**Published:** January 25, 2019

### Abstract

**Background:** Genetic repeat elements (interspersed or tandem repeats) have diverse functions within cells and at different phases of the cell cycle. However, their investigation at a genome-wide scale is challenging due to their repetitive nature. Here, we describe a method to study the DNA replication kinetics of different repeat elements in single cells throughout the S-phase of the cell cycle.

**Methods:** Mouse major satellite, minor satellite and telomere repeat elements as well as human LINE-1 and Alu repeats were detected in fluorescence *in situ* hybridization (FISH) assays using specific hybridization probes containing biotinylated, digoxigenin-conjugated or fluorescently-labeled nucleotides. We combined their visualization with the detection of cellular DNA replication signals at the single cell level.

**Results:** Specific hybridization of all probes to their intended target sequence and their location on individual chromosomes were verified by microscopic analysis. DNA replication



© 2019 by the author. This is an open access article distributed under the conditions of the [Creative Commons by Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium or format, provided the original work is correctly cited.

*OBM Genetics* 2019; 3(1), doi:10.21926/obm.genet.1901062

was detected by labeling replisome components or nascent DNA synthesis by detecting incorporated thymidine analogs. The different spatial distribution of replication signals allowed us to classify the cells in different S-phase substages. We also measured changes in genomic DNA content during the progression through S-phase. Using high-throughput/high-content microscopic analysis, we then performed colocalization analysis of the degree of overlap between DNA replication and DNA probe signals to determine the kinetics of DNA replication of repeat elements in individual cells.

**Conclusions:** Application of these probes in 3-dimensionally preserved interphase nuclei with the simultaneous detection of newly synthesized DNA or components of the replication machinery allows determination of the replication kinetics of all elements during the synthesis phase of the cell cycle in human cells as well as in embryonic and somatic mouse cells.

### Keywords

3D-FISH; Alu; DNA replication timing; EdU; LINE-1; PCNA; genetic repeat elements; satellite DNA; single cell analysis; telomeres

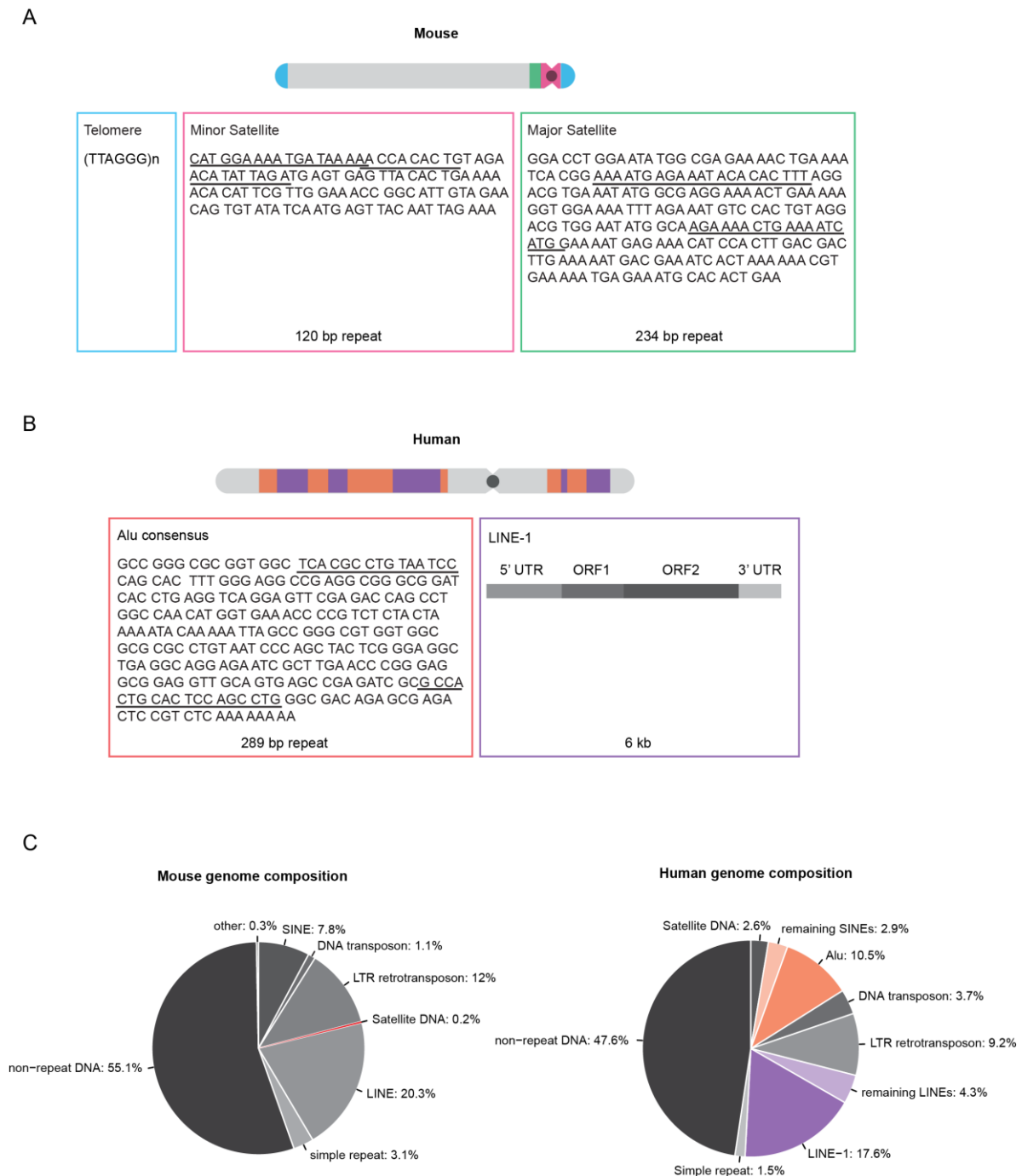
## 1. Introduction

DNA replication is a fundamental cellular process that, like DNA transcription and repair, requires tight control. Dysregulation of this process results in the introduction of mutations or chromosomal aberrations that lead to multiple disorders, including cancer. Thus, elucidation of the molecular mechanism responsible for the control of DNA replication dynamics is of great importance.

Many studies have shown that different chromatin types are replicated at distinct times during S-phase (reviewed in [1]). This is reflected by the changing spatial nuclear distribution of replication signals (visualized by thymidine analog incorporation or labeled replication machinery proteins) observed by light microscopy and termed replication patterns. Each pattern has been shown to correspond to the duplication of a different chromatin type, i.e., euchromatin, facultative and constitutive heterochromatin.

Recently established genome-wide replication timing protocols confirmed that the different chromatin types initiate DNA replication within specific zones (replication domains) along the genome [2]. It was further shown that a substantial proportion of the genome undergoes large-scale chromatin reorganization during embryonic development [2].

Repetitive DNA elements account for more than half of the human genome and up to 40% of the mouse genome (Figure 1C), whereas protein-coding sequences constitute only 1.2% and 1.4% of the human and mouse genomes, respectively [3, 4]. In the human genome, repetitive and transposable genetic elements include, among others, LINEs (long interspersed nuclear elements) and SINEs (short interspersed nuclear elements). Both elements rely on reverse transcription of an RNA intermediate for integration into new genomic locations [5].



**Figure 1** Overview of the localization and sequence of mouse and human repetitive elements. A: Schematic overview of an acrocentric mouse chromosome showing the localization of telomeric, minor and major satellite DNA. B: Schematic overview of a human chromosome showing the localization of Alu and LINE-1 elements. The sequence information for telomeres, minor and major satellite and Alu elements depict the consensus sequence of one repeat unit. Underlined sequence regions represent primer binding sites for PCR amplification of FISH probes. Full-length LINE-1 elements are 6 kb in length and consist of a 5' and a 3' untranslated region (UTR) and two open reading frames (ORF1 and ORF2). C: Pie charts showing the mouse (left) and human (right) genome composition (Repeat Masker). Approximately half of the genomes are composed of different classes of repetitive DNA elements.

*OBM Genetics* 2019; 3(1), doi:10.21926/obm.genet.1901062

LINE-1 elements represent a family of autonomous and active retroelements that account for approximately 17% of the human genome [3]. These elements are found in gene-poor and AT-rich regions, which correspond to dark G-bands on Giemsa-stained metaphase chromosomes [6]. The majority of LINE-1 elements are retrotransposition defective due to 5' truncations, rearrangements or mutations, resulting in 80-100 retrotransposition active LINE-1 elements out of the half a million LINE-1 copies in the human genome [3, 7]. Full-length LINE-1 elements have a total length of 6 kb and contain a 5' untranslated region (UTR), three open reading frames (ORF) and a 3' UTR ending in a AATAAA polyadenylation signal. ORF1 and ORF2 encode the p40 protein and a protein with endonuclease and reverse transcriptase activity that are necessary for retrotransposition [8, 9].

Alu elements are part of the SINE family and represent the most abundant class of non-autonomous retrotransposons. Alu elements cover approximately 11% of the human genome and are predominantly found in gene-rich regions, which, in contrast to LINE-1 elements, correspond to dark R-bands on metaphase chromosomes stained by reverse-Giemsa banding [3, 6]. Alu elements are approximately 290 bp long and have a dimeric structure, consisting of two highly similar left and right monomers separated by a linker sequence. The Alu element sequence does not encode proteins, although the left monomer contains a promoter for RNA polymerase III dependent transcription [3, 10]. In trans retrotransposition is mediated by LINE-1-encoded proteins [11, 12].

In mouse and human cells, telomeres are composed of non-coding, double-stranded TTAGGG array repeats bound by the shelterin complex [13, 14]. The ends of linear chromosomes mirror DNA damage breaks; however, their repair would lead to chromosome fusions. Shelterin prevents telomere recognition, DNA damage responses and chromosome end processing via repair pathways. Similar to telomerase, shelterin also serves to regulate telomere length [14].

In the mouse genome, centromeric and pericentromeric DNA is composed of highly conserved, tandem repeats of DNA consisting of multiple copies of the repeat unit organized in a head-to-tail manner and referred to as minor satellites (MiSat) and major satellites (MaSat) according to their size. MiSats occupy approximately 600 kb of the terminal centromeric regions of all acrocentric mouse chromosomes and are composed of 120-bp AT-rich repeat units [15]. The pericentric MaSat DNA spans 6 Mb adjacent to the MiSat and consists of 234-bp repeats that form clusters in mouse cells, visible as bright structures following DAPI counterstaining and referred to as chromocenters [16].

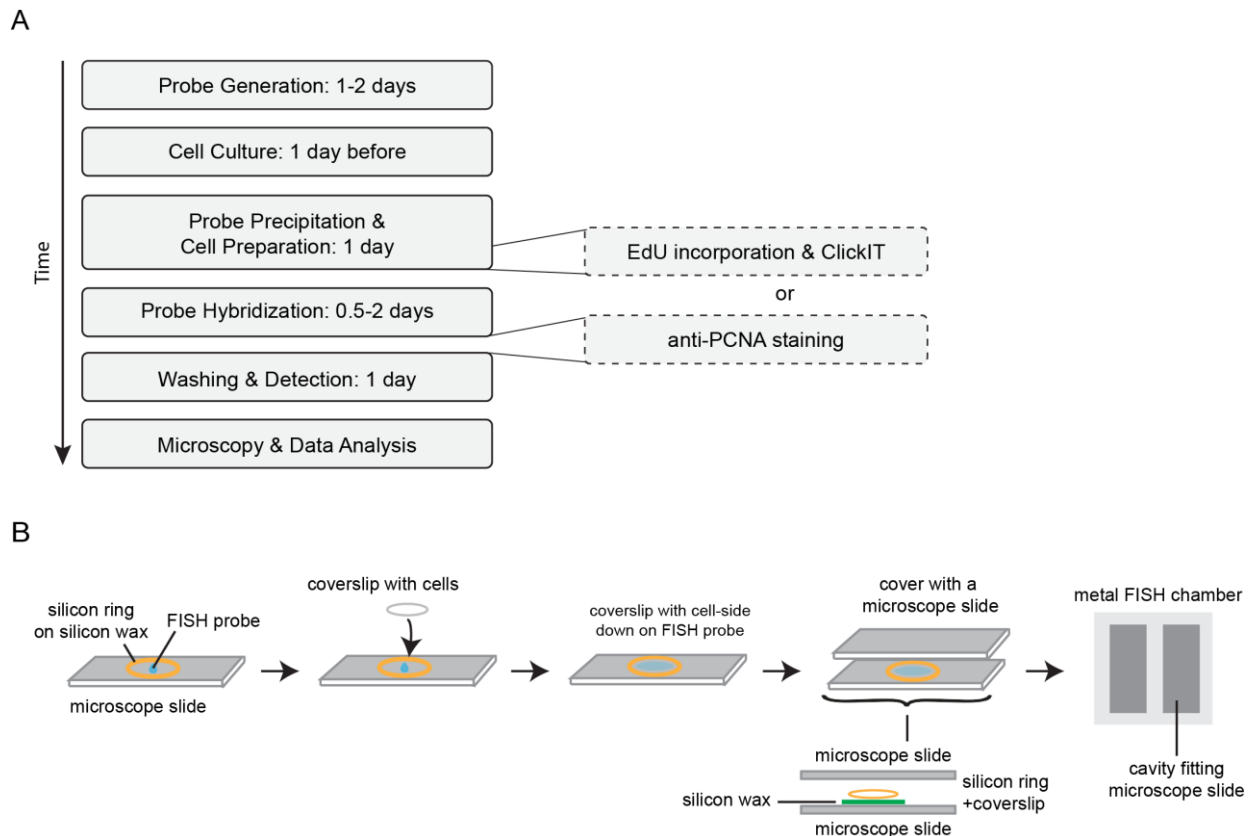
Although historically referred to as “junk”, “parasitic” or “selfish” DNA due to the lack of a clear function within the genome, repetitive elements are now believed to play a role in genome organization during the cell cycle and throughout development as well as in disease conditions [17-22]. Complete and error-free DNA duplication prior to mitosis is an absolute requirement for the preservation of genome integrity and ultimately, the development and maintenance of an organism. This applies not only to the protein-encoding part of the genome, but also holds true for the repetitive elements that are interspersed throughout the genome [23]. However, due to their repetitive nature, repetitive elements are difficult to investigate in genome-wide studies and, hence, the spatio-temporal organization of the replication of repetitive DNA elements remains to be fully elucidated.

Here, we present a protocol that combines fluorescence *in situ* hybridization (FISH) using repeat-specific probes with the detection of DNA replication signals (Repli-FISH) and DNA content.

This method can be applied to 3-dimensionally preserved cells and allows high-throughput microscopic analysis at the single cell level.

## 2. Materials and Methods

A timeline of a typical Repli-FISH experiment is outlined in Figure 2A.



**Figure 2** Representative timeline and setup of FISH experiments. A: Outline of the timelines of a typical replication FISH (Repli-FISH) experiment, showing that the combination of replication visualization and FISH can be performed over the course of four days. B: Experimental setup of the FISH denaturation and hybridization. A silicon wax mark is applied on a microscope slide and the coverslip is placed cell-side down on the probe; a silicon ring is used for additional sealing. A second microscope slide is then placed on top as a cover. For denaturation and hybridization, the entire apparatus is placed in a tightly sealed metal chamber.

### 2.1 Cell Culture

Human HeLa cervical carcinoma cells ([24], ATCC CCL-2) and mouse C2C12 myoblast cells ([25], ATCC CRL-1772) as well as mouse embryonic fibroblast cells (MEF W8, [26]) were cultured at 37°C under 5% CO<sub>2</sub> in high glucose Dulbecco's modified Eagle's medium (DMEM; Cat. No.: D6429, Sigma–Aldrich Chemie GmbH, Steinheim, Germany) supplemented with 1 μM gentamicin (Cat. No.: G1397, Sigma–Aldrich Chemie GmbH), 20 mM L-glutamine (Cat. No.: G7513, Sigma–Aldrich Chemie GmbH) and fetal calf serum [FCS; 10% (HeLa and MEF W8) or 20% (C2C12)]. Cells were passaged every two days.

*OBM Genetics* 2019; 3(1), doi:10.21926/obm.genet.1901062

Murine embryonic stem cells (mESCs, J1, [27], ATCC SCRC-1010) were cultured in high glucose DMEM (Cat. No.: D6429, Sigma–Aldrich Chemie GmbH) containing 16% FCS, 1x non-essential amino acids (Cat. No.: M7145, Sigma–Aldrich Chemie GmbH), 1x penicillin/streptomycin (Cat. No.: P4333, Sigma–Aldrich Chemie GmbH), 1x L-glutamine (Cat. No.: G7513, Sigma–Aldrich Chemie GmbH), 0.1 mM beta-mercaptoethanol (Cat. No.: 4227, Carl Roth, Karlsruhe, Germany), PD 0325901 (0.1  $\mu$ M; Cat. No.: Axon 1408, Axon Medchem BV, Groningen, The Netherlands), CHIR 99021 (0.3  $\mu$ M; Cat. No.: Axon 1386, Axon Medchem BV) and LIF (1,000 U/ml; Cat. No.: ALX-201-242, Enzo Life Sciences GmbH, Lörrach, Germany). Cell culture medium was changed every day and cells were split every two days at a ratio of 1:8 – 1:10.

For Repli-FISH experiments, cells were seeded on gelatin-coated glass coverslips (0.2% gelatin; Cat. No.: G2500, Sigma–Aldrich Chemie GmbH).

## **2.2 Isolation of Genomic DNA**

For the preparation of genomic DNA (gDNA), HeLa or C2C12 cells were harvested by centrifugation (10 min, 400 xg, 4°C) and incubated overnight at 50°C in TNES buffer (10 mM Tris [pH 7.5], 400 mM NaCl, 10 mM EDTA, 0.6% SDS and 1 mg/ml proteinase K (Carl Roth)). Subsequently, 0.6 mg/ml RNase A (Qiagen, Hilden, Germany) was added and DNA was incubated for 30 min at 37°C to remove RNA. NaCl (6 M) was added to the extracted gDNA at a final concentration of 1.25 M. After centrifugation (15 min, 16,000 xg, room temperature), gDNA was precipitated from the supernatant by the addition of ice-cold 100% ethanol and incubation at -20°C for 1 h. The gDNA was pelleted by centrifugation (10 min, 16,000 xg, 4°C), washed with 70% ethanol, dried at room temperature and dissolved in ddH<sub>2</sub>O.

## **2.3 Bio-16-dUTP Nucleotide Generation**

Bio-16-dUTP labeling reactions contained 4.5 mM aminoallyl-dUTPs (5-(-3-aminoallyl)-2'-deoxyuridine 5'-triphosphate, Cat. No: A 0410, Sigma–Aldrich Chemie GmbH), 44.5 mM carbonate-bicarbonate buffer (0.2 M NaHCO<sub>3</sub> stock with 0.5 M NaCl titrated to pH 8.3 with 0.2 M Na<sub>2</sub>CO<sub>3</sub>) with 0.5 M NaCl and 8.9 mM biotin-XX (6-((6-((biotinoyl)amino)hexanoyl)amino)hexanoic acid, Cat. No: B-1606, Molecular Probes) in a total volume of 45  $\mu$ l. Reactions were incubated for 3–4 hours at room temperature and stopped by the addition of 20 mM glycine (pH 8.0). Subsequently, 20 mM Tris-HCl (pH 7.75) was added to stabilize the nucleotides [28].

## 2.4 Probe Generation

Details of primers used for all PCR-based labeling reactions (except for the LINE-1 probe, which was generated by nick-translation) are shown in Table 1.

**Table 1** Primers used for FISH probe generation.

Primer name	Sequence [5'–3']	Reference
Alu-F	GGATTACAGGYRTGAGCCA*	[29]
Alu-R	RCCAYTGCACTCCAGCCTG*	[29]
MaSat-F	AAAATGAGAAACATCCACTTG	[30]
MaSat-R	CCATGATTTTCAGTTTTCTT	[30]
MiSat-F	CATGGAAAATGATAAAAACC	[31]
MiSat-R	CATCTAATATGTTCTACAGTGTG	[31]
Telo-F	TTAGGGTTAGGGTTAGGGTTAGGG	[32]
Telo-R	CCCTAACCTAACCTAACCTAACCTAA	[32]

\*With Y being either pyrimidine (C or T) and R being either purine (G or A)

*OBM Genetics* 2019; 3(1), doi:10.21926/obm.genet.1901062

### 2.4.1 Alu-Specific Probe

Alu-specific probes were generated via a two-step PCR. First, Alu elements were amplified from HeLa cell genomic DNA using partially degenerate Alu-specific primers (Table 1 and Figure 1A), followed by a PCR-based labeling with bio-16-dUTPs (Tables 2 and Tables 3). The product of the labeling PCR was subsequently purified with the QIAQuick PCR purification kit (Qiagen) according to the manufacturer's instructions.

**Table 2** Reaction setup for Alu-specific probe PCR.

PCR component	Template PCR (final concentration)	Labeling (final concentration)
PCR buffer	1x Taq buffer*	1x Taq buffer*
dATP/dGTP/dCTP	0.2 mM each**	0.2 mM each
dTTP	0.2 mM**	0.15 mM
Bio-16-dUTPs	-	0.05 mM
Primer F/R	1 $\mu$ M each	1 $\mu$ M each
Polymerase	1–2 U Taq	1–2 U Taq
DNA template	Human gDNA (150 ng)	1 $\mu$ l 1:50 PCR product
Final volume	To 50 $\mu$ l with ddH <sub>2</sub> O	To 50 $\mu$ l with ddH <sub>2</sub> O

\*10x Taq buffer: 100 mM Tris/HCl pH 8.3, 500 mM KCl, 15 mM MgCl<sub>2</sub>.

\*\* A dNTP mix containing all four nucleotides was used.

**Table 3** PCR cycling conditions for Alu-specific probe amplification.

Cycle step	Temperature	Time	Cycles
Initial denaturation	94°C	4 min	1
1. Denaturation	94°C	1 min	
2. Annealing	57°C/65°C*	1 min	35
3. Extension	72°C	3 min 30 s	
Final extension	72°C	10 min	1

\*Annealing at 57°C for the first amplification of Alu elements from gDNA (template PCR)/annealing at 65°C for the subsequent labeling PCR.

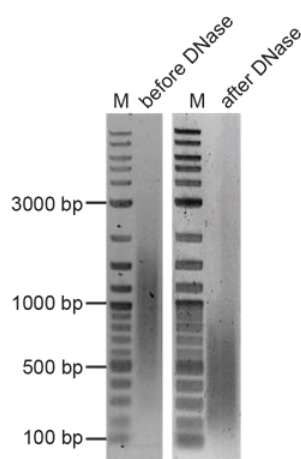
### 2.4.2 LINE-1-Specific Probe

LINE-1 probes were generated and labeled with bio-16-dUTPs by nick-translation of the pLRE3-eGFP plasmid containing the LRE wild-type LINE-1 sequence ([33], a kind gift from John V. Moran, University of Michigan Medical School, USA). Reactions contained 50 mM Tris-HCl pH 8, 5 mM MgCl<sub>2</sub>, 0.5 mg/ml BSA, 10 mM beta-mercaptoethanol, 0.04 mM biotin-16-dUTPs or digoxigenin-11-dUTPs, 0.05 mM dATP/dGTP/dCTP each, 0.32 U DNase I (D5025, Sigma–Aldrich Chemie GmbH), 10 U Klenow fragment (M0210, NEB, Ipswich, MA, USA), 1 µg pLRE3-eGFP plasmid DNA in a total volume of 100 µL. Reactions were incubated for 90 min at 15°C and stopped by adding 5 µl 0.5 M EDTA.

Alu- and LINE-1-specific probes were sheared with a Covaris S220 device (Covaris Inc., Woburn, MA, USA) to a final size of approximately 250 bp in microTUBES (520045, Covaris Inc.). Quality and size were evaluated by 1% agarose gel electrophoresis.

### 2.4.3 Major and Minor Satellite Repeat-Specific Probes

Both major and minor satellite repeat-specific probes were generated via PCR amplification using repeat-specific primers (Table 1 and Figure 1) and mouse genomic DNA from C2C12 mouse myoblasts as the template (Tables 4 and Tables 5). Since this amplification results in probes of varying lengths (detectable as a smear following agarose gel electrophoresis, see Figure 3), samples need to be digested using DNase I (1:250 dilution in ddH<sub>2</sub>O (stock: 2U/µl), 2 µl/reaction) for a maximum of 30 min at room temperature. The reaction was stopped by the addition of 1 µl 0.5 M EDTA and probe quality and size were evaluated by 1% agarose gel electrophoresis (15 µl of reaction mixture).



**Figure 3** Assessment of FISH probe quality. Agarose gel showing probes before and after shearing. Amplification of repetitive DNA elements via PCR or nick-translation leads to amplicons of different sizes (left); digestion with DNase I or mechanical shearing decreases the length of the probes (right). M = marker.

**Table 4** Reaction mixture for major and minor satellite-specific probe PCR amplification.

PCR component	PCR (final concentration)
PCR buffer	1x Taq buffer*
dATP/dGTP/dCTP	0.2 mM each
Bio-16-dUTPs/dig-11-dUTPs**	0.08 mM
Primer F/R	0.2 $\mu$ M
Polymerase	1–2 U Taq
DNA template	1 $\mu$ l mouse gDNA
Final volume	To 100 $\mu$ l with ddH <sub>2</sub> O

\* 10x Taq buffer: 100 mM Tris/HCl pH 8.3, 500 mM KCl, 15 mM MgCl<sub>2</sub>.

\*\* Digoxigenin-11-dUTPs alkali-labile (Cat. No: 11573152910, Sigma–Aldrich Chemie GmbH).

**Table 5** PCR cycling conditions for major and minor satellite-specific probe amplification.

Cycle step	Temperature	Time	Cycles
Initial denaturation	95°C	5 min	1
1. Denaturation	95°C	1 min	40
2. Annealing	56°C	1 min	
3. Extension	72°C	2 min	
Final extension	72°C	5 min	1

#### 2.4.4 Telomere Repeat-Specific Probe

Probes specific for telomeric repeats were generated by a two-step PCR using previously described self-annealing primers [32]. The first PCR reaction served to create elongated probe fragments with regular dNTPs, while the second reaction included biotinylated or cyanine dye-conjugated nucleotide analogues for efficient probe labeling (Table 6). For the initial amplification, a high-fidelity polymerase was used (e.g. Q5 polymerase (NEB, Cat#: M0491)). For labeling with bio-16-, Cy3- or Cy5-conjugated dUTPs in the subsequent PCR, however, Taq polymerase was required since the Q5 polymerase does not incorporate dUTPs. Identical cycling conditions were used in both reactions (Table 7).

**Table 6** Reaction mixture for telomere repeat-specific probe PCR amplification.

PCR component	PCR (final concentration)	Labeling (final concentration)
PCR buffer	1x Q5 buffer	1x Taq buffer*
MgCl <sub>2</sub>	0.5 mM	-
dATP/dGTP/dCTP	0.2 mM each**	0.05 mM each
dTTP	0.2 mM**	0.01 mM
Bio-16-dUTPs/Cy3 or Cy5-dUTPs***	-	0.1 mM/0.08 mM
Primer F/R	0.2 μM each	0.2 μM each
Polymerase	1–2 U Q5 DNA polymerase	1–2 U Taq
DNA template	-	1 μl of PCR product
Final volume	To 50 μl with ddH <sub>2</sub> O	To 50 μl with ddH <sub>2</sub> O

\* 10x Taq buffer: 100 mM Tris/HCl pH 8.3, 500 mM KCl, 15 mM MgCl<sub>2</sub>.

\*\* A dNTP mix containing all four nucleotides was used.

\*\*\* Cy3-dUTP and Cy5-dUTP: GE Healthcare (Chicago, IL, USA) PA53022 and PA55022 respectively.

**Table 7** PCR cycling conditions for telomere-specific probe amplification.

Cycle step	Temperature	Time	Cycles
Initial denaturation	95°C	5 min	1
1. Denaturation	98°C	10 s	40
2. Annealing	50°C	15 s	
3. Extension	72°C	20 s	
Final extension	72°C	2 min	1

## 2.5 Probe Precipitation and Fluorescence in Situ Hybridization (FISH)

NOTE: Co-detection of incorporated EdU must be performed prior to FISH (see Figure 2A and refer to Section 2.6.2).

*OBM Genetics* 2019; 3(1), doi:10.21926/obm.genet.1901062

### 2.5.1 Probe Precipitation

Probes were ethanol-precipitated with 0.15 M sodium-acetate and 50 µg/ml fish sperm DNA, washed with 70% ethanol, air-dried and resuspended in hybridization solution containing formamide (70% formamide, 2x SSC (30 mM Na-citrate, 300 mM NaCl), 10% dextran sulfate, pH 7) for MaSat-, MiSat- and telomere-specific probes, and without formamide (10 mM Tris-HCl, 3 mM MgCl<sub>2</sub>, 50 mM KCl, 10 µg/ml gelatin, 2x SSC, [34, 35]) for Alu- and LINE-1-specific probes. Probes were denatured for 5 min at 80°C and immediately placed on ice.

### 2.5.2 Cell Preparation

For subsequent hybridization and detection of replication signals, cells were washed twice with 1x PBS and immediately fixed with 3.7% formaldehyde/PBS for 10–15 min. After washing (2–3x) in PBS-T (0.02% Tween), cells were permeabilized in 0.5% Triton-X100/PBS for 10–15 min and washed again with PBS-T. At this point, cells may be used directly for hybridization, or stored for a few days in 1x PBS at 4°C.

For MaSat-, MiSat- and telomere-specific probe hybridization, cells were incubated for 15 min in 0.1 M HCl, washed 3x with 1x PBS. Cells were then equilibrated for 2 min in 2x SSC and finally for 20 min in 50% formamide/2x SSC. For Alu and LINE-1 FISH, cells were equilibrated in hybridization solution without formamide (see Section 2.5.1). All incubation steps were performed at room temperature.

For metaphase spreads, asynchronously growing cells were treated with colcemid (0.1 µg/ml final concentration) for 1 h, harvested by trypsinization and centrifuged at 500 xg for 5 min. After carefully removing the supernatant, the cell pellet was resuspended in pre-warmed KCl solution (0.075 M at 37°C) and incubated at 37°C for 5 min. To prevent clumping of the cells, the first 1–2 ml were added drop-wise and mixed well by flicking the tube. The rest of the solution was added to give a total volume of approximately 8 ml. After a second centrifugation step, the pellet was fixed by the initial drop-wise addition of freshly prepared fixative solution (methanol:acetic acid, 3:1), which was then filled up to 10 ml. After incubation at 4°C for at least 20 min (overnight incubation is preferred) the samples were again centrifuged and most of the supernatant was removed by decanting. The cell pellet was then finally resuspended in the remaining fixative solution by flicking. The chromosome preparations were made from 15–20 µl drops on microscope slides that were dipped into ddH<sub>2</sub>O.

For the subsequent hybridization, the slides were air-dried overnight, rehydrated for 10 min in ddH<sub>2</sub>O and digested at 37°C for 10 min in 0.005% pepsin (P6887, Sigma–Aldrich Chemie GmbH) in 0.01 M HCl. The slides were then dehydrated in 70% and 100% ethanol (5 min each) and air-dried overnight. Before hybridization, slides were equilibrated for 30 min in 50% formamide/2x SSC.

### 2.5.3 Hybridization

For hybridization of interphase cells, a circle/square the size of the glass coverslip carrying the cells was drawn onto a regular microscope glass slide (Menzel–Gläser Superfrost, Thermo Scientific, Cat. No: ABAA000080##32E) using an ImmunoPen (Calbiochem, Cat. No.: 402176-1EA). The hybridization solution containing the respective probe(s) was then applied to the center of this mark and cells were applied face down onto the hybridization solution. The coverslip was

*OBM Genetics* 2019; 3(1), doi:10.21926/obm.genet.1901062

additionally sealed with a silicon ring/square (self-made) and a second microscope slide was applied on top. The whole apparatus was then transferred to a hybridization chamber and tightly sealed (Figure 2B).

For hybridization of metaphase spreads, the area of the slide containing the spreads was encircled using an ImmunoPen. The respective probe(s) were then applied, covered with a coverslip to ensure even distribution of the probe and further sealed with a silicon square and a second microscope slide.

The chamber was subsequently incubated at 80°C for 5 min in a water bath to denature both the genomic DNA and probe. This was followed by incubation in a second water bath at 37°C for 24 h (MaSat, MiSat and telomeres) or overnight at 42°C (Alu and LINE-1).

#### 2.5.4 Washing and Blocking

After hybridization, cells were carefully removed from the microscope slide and washed three times for 5 min with 50% formamide/2x SSC pH 7 at 45°C followed by two washes for 5 min in 2x SSC at 45°C. For the Alu- and LINE-1-specific probes, the washing steps were performed with 2x SSC and 0.1x SSC without formamide. To block non-specific antibody binding sites, cells were incubated for 30 min in 1% BSA/4x SSC at room temperature.

NOTE: For PCNA detection, immunodetection must be performed before continuing with the probe detection (Section 2.6.1 and Figure 2).

#### 2.5.5 Probe Detection

Probes were detected by incubation with a rabbit anti-digoxigenin antibody (1:500, Cat. No.: 700772, Thermo Fisher Scientific), a fluorescently-labeled anti-rabbit IgG antibody or fluorescently-tagged streptavidin diluted in 1% BSA/4x SSC, for 20 min at room temperature and washed three times for 5 min with 0.05% Tween/4x SSC.

A post-fixation step with 1%–4% formaldehyde/PBS for 10 min further stabilized the resulting antibody complexes.

DNA counterstaining was performed by incubating cells for 10 min at room temperature with DAPI (10 µg/ml, 4',6-diamidin-2-phenylindol) and slides were mounted using Mowiol (Merck KGaA, Cat. No.: 81381-250G).

## 2.6 Replication Staining

### 2.6.1 Staining of the Replisome Component PCNA.

Immunofluorescent detection of PCNA usually requires additional fixation with methanol. This step ensures accessibility of the anti-PCNA antibody to the epitope, which is located at the interface between the monomeric subunits of the homotrimeric PCNA ring-complex (Figure S1). However, we noticed that this step could be omitted when the hybridization procedure has been performed before.

To achieve good signals, antibody detection of PCNA must be performed after hybridization and washing but before detection of the probe (Figure 2A). After blocking as described in Section 2.5.4, cells were incubated for 1–2 h at room temperature with an antibody specific for PCNA (mouse anti-PCNA, mAb, DABCO, Cat#: M0879, 1/200) diluted in blocking buffer (1x PBS/0.02%

*OBM Genetics* 2019; 3(1), doi:10.21926/obm.genet.1901062

Tween20/2% BSA). After washing (2–3x) in PBS-T (0.02% Tween20), detection with secondary antibodies conjugated with fluorophores suitable for subsequent microscopic detection was performed, before continuing with the probe detection (see Section 2.5.5).

### 2.6.2 EdU Labeling and Visualization of Replicated DNA

EdU labeling and detection must be carried out before cells are equilibrated in SSC and hybridization is performed.

For detection and staining of active DNA synthesis sites, cells were incubated for 15 min with 10  $\mu$ M EdU (5-ethynyl-2'-deoxyuridine, Thermo Fisher Scientific, Waltham, MA, USA), washed with 1x PBS and fixed with 3.7% formaldehyde/PBS for 10 min. After washing (3x) with PBS-T (0.02% Tween20), cells were permeabilized with 0.5% Triton-X100/PBS for 15 min and pre-denatured with 0.1 M HCl for 15 min, followed by a second permeabilization step with 0.5% Triton-X100/PBS for 15 min.

EdU was detected using ClickIT chemistry (EdU Click-ROTI kit, Carl Roth, Karlsruhe, Germany) according to the manufacturer's instruction. The 6-FAM and the 3-azido-7-hydroxycoumarin azides were used in final dilutions of 1:1,000 and 1:200, respectively. Cells were post-fixed with 1% formaldehyde/PBS for 10 min before proceeding with equilibration in SSC and hybridization buffer, followed by FISH and probe detection (see sections 2.5.2–2.5.5).

## 2.7 Confocal Microscopy and Line-Profile Colocalization Analysis

Confocal z-stacks of Repli-FISH stained cells were acquired using the UltraVIEW VoX spinning disc microscopy system (PerkinElmer Life Sciences, Waltham, MA, USA) equipped with an oil immersion 60x Plan Apochromat NA 1.45 objective lens.

To measure and compare the overlap between the major satellite-specific FISH probe signals and the replication signals (EdU), line intensity profile analysis was performed in the mid-focal planes of individual nuclei from cells in different stages of the cell cycle (non-S versus early, mid and late S-phase) using ImageJ (<https://imagej.nih.gov/ij/>). The intensities of both signals were measured along a line drawn manually through the nuclei and covering both euchromatic and heterochromatic (chromocenters) portions within the cell nucleus. The intensity values of both channels were subsequently plotted against the distance along the line.

## 2.8 High-Content/High-Throughput Imaging and Analysis

High-content/high-throughput imaging and analysis were performed with the Operetta High-Content Imaging System (Perkin Elmer, Part Number: HH12000000). Multiple fields-of-view (169 in total) were acquired with an immersion free 60x high NA objective (NA 0.9) and automatic image analysis was performed with the associated *Harmony* software (v3.5). The analysis pipeline involved the initial automatic detection and segmentation of individual cell nuclei based on DAPI counterstaining. Cells overlapping the edge of the image or otherwise deformed cell nuclei were excluded from the analysis. Subsequently, segmentation of major satellite repeats within the individual cell nuclei was performed based on the fluorescence signal of the hybridized MaSat-specific probe. Based on this object segmentation, the fluorescence intensity of the replication signal (PCNA in this case) was measured within all spots per individual cell nucleus. This allowed

*OBM Genetics* 2019; 3(1), doi:10.21926/obm.genet.1901062

the comparison of overlapping signals at the different S-phase substages with each other and also with non-S-phase cells (G1- and G2-phase) from the whole population. For this, the cells were sorted according to the different cell cycle stages based on the PCNA patterns in ImageJ using a custom written macro (available upon request). In addition, the DAPI intensity of each nucleus was measured to compare DNA content within the different subpopulations. Boxplots were used to plot (1) the increase in DNA content over the cell cycle (non-S-phase, early and late S-phase) and (2) the accumulation of PCNA at chromocenters in cells at the different cell cycle stages using RStudio (Version 0.99.893).

Harmony analysis pipeline:

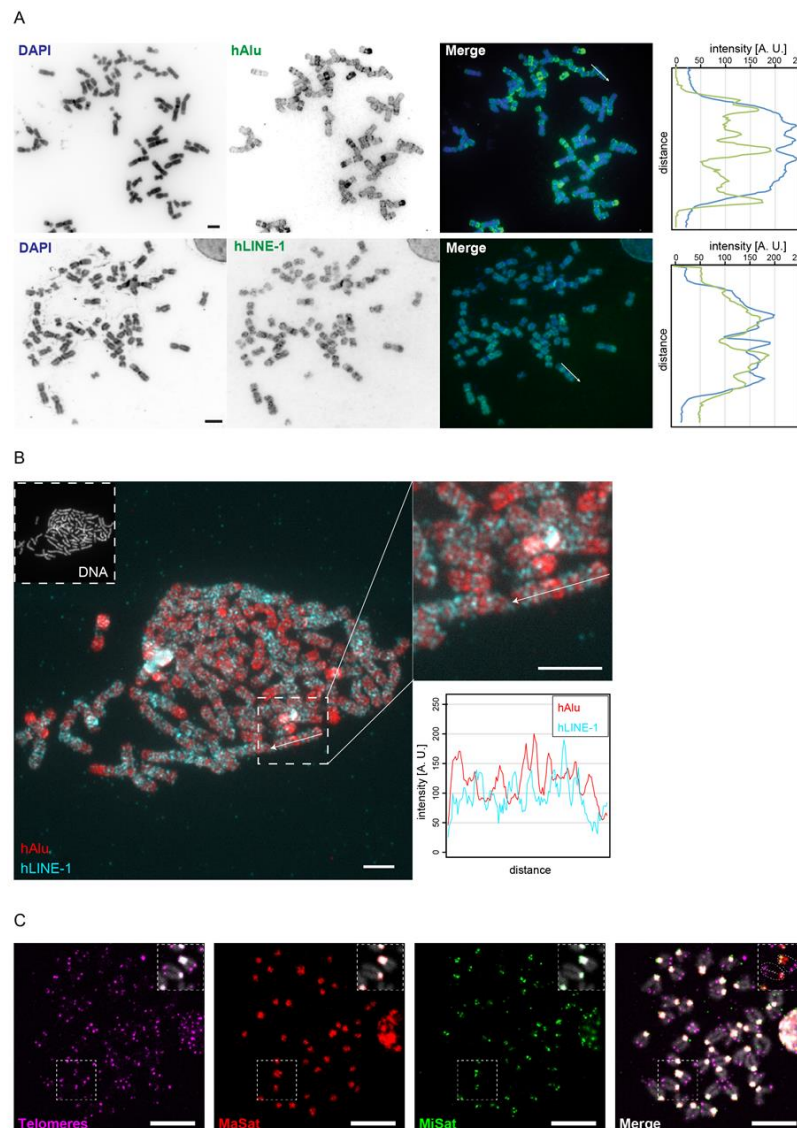
- 1) Input image:
  - Stack processing: Max. projection
  - Flatfield correction: Basic
- 2) Find Nuclei (2):
  - Channel: DAPI
  - Method A
 => Output population: Nuclei
- 3) Find spots:
  - Channel: Alexa 488
  - Population: Nuclei
  - Region: Nucleus
  - Method: A
 => Output population: Alexa488 spots
- 4) Calculate morphological properties:
  - Population: Nuclei
  - Region: Nucleus
  - Method: Standard
- 5) Select population (2):
  - Population: Nuclei
  - Method: Common filters
  - Remove border objects
 => Output population: Nuclei inside [field-of-view]
- 6) Select population:
  - Population: Nuclei inside
  - Method: Filter by property
    - Properties roundness: 0.8
    - Properties area [ $\mu\text{m}^2$ ] <300
    - Properties area [ $\mu\text{m}^2$ ] >60
 => Output population: Round nuclei inside (exclude deformed nuclei)
- 7) Calculate intensity properties (in whole nucleus and in spots):
  - DAPI, MaSat, PCNA

### 3. Results

Repetitive DNA elements are difficult to investigate using genomic approaches due to their repetitive nature, which hinders their mapping to reference genomes. Here, we describe the use of probes for five different genetic repeat elements in different species (mouse and human) and at different developmental stages for the study of their DNA replication kinetics. The probes presented here are commonly used in our laboratory and were generated via standard procedures (PCR and nick-labeling) as described in Section 2 (*Material and Methods*). However, we also emphasize the use of other probes that are suitable for the detection of specific DNA sequences and that have been proven useful in cytogenetic studies, e.g. peptide nucleic acids (PNAs; [36]), probes generated via *primed in situ synthesis* (PRINS; [37, 38]) or from bacterial artificial chromosomes (BACs).

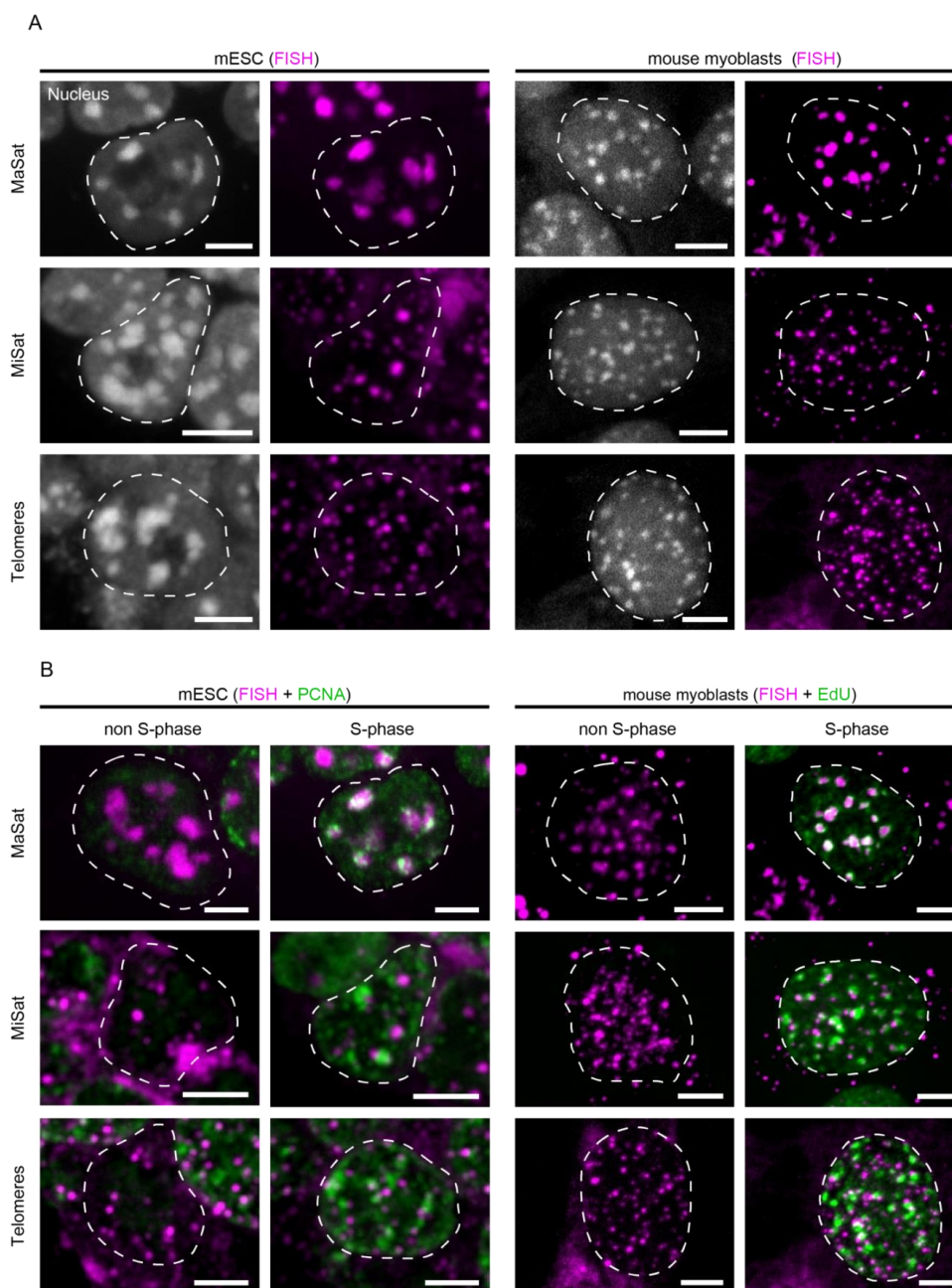
In order to confirm their specificity, we performed metaphase FISH with all DNA probes. Figure 4A shows the results for human Alu- and LINE-1-specific probes and their distribution along DAPI-counterstained chromosomes as measured by line-profile colocalization analysis (Figure 4A). Alu elements are abundant in GC-rich regions, whereas LINE-1 elements are predominantly found in AT-rich (GC-poor) regions. We found the distribution of Alu elements correlated negatively with the DAPI staining, which preferentially binds AT-rich regions of the DNA, whereas LINE-1 and DAPI signals correlated positively. Furthermore, double hybridization of mitotic chromosomes with biotin-labeled Alu- and digoxigenin-labeled LINE-1-specific probes showed mutually exclusive binding of the two elements (Figure 4A & Figure 4B), emphasizing their existence in distinct regions of the genome (Figure 1B).

The localization and specificity of the mouse probes against MaSat, MiSat and telomeres were evaluated by triple-color DNA FISH on chromosome preparations. For this, the probes were labeled via PCR using dUTPs conjugated to either digoxigenin-11, biotin-16- or Cy5. The results are summarized in Figure 4C. The telomere probe was found to hybridize specifically to the ends of chromatids, resulting in four signals per chromosome. Major satellite probes displayed only one large signal on each chromosome overlapping with DAPI-intense pericentromeric segments. In contrast, the minor satellite probe hybridization resulted in two distinct spots per chromosome; one per chromatid. In accordance with their chromosomal distribution, the centromeric (MiSat) and pericentromeric (MaSat) DNA signals were found in close proximity. Due to the acrocentric organization of mouse chromosomes (Figure 1A), a similar distribution was observed for proximal telomeres and minor satellite DNA. The clustering of pericentromeric and centromeric heterochromatin as well as the organization of chromosomal ends was also visualized by the simultaneous hybridization of all probes to interphase cells as shown in Figure S2. In summary, these data clearly confirmed the specificity of each probe for its intended target structure.



**Figure 4** Validation of FISH probe specificity by hybridization to metaphase chromosome preparations. **A:** Whole metaphase spreads from HeLa cells were hybridized with human Alu-specific (upper row) and human LINE-1-specific (lower row) FISH probes. Chromosomes were counterstained with DAPI. Color line profiles represent the expected overlap between Alu with GC-rich (gene-rich) and LINE-1 with AT-rich (gene-poor) regions, which are differentially stained by DAPI. Scale bar: 5  $\mu$ m. **B:** A metaphase spread hybridized to both human Alu-specific (red) and human LINE-1-specific (cyan) probes simultaneously reflects their widespread distribution along eu- and heterochromatic DNA segments (left). The two probes hybridize to mutually exclusive segments along chromosomes as represented by line-profile analysis of a single representative chromosome (right). Scale bar: 5  $\mu$ m. **C:** Triple-color FISH combining major satellite-, minor satellite- and telomere-specific probes, each labeled with a different nucleotide analog and hybridized to metaphase chromosome preparations from diploid mouse cells. All probes bind specifically to the chromosomal regions in which the target sequences are organized (see also enlarged insets in each image), i.e. chromosome ends (telomeres, magenta), DAPI-dense pericentromeric regions (MaSat, red) and centromeric regions (MiSat, green). Scale bar: 10  $\mu$ m.

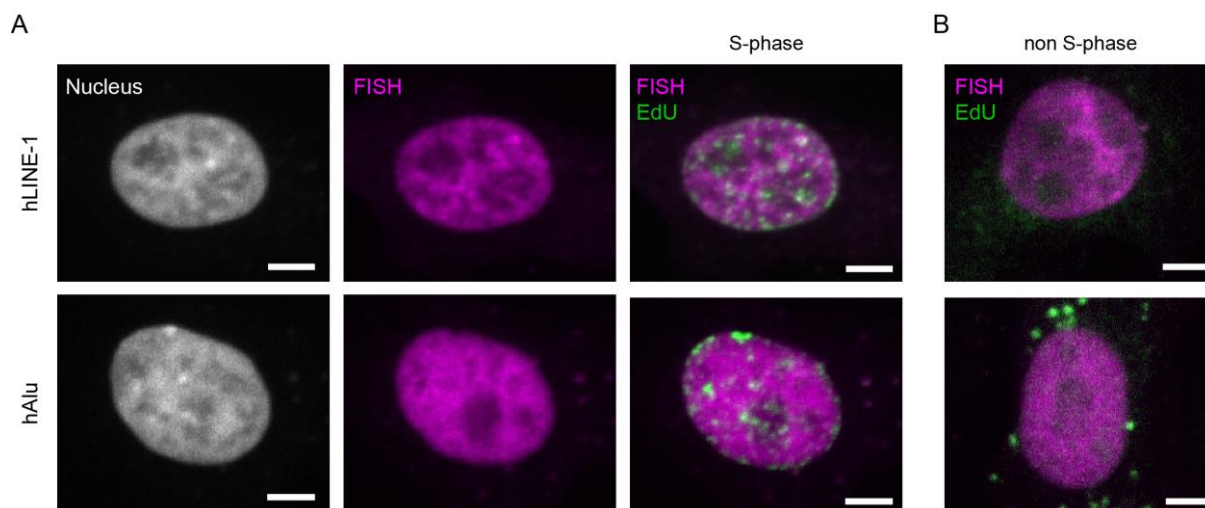
In both mouse ESC and myoblast interphase nuclei, major satellite DNA was detected as large spots co-localizing with DAPI-intense regions (Figure 5A, upper row). In contrast, minor satellite signals were found in close proximity to the major satellite clusters, forming several distinct spots at their periphery, as expected from their physical location on individual chromosomes (Figure 5A, middle row and Figure 1A). Telomere-specific FISH signals were detected throughout the nucleus, with some localizing adjacent to pericentric heterochromatin (Figure 5A, lower row) due to the acrocentric nature of mouse chromosomes (see Figure 1A).



**Figure 5** Overview of the localization of MaSat, MiSat and telomeres in murine interphase cells. **A:** MaSat-, MiSat- and telomere-specific FISH was performed on interphase cells of mouse embryonic stem cells (mESC, left) and mouse myoblasts (right). DNA was visualized by DAPI counterstaining. **B:** DNA replication was visualized by PCNA co-staining or EdU incorporation and detection. Scale bar: 5  $\mu$ m.

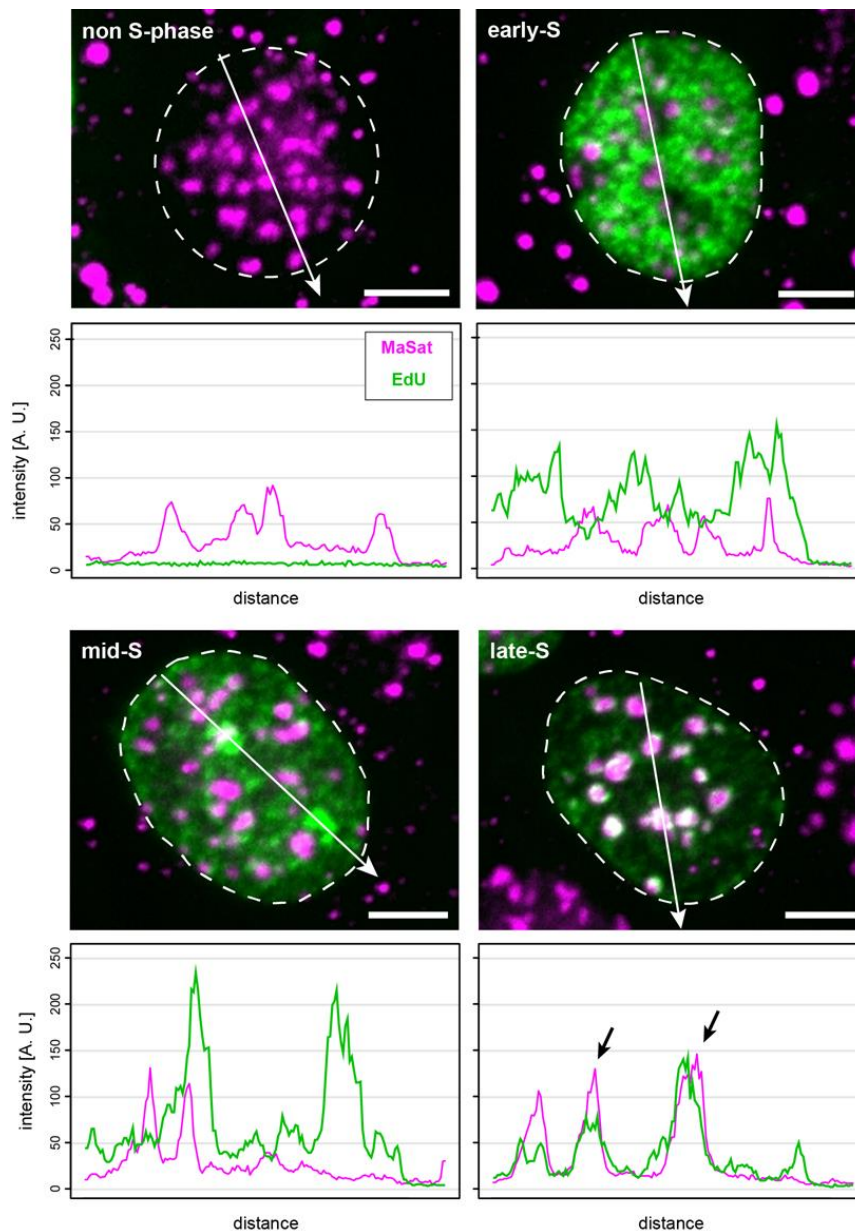
As interphase chromosomes distribute throughout the nuclear volume, signals for both LINE-1 and Alu interspersed repeats were found throughout the nuclei of human cells, albeit excluded from the nucleoli (Figure 6A). This reflects the fact that both classes of elements cover a large proportion of the genome (e.g. approximately 11% for Alu repeats [39]) as represented by the distribution of LINE-1 and Alu elements on metaphase chromosomes (Figure 1B, Figure 1C and Figure 4B). Notably, the correlation of LINE-1 signals with heterochromatin was stronger than the correlation of Alu signals as judged by the distribution of the DNA and replication signals (note that chromocenters do not form in human cells) (Figure 6A and Figure B). A detailed analysis of the distribution of Alu and LINE-1 signals throughout S-phase as well as their replication timing are reported in [40]. In brief, it was shown that Alu elements are replicated during early S-phase, whereas LINE-1 elements are replicated throughout S-phase.

Next, we evaluated the DNA replication kinetics of Alu and LINE-1 elements. We were able to show that both signals, either PCNA or EdU, facilitated discrimination between replicating and non-S-phase cells within an asynchronous cell population (Figure 5B, Figure 6A & Figure 6B). While replicating cells are characterized by the dynamic focal accumulation of the respective signal throughout S-phase, cells in either the G1- or G2-phase of the cell cycle can be identified by the lack of such a pattern. Due to the different types of replication markers used in our approach, non-S-phase cells co-stained for PCNA showed a diffuse nuclear signal representing the unbound nuclear pool of the protein (Figure 5B, left panel, [41-44]). In contrast, cells in which nascent DNA is marked by the incorporation of the thymidine analog EdU, were only labeled during S-phase and no signals were detected in the G1 and G2 phases (Figure 5B, right panel). Notably, the nuclear localization of the different repetitive elements did not change within the nucleus of S-phase cells, to which the analysis was restricted as can be seen by comparison of the FISH signal distribution in cells at the different stages of S-phase (Figure 5B). Taken together, our results showed that all probes used in this study were highly specific for their intended target sequence (Figure 4), resulting in clear signals with high signal-to-noise ratio when hybridized to interphase nuclei. Furthermore, both nascent DNA as well as replication factors were co-visualized successfully (Figure 5 & Figure 6).



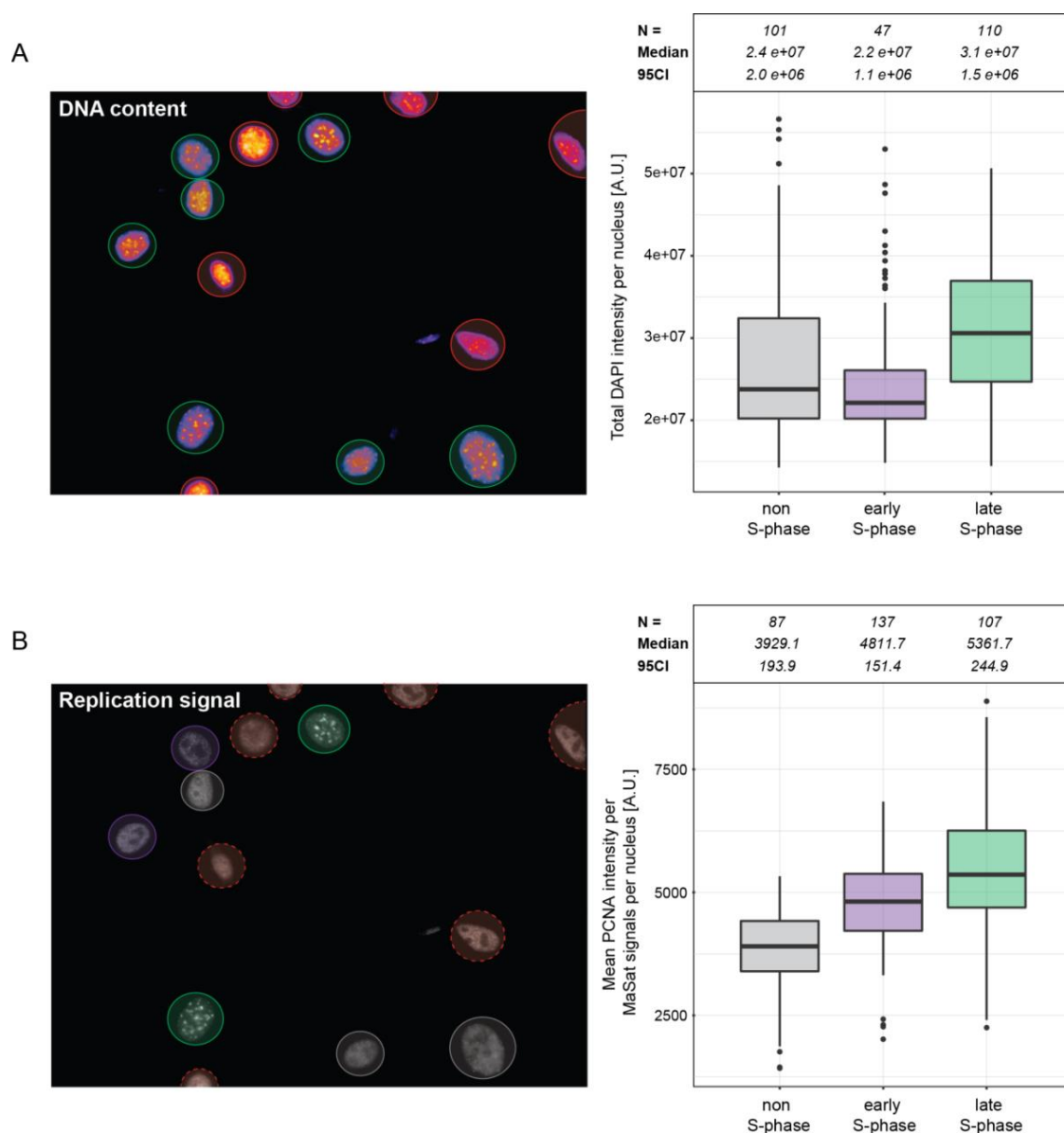
**Figure 6** Overview showing the localization of Alu and LINE-1 elements in human interphase cells. **A:** HeLa interphase cells showing the nuclear distribution of Alu- and LINE-1-specific FISH signals. DNA was visualized by DAPI counterstaining. Alu- and LINE-1-specific FISH signals visualized in combination with EdU in S-phase cells. **B:** The same probes were observed in cells that are not in S-phase for comparison of their respective distribution in the cell nucleus. Scale bar: 5  $\mu$ m.

Colocalization analysis (e.g., color line-profile analysis) of the overlapping FISH and replication signals serve as a measure of ongoing DNA replication during the substages of S-phase and, thus, indicate replication timing. Figure 7 shows the outcome of such representative measurements performed on mouse myoblast cells hybridized with the MaSat-specific probe. Again, non-S-phase and replicating cells can be discriminated by the absence or presence of the EdU signal, respectively. Furthermore, the replication of euchromatin (early S), as well as facultative (mid-S) and constitutive heterochromatin (late S), are represented by the changing EdU patterns. The overlap between FISH and replication signals along the analysis path (white arrow) were strongest at late S-phase, which is in complete accordance with previous reports [45, 46]. Figure S3 shows the results from a similar analysis performed with cells hybridized to either MiSat- or telomere-specific probes. Here, we found minor satellite DNA replicated asynchronously throughout S-phase, indicated by the overlap of the MiSat FISH signal and replication foci (EdU) shown by line-profile analysis and similar to the results published by [47]. Furthermore, telomeres were also found to replicate throughout the course of S-phase, which is in accordance with previous findings [48-50].



**Figure 7** MaSat probe localization in mouse myoblast cells at different S-phase substages. Cells were classified as non-S-phase, early, mid and late replication substage according to their EdU pattern. Colocalization of MaSat FISH (magenta) and replication (green) was analyzed via a color line-profile (black arrows). Scale bar: 5  $\mu$ m.

These results were also validated and extended by high-content/high-throughput imaging and analysis of asynchronous cell populations (Figure 8). Here, suitable cells were automatically identified based on DNA counterstaining with DAPI (Figure 8A, left). The progression through S-phase was reflected not only by the replication signal pattern, but also by the increase in genomic DNA content over time as measured by the increasing DAPI intensity from cells at the different S-phase substages (Figure 8A, right). The broader signal distribution of non-S-phase cells can be explained by the fact that cells from both G1- and G2-phases of the cell cycle were grouped in this fraction. Cells were classified as non-S-phase cells and cells at either early or late S-phase based on the replication signal distribution (Figure 8B, left). The latter showed a similar increase in the replication signal overlap (PCNA) with major satellite repeats at late S-phase (Figure 8B, right), thus, supporting the previous results of individual cell measurements (Figure 7).



**Figure 8** High-content/high-throughput Repli-FISH analysis of asynchronous cell populations. MEF W8 cells hybridized with a major satellite-specific probe and co-stained for PCNA were imaged with a high-content imaging system (Operetta High-content Imaging System, Perkin Elmer, 60x 0.9 NA air objective). **A:** Image analysis was performed with the *Harmony* software. Automatic segmentation of cell nuclei was performed based on DNA counterstaining with DAPI (left, green circles). Cells out of focus or on the edge of the image were excluded from the analysis (left, red circles). Increasing DNA content during S-phase is represented by the increase in total DAPI intensity measured in cells at early or late S-phase compared to non-S-phase cells (right). **B:** Cells were classified by the PCNA signal distribution and sorted into non-S-phase cells (left, gray circles) and cells at either early (left, purple circles) or late S-phase (left, green circles). Comparison of the mean PCNA intensity within all major satellite spots per cell nucleus showed a clear increase toward late S-phase (right, green boxplot), allowing discrimination from the other classes in the population and confirmed the results of individual cell measurements shown in Figure 7.

#### 4. Discussion

Modern next-generation sequencing methods have become increasingly cost and time-efficient. However, available software is still challenged by the correct mapping of repetitive elements both in resequencing and *de novo* assembly projects. This can be partially explained by the read-length that is now achievable (50–150 bp by next-generation sequencing compared to 800–900 bp obtained by Sanger sequencing), which is substantially shorter than the length of most repeats covering the genome. This results in reads that map to multiple locations within the genome (multi-reads) in resequencing experiments or gaps in *de novo* genome assemblies [39].

At the same time, several methods have been developed that combine the isolation of nascent DNA from replicating cells with subsequent deep-sequencing (Repli-seq) of the underlying DNA sequence [51]. This method can be used to study the replication timing from cell populations on a genome-wide scale and further offers the ability to relate the sequence information to underlying DNA and chromatin modifications, providing correlative information on the mechanism by which DNA replication is regulated.

Due to the length of time required to sufficiently label and precipitate hundreds of kilobases of labeled DNA (>60 min), these methods have a rather poor temporal resolution. Because of cell population averaging, the analysis focuses mainly on early versus late replicating cells, thereby masking intermediate replication states. Thus, given the fast underlying kinetics and potential variability of the replication process in individual cells, not all questions can be addressed with such methods. In addition, this method is problematic in the alignment of repetitive elements to a reference sequence as described previously.

Although described specifically for different repetitive elements in mouse and human cell lines, the Repli-FISH method can be easily extended using other probes, such as satellite repeat elements [40], rDNA repeat loci [52] or individual genes, e.g. via labeled bacterial artificial chromosomes (BACs) with sequence complementarity to specific genes [53]. In addition, the hybridization procedure can be combined with the detection of other signals of interest, including specific proteins and histone or DNA modifications.

Finally, individual cells can be analyzed either in 3-dimensions at the single cell level or using high-throughput/high-content microscopy techniques and at all levels of microscopical resolution (wide-field, confocal and super-resolution). Of note, the procedure described in this report preserves nuclear integrity (Figure S4), similar to other FISH procedures that include, for example, repeated freezing and thawing in liquid nitrogen [28].

In summary, the Repli-FISH method described here is an efficient and complementary approach to the study of replication kinetics of a broad range of DNA sequences that are otherwise difficult to map and study with the currently available next-generation sequencing methods.

#### Acknowledgments

We are indebted to Anne Lehmkuhl and Stephanie Meyer for excellent technical assistance. We thank John V. Morgan (U. Michigan Medical School, USA) for providing the pLRE3-eGFP plasmid.

*OBM Genetics* 2019; 3(1), doi:10.21926/obm.genet.1901062

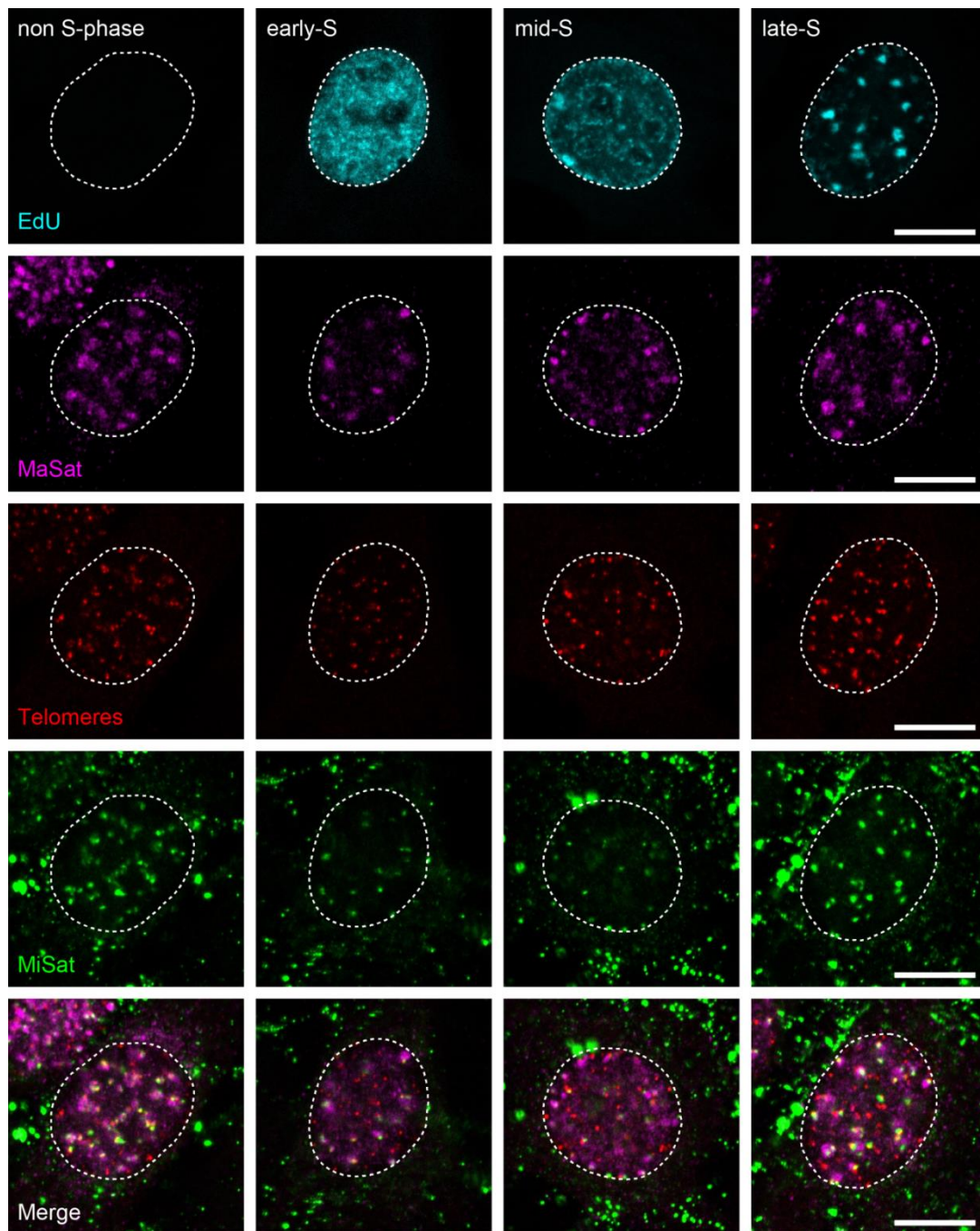
## Additional Materials

1. Figure S1: The epitope for the monoclonal anti-PCNA antibody (PC10) maps to the interface of the PCNA monomers.



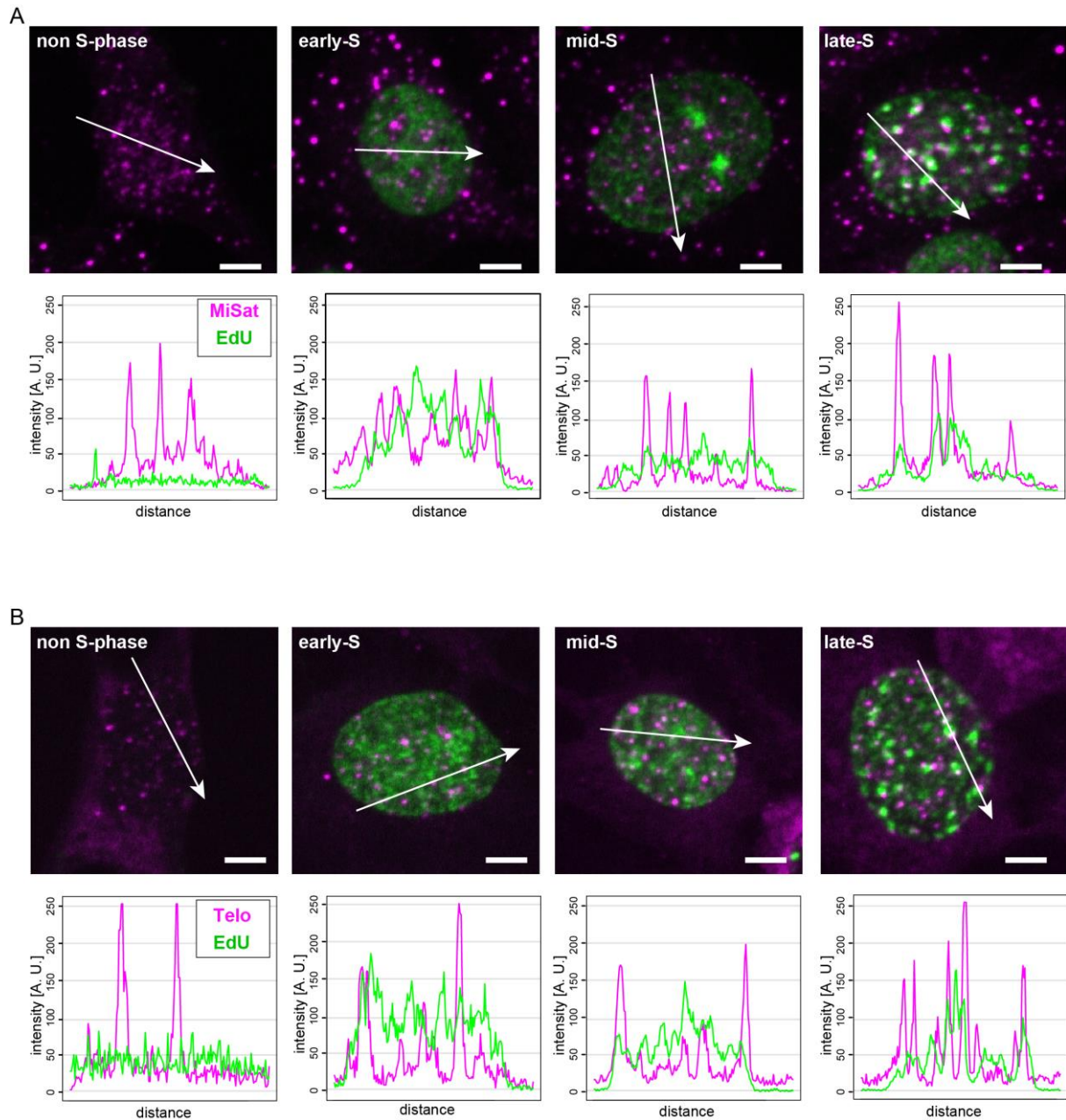
**Figure S1** The epitope for the monoclonal anti-PCNA antibody (PC10) maps to the interface of the PCNA monomers. The epitope of the PC10 anti-PCNA monoclonal antibody [54] spans amino acids 111–125 (A, red stretch, see [55]), which are located right at the interface between neighboring monomers within the homotrimeric PCNA ring-complex (B and C, PDB: 1AXC). Due to this, immunodetection of PCNA requires partial denaturation of the complex, e.g. via alcohol (post-) fixation or, as is the case during the *in situ* hybridization procedure, by heat-induced denaturation.

## 2. Figure S2: Triple-color FISH combined with replication staining.



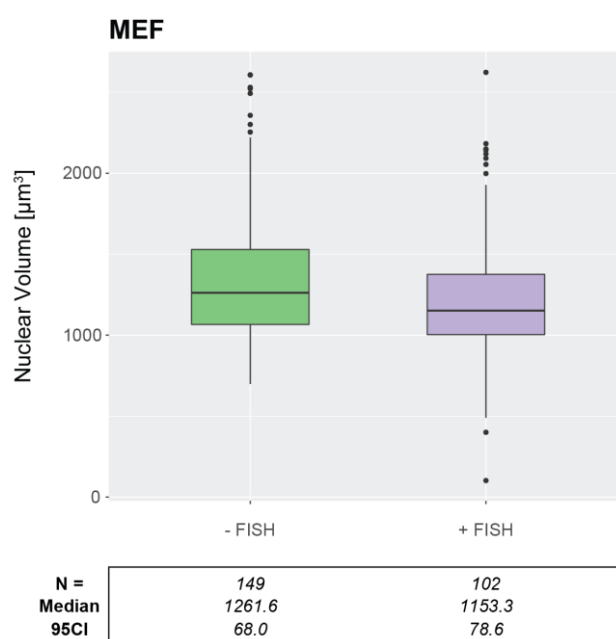
**Figure S2** Triple-color FISH combined with replication staining. Telomere- (red), MaSat- (magenta) and MiSat-specific (green) FISH probes were simultaneously hybridized to interphase cells and detected in combination with EdU, the latter marking sites of nascent DNA (cyan, upper row). Like the results presented in Figure 4, this confirms that multiple probes can be detected and studied simultaneously within the same cell nucleus. Scale bar: 10  $\mu$ m.

## 3. Figure S3: Localization of MiSat- and telomere-specific probes in mouse myoblast cells at different S-phase substages.



**Figure S3** Localization of MiSat- and telomere-specific probes in mouse myoblast cells at different S-phase substages. As in Figure 7, cells were categorized based on the nuclear distribution of the EdU signal into non-S-phase or early, mid and late replicating cells. The colocalization between MiSat FISH (A) and telomere FISH (B) signals with EdU (green) was evaluated by line profile analysis along the white arrows in each cell nucleus. Scale bar: 5  $\mu$ m.

## 4. Figure S4: Effect of the hybridization procedure on nuclear integrity.



**Figure S4** Effect of the hybridization procedure on nuclear integrity. Comparison of mouse embryonic fibroblast cells used in *in situ* hybridization experiments (+ FISH) and unhybridized control samples (- FISH) shows that the hybridization procedure, as described here, only slightly affects the nuclear volume of the cells. N = number of cells from one (+ FISH) and two (- FISH) biological replicates. Boxplots represent the median (bold horizontal line), first and third quantiles (lower and upper boundaries of the box), lower and upper whiskers (vertical lines extending the box boundaries) as well as outliers (black dots).

### Author Contributions

C.R., P.W. and M.C.C. wrote the manuscript. A.S. performed experiments in HeLa cells with Alu and LINE-1 specific probes. C.R. and P.W. performed experiments in mouse myoblast and ES cells, respectively using probes for major, minor and telomeric repeats and in metaphase cells. C.R. and P.W. analyzed the data.

### Funding

This work was supported by grants of the Deutsche Forschungsgemeinschaft (GRK1657/TP1B and DFG CA198/9-1 and 2) to M.C.C..

### Competing Interests

The authors have declared that no competing interests exist.

### References

1. Chagin VO, Stear JH, Cardoso MC. Organization of DNA replication. Csh Perspect Biol. 2010; 2: a000737.

*OBM Genetics* 2019; 3(1), doi:10.21926/obm.genet.1901062

2. Hiratani I, Ryba T, Itoh M, Yokochi T, Schwaiger M, Chang CW, et al. Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol.* 2008; 6: e245.
3. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature.* 2001; 409: 860-921.
4. Mouse Genome Sequencing C, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature.* 2002; 420: 520-562.
5. Batzer MA, Deininger PL. Alu repeats and human genomic diversity. *Nat Rev Genet.* 2002; 3: 370-379.
6. Korenberg JR, Rykowski MC. Human genome organization: alu, lines and the molecular structure of metaphase chromosome bands. *Cell.* 1988; 53: 391-400.
7. Ostertag EM, Kazazian HH, Jr. Biology of mammalian L1 retrotransposons. *Annu Rev Genet.* 2001; 35: 501-538.
8. Martin SL, Bushman FD. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol Cell Biol.* 2001; 21: 467-475.
9. Kazazian HH, Jr., Moran JV. The impact of L1 retrotransposons on the human genome. *Nat Genet.* 1998; 19: 19-24.
10. Chu WM, Liu WM, Schmid CW. RNA polymerase III promoter and terminator elements affect Alu RNA expression. *Nucleic Acids Res.* 1995; 23: 1750-1757.
11. Deininger P. Alu elements: know the SINEs. *Genome Biol.* 2011; 12: 236.
12. Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet.* 2003; 35: 41-48.
13. Moyzis RK, Buckingham JM, Cram LS, Dani M, Deaven LL, Jones MD, et al. A highly conserved repetitive DNA sequence, (TTAGGG)<sub>n</sub>, present at the telomeres of human chromosomes. *Proc Natl Acad Sci USA.* 1988; 85: 6622-6626.
14. de Lange T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.* 2005; 19: 2100-2110.
15. Wong AK, Rattner JB. Sequence organization and cytological localization of the minor satellite of mouse. *Nucleic Acids Res.* 1988; 16: 11645-11661.
16. Joseph A, Mitchell AR, Miller OJ. The organization of the mouse satellite DNA at centromeres. *Exp Cell Res.* 1989; 183: 494-500.
17. Ohno S. So much "junk" DNA in our genome. *Brookhaven Symp Biol.* 1972; 23: 366-370.
18. Orgel LE, Crick FH. Selfish DNA: the ultimate parasite. *Nature.* 1980; 284: 604-607.
19. Doolittle WF, Sapienza C. Selfish genes, the phenotype paradigm and genome evolution. *Nature.* 1980; 284: 601-603.
20. Choo KHA. Domain organization at the centromere and neocentromere. *Dev Cell.* 2001; 1: 165-177.
21. Lyon MF. Do LINEs have a role in X-chromosome inactivation? *J Biomed Biotechnol.* 2006.
22. Rubin CM, Kimura RH, Schmid CW. Selective stimulation of translational expression by Alu RNA. *Nucleic Acids Res.* 2002; 30: 3253-3261.
23. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet.* 2007; 8: 272-285.
24. Gey GO, Coffman WD, Kubicek MT. Tissue culture studies of the proliferative capacity of cervical carcinoma and normal epithelium. *Cancer Res.* 1952: 264-265.

*OBM Genetics* 2019; 3(1), doi:10.21926/obm.genet.1901062

25. Yaffe D, Saxel O. Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature*. 1977; 270: 725-727.
26. Peters AH, O'Carroll D, Scherthan H, Mechtler K, Sauer S, Schofer C, et al. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell*. 2001; 107: 323-337.
27. Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*. 1992; 69: 915-926.
28. Cremer M, Grasser F, Lanctot C, Muller S, Neusser M, Zinner R, et al. Multicolor 3D fluorescence in situ hybridization for imaging interphase chromosomes. *Methods Mol Biol*. 2008; 463: 205-239.
29. Liu P, Siciliano J, Seong D, Craig J, Zhao Y, de Jong PJ, et al. Dual Alu polymerase chain reaction primers and conditions for isolation of human chromosome painting probes from hybrid cells. *Cancer Genet Cytogenet*. 1993; 65: 93-99.
30. Frauer C, Rottach A, Meilinger D, Bultmann S, Fellingner K, Hasenoder S, et al. Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. *PLoS One*. 2011; 6: e16627.
31. Lehnertz B, Ueda Y, Derijck AA, Braunschweig U, Perez-Burgos L, Kubicek S, et al. Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr Biol*. 2003; 13: 1192-1200.
32. Anton T, Bultmann S, Leonhardt H, Markaki Y. Visualization of specific DNA sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system. *Nucleus*. 2014; 5: 163-172.
33. Garcia-Perez JL, Morell M, Scheys JO, Kulpa DA, Morell S, Carter CC, et al. Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. *Nature*. 2010; 466: 769-773.
34. Celeda D, Aldinger K, Haar FM, Hausmann M, Durm M, Ludwig H, et al. Rapid fluorescence in situ hybridization with repetitive DNA probes: quantification by digital image analysis. *Cytometry*. 1994; 17: 13-25.
35. Celeda D, Bettag U, Cremer C. A simplified combination of DNA probe preparation and fluorescence in situ hybridization. *Z Naturforsch C*. 1992; 47: 739-747.
36. Paulasova P, Pellestor F. The peptide nucleic acids (PNAs): a new generation of probes for genetic and cytogenetic analyses. *Ann Genet*. 2004; 47: 349-358.
37. Koch JE, Kolvraa S, Petersen KB, Gregersen N, Bolund L. Oligonucleotide-priming methods for the chromosome-specific labelling of alpha satellite DNA in situ. *Chromosoma*. 1989; 98: 259-265.
38. Pellestor F. PRINS and in situ PCR protocols. Totowa, NJ: Humana Press; 2006.
39. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet*. 2011; 13: 36-46.
40. Natale F, Scholl A, Rapp A, Yu W, Rausch C, Cardoso MC. DNA replication and repair kinetics of Alu, LINE-1 and satellite III genomic repetitive elements. *Epigenetics Chromatin*. 2018; 11: 61.
41. Bravo R, Frank R, Blundell PA, Macdonald-Bravo H. Cyclin/PCNA is the auxiliary protein of DNA polymerase-delta. *Nature*. 1987; 326: 515-517.

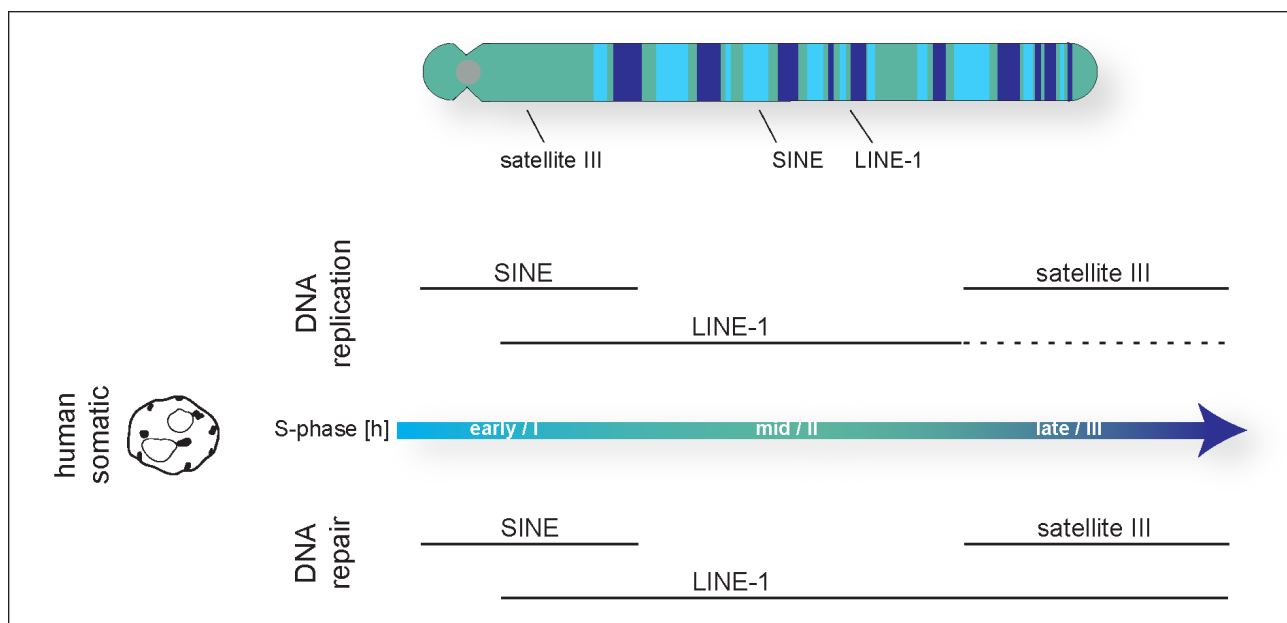
*OBM Genetics* 2019; 3(1), doi:10.21926/obm.genet.1901062

42. Celis JE, Madsen P. Increased nuclear cyclin/PCNA antigen staining of non S-phase transformed human amnion cells engaged in nucleotide excision DNA repair. *FEBS Lett.* 1986; 209: 277-283.
43. Celis JE, Madsen P, Nielsen S, Celis A. Nuclear patterns of cyclin (PCNA) antigen distribution subdivide S-phase in cultured cells--some applications of PCNA antibodies. *Leuk Res.* 1986; 10: 237-249.
44. Leonhardt H, Rahn HP, Weinzierl P, Sporbert A, Cremer T, Zink D, et al. Dynamics of DNA replication factories in living cells. *J Cell Biol.* 2000; 149: 271-280.
45. O'Keefe RT, Henderson SC, Spector DL. Dynamic organization of DNA replication in mammalian cell nuclei: spatially and temporally defined replication of chromosome-specific alpha-satellite DNA sequences. *J Cell Biol.* 1992; 116: 1095-1110.
46. Casas-Delucchi CS, van Bommel JG, Haase S, Herce HD, Nowak D, Meilinger D, et al. Histone hypoacetylation is required to maintain late replication timing of constitutive heterochromatin. *Nucleic Acids Res.* 2012; 40: 159-169.
47. Weidtkamp-Peters S, Rahn HP, Cardoso MC, Hemmerich P. Replication of centromeric heterochromatin in mouse fibroblasts takes place in early, middle, and late S phase. *Histochem Cell Biol.* 2006; 125: 91-102.
48. Arnoult N, Schluth-Bolard C, Letessier A, Drascovic I, Bouarich-Bourimi R, Campisi J, et al. Replication timing of human telomeres is chromosome arm-specific, influenced by subtelomeric structures and connected to nuclear localization. *PLoS Genet.* 2010; 6: e1000920.
49. Wright WE, Tesmer VM, Liao ML, Shay JW. Normal human telomeres are not late replicating. *Exp Cell Res.* 1999; 251: 492-499.
50. Zou Y, Gryaznov SM, Shay JW, Wright WE, Cornforth MN. Asynchronous replication timing of telomeres at opposite arms of mammalian chromosomes. *Proc Natl Acad Sci USA.* 2004; 101: 12928-12933.
51. Marchal C, Sasaki T, Vera D, Wilson K, Sima J, Rivera-Mulia JC, et al. Genome-wide analysis of replication timing by next-generation sequencing with E/L Repli-seq. *Nat Protoc.* 2018; 13: 819-839.
52. Dimitrova DS. DNA replication initiation patterns and spatial dynamics of the human ribosomal RNA gene loci. *J Cell Sci.* 2011; 124: 2743-2752.
53. Shimizu N, Shingaki K. Macroscopic folding and replication of the homogeneously staining region in late S phase leads to the appearance of replication bands in mitotic chromosomes. *J Cell Sci.* 2004; 117: 5303-5312.
54. Waseem NH, Lane DP. Monoclonal antibody analysis of the proliferating cell nuclear antigen (PCNA). Structural conservation and the detection of a nucleolar form. *J Cell Sci.* 1990; 96: 121-129.
55. Roos G, Landberg G, Huff JP, Houghten R, Takasaki Y, Tan EM. Analysis of the epitopes of proliferating cell nuclear antigen recognized by monoclonal antibodies. *Lab Invest.* 1993; 68: 204-210.

## 5.2 CHAPTER II: DNA Replication and Repair Kinetics of Alu, LINE-1 and Satellite III Genomic Repetitive Elements

C. Rausch performed the line profile and Pearson's correlation analyses of the double metaphase FISH for Alu and L1 and prepared the corresponding figure part of Supplementary Figure S3a.

### Graphical Abstract






## RESEARCH

## Open Access



# DNA replication and repair kinetics of Alu, LINE-1 and satellite III genomic repetitive elements

Francesco Natale<sup>1,2†</sup>, Annina Scholl<sup>1†</sup>, Alexander Rapp<sup>1†</sup>, Wei Yu<sup>1,3</sup>, Cathia Rausch<sup>1</sup> and M. Cristina Cardoso<sup>1\*</sup> 

## Abstract

**Background:** Preservation of genome integrity by complete, error-free DNA duplication prior to cell division and by correct DNA damage repair is paramount for the development and maintenance of an organism. This holds true not only for protein-encoding genes, but also it applies to repetitive DNA elements, which make up more than half of the human genome. Here, we focused on the replication and repair kinetics of interspersed and tandem repetitive DNA elements.

**Results:** We integrated genomic population level data with a single cell immunofluorescence in situ hybridization approach to simultaneously label replication/repair and repetitive DNA elements. We found that: (1) the euchromatic Alu element was replicated during early S-phase; (2) LINE-1, which is associated with AT-rich genomic regions, was replicated throughout S-phase, with the majority being replicated according to their particular histone marks; (3) satellite III, which constitutes pericentromeric heterochromatin, was replicated exclusively during the mid-to-late S-phase. As for the DNA double-strand break repair process, we observed that Alu elements followed the global genome repair kinetics, while LINE-1 elements repaired at a slower rate. Finally, satellite III repeats were repaired at later time points.

**Conclusions:** We conclude that the histone modifications in the specific repeat element predominantly determine its replication and repair timing. Thus, Alu elements, which are characterized by euchromatic chromatin features, are repaired and replicated the earliest, followed by LINE-1 elements, including more variegated eu/heterochromatic features and, lastly, satellite tandem repeats, which are homogeneously characterized by heterochromatic features and extend over megabase-long genomic regions. Altogether, this work reemphasizes the need for complementary approaches to achieve an integrated and comprehensive investigation of genomic processes.

**Keywords:** Alu, ChIP-Seq, DNA repair, DNA repetitive elements, DNA replication, Genome-wide analysis, Immunofluorescence, LINE, Phosphorylated H2AX, Satellites

## Background

Preservation of genome integrity by complete, error-free DNA duplication prior to cell division and by correct DNA damage repair is paramount for the development and maintenance of an organism. This holds true not only for the protein-encoding genes, but also applies for

the repetitive DNA elements [1]. More than half of the human genome is made up of repetitive DNA elements. This fraction is remarkably large when compared to the ~1.2% protein coding DNA [2]. In mouse, the proportion is similar, with repetitive DNA elements and coding regions making up to 40% and 1.4% of the murine genome, respectively [3].

Long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs) and LTR retrotransposons are transposable DNA elements. These elements insert into new genomic locations by reverse

\*Correspondence: cardoso@bio.tu-darmstadt.de

†Francesco Natale, Annina Scholl, Alexander Rapp: Co-first authors.

<sup>1</sup> Department of Biology, Technische Universität Darmstadt, 64287 Darmstadt, Germany

Full list of author information is available at the end of the article



© The Author(s) 2018. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

transcription of an RNA intermediate. LINEs are found in all vertebrate species. The LINE-1 (L1) family of transposable elements represents the only group of autonomous non-LTR retrotransposons in the human genome [4]. Functional L1 elements encode a consensus sequence of about 6 kbp, including two open reading frames encoding for proteins that are necessary for the retrotransposition [5–8]. L1 retrotransposition requires transcription of L1 RNA, its transport to the cytoplasm, and translation of its two open reading frames. Both L1-encoded proteins (ORF1p and ORF2p) are thought to preferentially associate with their own encoding RNA and form a ribonucleoprotein complex, which is a proposed retrotransposition intermediate [9]. The latter must then access the nucleus, where the L1 endonuclease cleaves the genomic DNA at a degenerate consensus sequence. The resulting free 3' hydroxyl residue is subsequently used by the L1 reverse transcriptase as a primer to copy the L1 sequence in situ. Such process is termed “target-primed reverse transcription” [10]. Finally, the resulting L1 cDNA is joined to the target DNA. L1 elements alone make up about 17% of the human genome, and they are preferentially found at AT-rich and gene poor regions, corresponding to G-bands and DAPI-bright bands of metaphase chromosomes [2, 11]. However, the vast majority (>99%) of L1s are, on average, 1400 bp long and inactive because of point mutations, truncations and other rearrangements; it is estimated that the average diploid human genome contains about 100 retrotransposition-competent L1s [10].

Alu repetitive DNA elements are among the most abundant SINEs. They are about 280 bp long and are specific to primates. Similar SINEs can be found in other organisms, like B1 elements in rodents [12]. Alu elements do not encode proteins but contain a RNA polymerase III promoter [2], and it was demonstrated that they use L1-encoded proteins for their retrotransposition in trans [13, 14]. There exist more than  $10^6$  Alu elements in the human genome covering approximately 11% of the genomic DNA, and they are preferentially distributed in gene-rich genomic regions corresponding to R-bands in metaphase chromosomes [11]. Therefore, based on their genomic distribution, L1 and Alu elements represent chromatin compartments with opposing features. Furthermore, L1 and, potentially, Alu activity represents a potential threat for the integrity and stability of the genome, in both dividing and nondividing cells. By direct insertion of the transposable element into or close to a gene, L1 might interfere with gene activity, disrupting exons or influencing splicing [4]. Furthermore, because of their abundance, their sequences may be used in homologous recombination (HR) in a non-allelic fashion, leading to insertions or deletions in the damaged region [15, 16].

Indeed, insertions of L1 elements have been reported in tumor suppressor genes in several cancer types [17–19].

Satellite DNA elements consist of very large arrays of tandemly repeating, non-coding DNA. They are the main component of functional centromeres and form the main component of constitutive heterochromatin. Human satellite III and murine major satellite are examples of pericentromeric heterochromatin, while human alpha satellite and mouse minor satellite can be found in centromeres [20–24]. Mouse major satellite reaches up to 8 Mbp and is made up of 234 bp-long AT-rich units. It is found in pericentromeric regions of all chromosomes, except for the Y chromosome. In interphase nuclei, major satellite DNA can be found at bright DAPI-stained DNA regions. The latter consist of clusters of heterochromatic regions from several chromosomes and are known as “chromocenters” [3, 20]. Human satellite III consists of a 5-bp-long unit and its presence was shown in seven autosomes (chromosomes 1, 9, 13, 14, 15, 21 and 22) and the Y chromosome [21]. Despite their epigenetic state, pericentromeric satellites (e.g., satellite III) have been shown to be transcriptionally active in response of various stressors (UV-C, genotoxic chemicals, osmotic imbalance, oxidative stress and hypoxia). Transcription of these elements not only can stabilize pericentromeric regions, but also it can promote recovery from stress by activating alternative splicing and, thus, modulating critical stress response genes [25–27]. Finally, a large number of repetitive DNA elements can exist in at least two conformations: a right-handed B form (the most abundant) with canonical Watson–Crick base pairing and non-B conformations, possibly transiently formed at specific sequence motifs. The latter may arise from supercoil density, partly generated by transcription or protein binding, and are involved in genome susceptibility to DNA damage [28].

Overall, the DNA duplication and repair of repetitive DNA elements before cell division are paramount to genome integrity. However, the spatio-temporal organization of DNA duplication and repair of repetitive elements is yet to be fully elucidated. In this study, we investigated the DNA replication timing and DNA double-strand break repair kinetics of different repetitive DNA elements. We integrated publicly available genomic data (ChIP-Seq, Repli-Seq and other repetitive and non-B DNA sequences) with a combined immunofluorescence in situ hybridization (FISH) analysis to visualize DNA replication or DNA damage response (DDR) sites in the context of repetitive DNA elements. Our results reemphasize the need for complementary approaches to achieve an integrated and comprehensive investigation of any genomic process.

## Results

### Genome-wide replication timing of repetitive DNA elements correlates with GC content, gene density and chromatin state

First, we asked whether the replication timing of repetitive DNA elements may depend on the genomic distribution and, thus, the chromatin state of such elements. To characterize such relation, we retrieved publicly available genomic data of 12 different human repetitive elements, covering all different types: direct, inverted, mirror, tandem (microsatellite/SSRs), low complexity (AT and GC) and interspersed elements. The latter were dissected further into SINEs (Alu and MIR), LINEs (L1 and L2) and LTR/DNA transposons (MER) without further subdivision into subfamilies (Table 1; [29, 30]). These classifications utilize the reference genome assembly and are not ploidy-adjusted.

DNA sequence composition is one of the genomic features dictating the distribution of repetitive DNA elements in the genome. Hence, we started with computing the abundance of the 12 different repetitive DNA elements (i.e., their number) in 10 kbp genomic intervals, which is the genomic resolution we adopted in this study (example tracks are given in Additional file 1: Fig. S1). Then, we calculated the genome-wide Spearman's rho correlation coefficient between each repetitive DNA element abundance and different GC (or AT) content. Alu, GC low complexity, MIR and, to a minor extent, direct repeats were positively correlated with GC content, whereas AT low complexity and L1 were negatively

correlated to GC content (Fig. 1a, left column). Repetitive DNA elements showing positive correlation with GC content also showed a gradually decreasing correlation (or an anti-correlation) with decreasing GC content (Fig. 1a). The inverted case was observed for those repetitive DNA elements showing a negative correlation with GC content, instead. Little to no correlation with GC content was observed for microsatellites/SSRs and mirror repeats (Fig. 1a).

Next, we investigated the relation between the above-mentioned repetitive elements and diverse histone modifications retrieved from publicly available databases (HeLa S3 cells, [ENCODE tier 2]). We computed the genome-wide Spearman's rho correlation coefficient between each DNA repetitive element and each given histone modification. Repetitive DNA elements scoring an (anti-)correlation with GC content also scored a strong (anti-)correlation with the majority of the histone modifications we tested. For example, Alu elements were abundant on genomic locations whose chromatin was decorated by typical active promoter (H3K4me3/2, H3K9ac, H3K27ac) or gene body (H3K36me3, H3K79me2) modifications. Conversely, Alu elements were scarce on genomic regions whose chromatin was marked by H3K9me3 (Fig. 1b). Overall, Alu abundance positively correlated with genic regions (Fig. 1b, right column). L1 elements showed the opposite behavior with an anti-correlation to most euchromatic marks and genic elements and a weak positive correlation to H3K9me3.

**Table 1 Overview of human repetitive and non-B DNA elements**

DNA sequence element	Type of sequence	Data source [29, 30]
Alu sequence	Short interspersed nuclear element	RepeatMasker
MIR	Short interspersed nuclear element	RepeatMasker
LINE1	Long interspersed nuclear element	RepeatMasker
LINE2	Long interspersed nuclear element	RepeatMasker
MER	LTRs/DNA transposons	RepeatMasker
AT low complexity repeats		RepeatMasker
GC low complexity repeats		RepeatMasker
Simple sequence repeats (SSRs)	Tandem repeats	RepeatMasker
G-Quadruplex forming repeats	non-B DNA structure	RepeatMasker
Z-DNA motif	non-B DNA structure	Non-B DataBase
Inverted repeats	Repetitive DNA element	Non-B DataBase
Cruciform motif	non-B DNA structure	Non-B DataBase
Direct repeats	Repetitive DNA element	Non-B DataBase
Slipped motif	non-B DNA structure	Non-B DataBase
Mirror repeats	Repetitive DNA element	Non-B DataBase
Triplex motif	non-B DNA structure	Non-B DataBase
A-Phased motif	non-B DNA structure	Non-B DataBase
Microsatellite	Tandem repeats	RepeatMasker

This observation may have implications in the DDR and is discussed below.

Finally, we correlated the abundance of the above-mentioned repetitive DNA elements with the stages of the S-phase of the cell cycle obtained by publicly available Repli-Seq experiments from HeLa S3 cells (ENCODE, tier 2) [31]. The Spearman's correlation coefficient computed between Repli-Seq data and the above-mentioned repetitive DNA elements indicated that Alu elements were abundant on chromatin regions whose DNA was duplicated in the early G1b and S1 stages, but poorly represented in those regions duplicated in the S3, S4 and G2 late stages (Fig. 1c). Interestingly, despite showing a negative correlation with transcription-permissive histone modifications, L1 showed little to no correlation to any S-phase replication substage with L1 DNA being duplicated throughout S-phase with only a slight increase in mid and late S-phase (Fig. 1c).

We then asked whether the chromatin landscape played a role in the DNA replication of L1 elements and, specifically, in genomic regions where L1 elements are abundant. To this end, we segmented the genome into L1-rich (containing more than 10 L1 elements per 10 kpb interval) and L1-poor (no L1 elements), and then we computed the correlation coefficients between histone marks and replication substages. All transcription-permissive histone modifications showed a positive correlation with early S-phase (G1b and S1), independent of the abundance of L1 elements (Fig. 1d). This indicates that chromatin regions decorated with transcription-permissive marks were replicated during the early stage of S-phase. For these regions, we observed a transition phase during the mid-stage of S-phase, wherein the positive correlation shifted toward a negative correlation (Fig. 1d, S2 and S3), the latter persisting through the late substages of S-phase (Fig. 1d, S3 and S4). Heterochromatin-associated histone modifications (H3K27me3 and H3K9me3) showed a marked difference between L1-poor and L1-rich genomic regions, instead. Specifically, the latter presented reduced correlations throughout the cell cycle, suggesting an independent DNA replication mechanism throughout S-phase, once DNA replication commenced. Genomic regions devoid of L1

elements presented a pattern similar to that we observed for the whole genome, with H3K27me3/H3K9me3-decorated chromatin being replicated in the late stage of the S-phase.

To test the generality of our observations, we analyzed two additional cell lines: the lymphoblastoid GM12878 cells and the hepatocellular carcinoma HepG2. Similar to our observations in HeLa cells, an identical temporal pattern was observed for transcription-permissive histone modifications (Additional file 1: Fig. S2). Heterochromatin-associated histone modifications presented marked cell-specific differences, instead. Specifically, HepG2 cells presented a late replication of H3K9me3-decorated chromatin (Additional file 1: Fig. S2c). The latter was replicated much earlier in GM12878 cells, instead (Additional file 1: Fig. S2b). Further, L1 abundance seemed to have a cell-specific impact on the correlation between heterochromatin-associated histone modifications and replication timing: while we observed a difference between L1-poor and L1-rich genomic regions in GM12878 cells, no difference was observed in HepG2 cells. For the former, it was the absence of L1 elements that led to a decrease in the correlation coefficient throughout the S-substages.

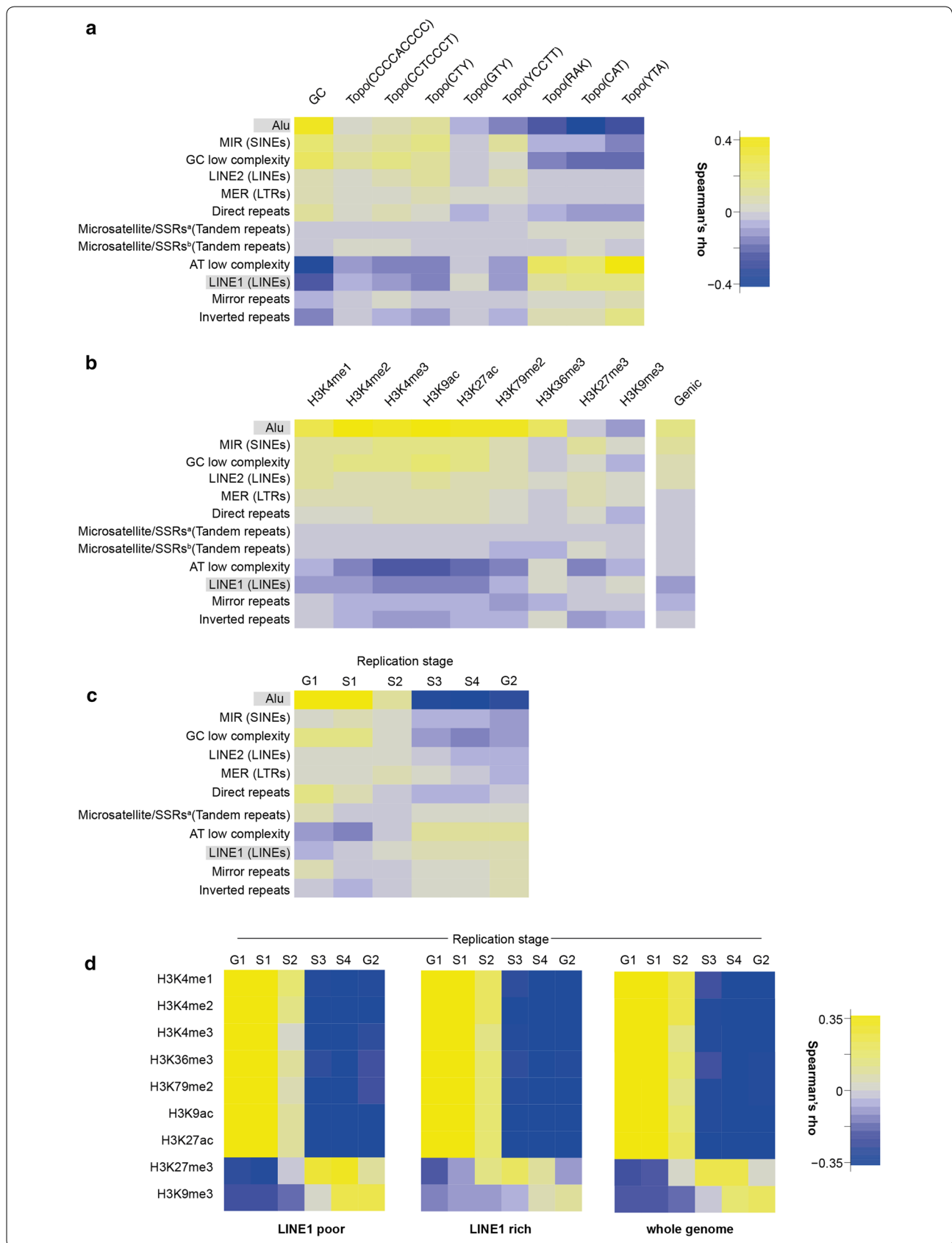
Taken together, these observations reveal that the histone modifications predominantly dictate the replication timing. Yet, the presence of L1 elements—and possibly their transcriptional context—locally perturbs the replication program.

#### FISH-based replication timing assessment of interspersed and tandem repeats

Despite their high throughput and the high attainable read depth, it proves challenging to quantitatively assess highly repetitive DNA regions such as (peri-)centromeric chromatin by next-generation sequencing approaches. Sequences at the boundary of a given highly repetitive DNA region can be mapped by utilizing the non-repetitive (and, thus, mappable) adjacent sequences. However, peri- and centromeric regions, covering up to 15% of the genome, are harder to be probed by sequencing methods, especially for quantitative analyses. Therefore, to assess the DNA duplication prior to cell division of (peri-)

(See figure on next page.)

**Fig. 1** Genomic features of repetitive DNA elements. Spearman's rho correlation matrix. The number of each repetitive DNA element copies, or the amount of a given genomic feature is counted in each 10 kb genomic interval. The correlation coefficient is calculated for comparison of repetitive DNA elements with GC content (**a**), histone modifications (from HeLa, ENCODE tier 2) or genic regions (**b**) and replication timing from Repli-Seq data (HeLa, ENCODE tier 2) [31] (**c**). Data are from >290,000 genomic intervals. For each correlation,  $P < 2.2 \times 10^{-16}$ . In **a**, Topo(...): from left to right, topoisomerase I consensus sequences at decreasing GC content. Highlighted Alu and L1 repetitive elements are arbitrarily chosen to define chromatin compartments with opposing chromatin features, and are further investigated in FISH experiments. **d** Correlation matrix of histone modifications and replication timing in HeLa cells, for L1-rich (> 10 counts per genomic interval) L1-poor (> 1 count per genomic element) genomic regions



centromeric chromatin, we established an immunofluorescence (FISH)-based method. Together with (peri-)centromeric chromatin, we also probed Alu and L1 repetitive DNA elements, as they recapitulate chromatin compartments with opposing functional features (e.g., euchromatin versus heterochromatin) (Fig. 1b). Specifically, we investigated whether the repetitive DNA elements showed a temporal replication similar to that of the chromatin compartment they are mainly associated with. To address this question, we combined FISH, with probes for the repetitive DNA elements, with the detection of incorporated thymidine analogues to dissect the S-phase stages. Different chromatin types are replicated at different stages of the S-phase, which are identifiable by their spatial patterns [32]. Thus, we incubated unsynchronized HeLa (and C2C12) cells with 10  $\mu$ M EdU for 15 min to label the replicating DNA, and then probed Alu, L1 and satellite III (and major satellite) elements with specific DNA probes.

First, we validated the specificity of the hybridization probes on mitotic chromosomes. As mentioned before, Alu elements are predominantly found in GC-rich regions and, thus, R-bands, whereas L1 are more abundant in GC-poor regions (or AT-rich). To counterstain the DNA, we employed DAPI, which preferentially binds to AT-rich DNA sequences. We simultaneously hybridized a biotin-labeled Alu probe and a digoxigenin labeled L1 probe on HeLa metaphases. Alu and L1 were indeed found to negatively correlate in color line profiles and Pearson's correlation analysis showed anti-correlation of Alu and L1 with a  $\rho$  value of  $-0.23$  (Additional file 1: Fig. S3a). In addition, we probed human satellite III DNA, which forms constitutive heterochromatin with a probe specific for satellite III sequences found on chromosome 1. According to previous spectral karyotype analysis of HeLa cells [33], about four copies of satellite III per cell are expected. We observed more than four hybridization signals, indicating that other satellite III (or satellite II) sequences (most likely on chromosomes 9 and 16) were probed under our experimental conditions (Additional file 1: Fig. S3a). As a control, we also probed murine major satellite DNA in C2C12 myoblast cells. Each telocentric mouse chromosome possesses a major satellite

DNA sequence, which was efficiently labeled by the probe (Additional file 1: Fig. S3a).

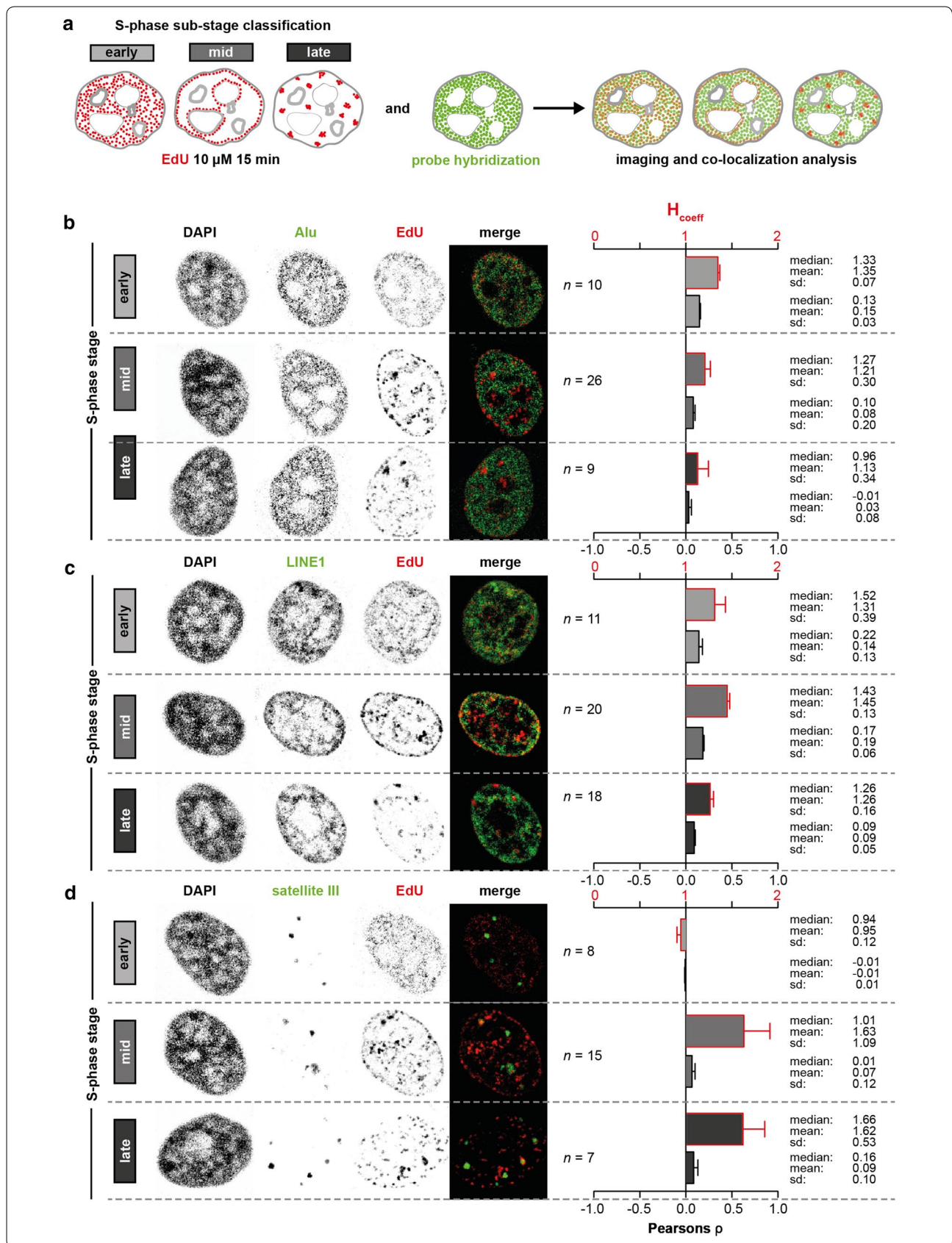
To investigate the replication timing of the above-mentioned repetitive elements in interphase, we measured the degree of colocalization by means of the  $H$  coefficient ( $H_{\text{coefficient}}$ ) as well as the Pearson's correlation coefficient ( $\rho$ ) [34]. The greater the  $H_{\text{coefficient}}$  is ( $>1$ ), the more the two signals colocalize.  $H_{\text{coefficient}}$  values lower than 1 indicates the two signals are randomly distributed. For example, L1 FISH signal (AT-rich) was more correlated with the DAPI signal than Alu FISH signal in interphase nuclei, and, thus, it showed a greater  $H_{\text{coefficient}}$  (Additional file 1: Fig. S3b). The Pearson's coefficient ranges between  $-1$  (anti-correlated signals) to  $+1$  (correlated signals).

The EdU pulse labeling allowed us to identify three different substages of the S-phase [35]: the early S-phase presented nuclear EdU foci distributed throughout the nuclear interior; the mid S-phase mainly showed foci at the peri-nuclear and peri-nucleolar regions; the late S-phase exhibited distinct nuclear spots corresponding to highly compacted chromatin (i.e., heterochromatin) in HeLa cells (Fig. 2a). We validated the S-phase classification using fluorescence staining of incorporated EdU by measuring the total genomic DNA content of the staged cells and comparing with the DNA content of the EdU negative cells in the population. The latter were further subdivided into G1 and G2 based on their nuclear volume and DAPI content. We observed a steadily increase in the DNA content of the S-phase staged cells from early via mid-to-late S-phase cells, with all three populations exhibiting a larger DNA content than G1 cells and a smaller compared to G2 cells (Additional file 1: Fig. S3c). After probe hybridization, cells were imaged, deconvolved and staged according to their S-phase pattern. To measure the colocalization of the EdU and FISH signals, first a nuclear mask was generated based on the DAPI (DNA) channel. Then, to remove background signal, a local mean filter was applied to the channels to be compared. Finally, the Pearson's and  $H_{\text{coefficient}}$  were calculated for each z-plane (Additional file 1: Fig. S4a).

Similar to the genomic data, Alu DNA duplication prior to cell division was strongly associated to the early substage of the S-phase and anti-correlated to the late

(See figure on next page.)

**Fig. 2** Replication timing of repetitive DNA elements analyzed by FISH and S-phase substages classification. **a** Schematics of the experiment. HeLa cells were pulse-labeled with EdU for 15 min to allow the classification of different substages of the S-phase of the cell cycle (early, mid and late). Cells are then fixed, the probe is hybridized and microscopy is performed. **b** (left) Representative confocal and deconvolved micrographs of HeLa cells depicting the DAPI, Alu elements and EdU as inverted gray channels, at the three different S-phase substages. Merge is shown in pseudo-colors. Scale bar: 5  $\mu$ m. (right) Colocalization analysis of FISH and EdU signal at the three different S-phase substages via  $H_{\text{coefficient}}$  and Pearson's correlation coefficient as indicated. Error bars show the standard error of the mean. Data are from three independent experiments.  $n$  combined total number of cells analyzed.  $sd$  standard deviation. **c, d** Represent the same as in **b** for L1 and satellite III, respectively



S-phase (Fig. 2b). L1 DNA replication was associated with all the substages of the S-phase (Fig. 2c). Satellite III DNA duplication showed weak anti-correlation with the early S-phase and strong positive correlation with mid and late S-phase, instead (Fig. 2d). We obtained similar results when we probed the major satellite elements in C2C12 mouse cells, whereby the highest colocalization was detected during late S-phase (Additional file 1: Fig. S5; [36]). Both, H and Pearson's, coefficients showed similar outcomes. Overall, we observed an euchromatin-to-heterochromatin temporal trend, which can be recapitulated by Alu/L1 and satellite III elements. The outcome of the immuno-FISH analysis strongly supports the genomic data on Alu replication kinetics and, to a lesser extent, also the L1 element genomic data, which we demonstrated to be associated with early and late replicating loci in the genome and be rather dictated by the chromatin modifications they are embedded in. In addition, it extends the investigation to satellites, thus highlighting the importance of performing such combined analysis integrating the benefits of both sequencing and hybridization methodologies to achieve a complete genomic coverage.

#### Repair of repetitive DNA elements follows an euchromatin-to-heterochromatin spatio-temporal trend

DNA replication is one of the major causes of endogenous DNA double-strand breaks (DSBs) at collapsed replication forks. Repair and resolution of these lesions is paramount for an error-free cell division. Because DNA replication is temporally and spatially organized [32, 37], we next investigated whether the DNA repair of the above-mentioned repetitive DNA elements followed the same trend we observed for the replication.

To evaluate the cellular DNA damage response (DDR) after exposure to ionizing radiation (IR), we assessed the genomic distribution of histone H2AX phosphorylation ( $\gamma$ H2AX) by ChIP-Seq at early (0.5 h), mid (3 h) and late (24 h) time points in HeLa cells. Similarly to the analysis of the previously described histone modifications, we computed  $\gamma$ H2AX abundance in 10 kbp genomic intervals. Then, we calculated the Spearman's correlation coefficient between  $\gamma$ H2AX abundance and the number of the above-mentioned repetitive DNA elements per genomic interval. The outcome of this analysis indicated that  $\gamma$ H2AX was enriched within genomic sequences in which Alu, GC low complexity repeats and direct repeats were abundant (Fig. 3a) at early time points after irradiation. Conversely, AT low complexity repeats and L1 elements presented a lower  $\gamma$ H2AX signal (Fig. 3a). This trend was inverted at 24 h post-IR for most of the probed repetitive elements. Also, the abundance of the substrate

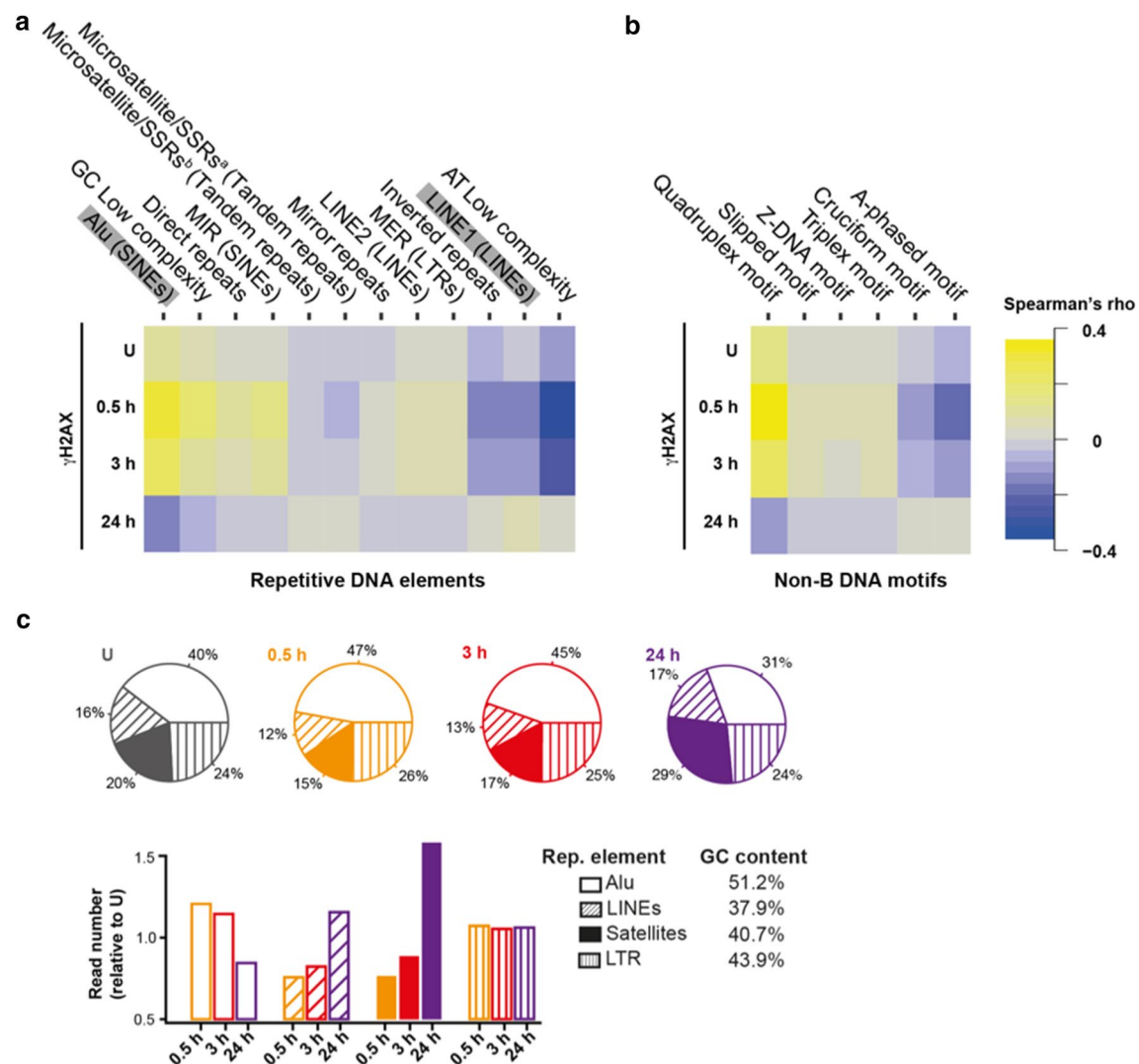
histone (H2AX) was not relevant for the outcome of the analysis, as only minor differences were observed by comparing input DNA-normalized and H2AX-normalized data (Additional file 1: Fig. S6). Sample genomic loci with tracks showing the repetitive elements and the  $\gamma$ H2AX density over the time of the DNA damage response are shown in Additional file 1: Fig. S7.

A large number of repetitive DNA elements can exist in at least two conformations: B and non-B forms. These structures—especially the latter—are involved in genome susceptibility to DNA damage [28]. To investigate which non-B DNA structure was more likely to be formed by a given repetitive DNA element, we first retrieved the genomic distribution of six non-B DNA motifs (cruciform, slipped, triplex, G-quadruplex, Z-DNA and A-phased) from “non-B DB” (<https://nonb-abcc.ncifcrf.gov/>)—a database for integrated annotation and analysis of non-B DNA forming motifs—and, next, we computed the number of such structures in 10 kbp genomic intervals. Then, we calculated the Spearman's correlation coefficient between the non-B forms and the previously investigated repetitive DNA elements (Additional file 1: Fig. S8). Microsatellites were strongly associated with Z-DNA and slipped motifs, while inverted repeats showed no correlation with triplex and Z-DNA motifs, but were highly correlated with cruciform motifs (Additional file 1: Fig. S8).

Interestingly, A-phased motifs—formed at A-rich tracts and involved in helix bending and transcription regulation [38, 39]—showed a complex relation with interspersed repetitive DNA elements: Alu and L1 elements were positively correlated with A-phased motifs, yet their respective cognates, MIR and L2, showed a negative correlation (Additional file 1: Fig. S8).

Next, we correlated the abundance of  $\gamma$ H2AX with the retrieved non-B DNA motifs, over time. In the absence of IR, the endogenous  $\gamma$ H2AX signal was slightly correlated with G-quadruplex motifs (Fig. 3b). The latter presented the highest  $\gamma$ H2AX levels directly after irradiation (Fig. 3b). Cruciform motifs, which are mainly formed by inverted and direct repeats and are associated with H3K9me3, presented low  $\gamma$ H2AX signal, compared to G-quadruplex. Slipped motifs, which are also found in inverted repeats-rich genomic regions, presented no correlation with  $\gamma$ H2AX, instead. This divergent behavior highlights the complexity of the genomic compartmentalization and collocates slipped motifs in regions of chromatin responding to DNA DSB signaling at earlier stages of DDR.

We observed yet another divergent behavior pertaining to the Alu and L1 elements. Both elements are positively associated to A-phased motifs, yet the latter are negatively correlated with  $\gamma$ H2AX (Fig. 3b). This observation



**Fig. 3** Genome-wide DNA repair kinetics of non-B and repetitive DNA elements. **a** Spearman's rho correlation matrix between repetitive **a** and non-B **b** DNA elements and  $\gamma$ H2AX levels before and after (0.5, 3 and 24 h) IR in HeLa cells. Calculation of the correlation coefficient is as in Fig. 1. **c** (top) Pie-charts showing the distribution of read counts for Alu, LINES, satellites and LTR repetitive DNA elements, before and after (0.5, 3 and 24 h) IR. (bottom) Bar-plots showing the relative enrichment for the repeat element indicated after (0.5, 3 and 24 h) IR. The number of reads for a given repetitive element and at a given time point was normalized over the corresponding number of reads before IR (for details see Materials and methods section). The respective GC content of the repeat is indicated (whole human genome GC content is 43%)

may underline a specific regulation of DDR in A-phased-rich regions.

To measure the total  $\gamma$ H2AX signal mapped to Alu, LINES, satellites and LTRs repetitive elements in an unbiased fashion, we made use of the raw sequence data, as these elements are not filtered for unique mappability. First, we mapped the quality-filtered raw reads to the corresponding repetitive elements as annotated in RepeatMasker. Then, the number of reads in each class was normalized to the total number of repetitive elements reads, containing only signatures of a single repetitive

element type, so that the resulting fraction represents the genome-wide  $\gamma$ H2AX coverage in a given repetitive element, which we deemed "metarepetitive element." The analysis of the metarepetitive elements revealed that the Alu signature increased by about 7% at 0.5 h post-IR (47%), with a concomitant decrease in the LINES (-4%) and satellites (-5%) signatures, compared to the unirradiated control (Fig. 3c). Conversely, 24 h post-IR the Alu signature decreased to 31%, whereas the satellites signature increased from 15 to 29% (compared to 0.5 h

post-IR). The read counts containing LTR signature mainly remained unvaried (Fig. 3c).

#### FISH-based repair timing assessment of interspersed and tandem repeats

We, then, investigated the DDR in repetitive elements by means of immuno-FISH, as we did for replication. Because H2AX distribution can affect, per se, the spreading of  $\gamma$ H2AX, we first performed a correlation analysis between the H2AX histone distribution and the Alu, L1 or satellite III DNA repetitive elements using the  $H_{\text{coef}}$  coefficient and the Pearson's coefficient. HeLa cells were fixed and immunostained for H2AX, and then incubated with Alu/L1/satellite III probes for the hybridization. The analysis was then performed as previously described (Additional file 1: Fig. S4b). While Alu and L1 showed a clear positive correlation with H2AX histone distribution, satellite III signal showed no correlation with H2AX distribution (Additional file 1: Fig. S9). This is in line with the outcome of the genomic data.

Next, we investigated  $\gamma$ H2AX response and its association with Alu, L1 or satellite III DNA repetitive elements. The cells were irradiated with 2 Gy X-rays and incubated for 0.5, 3 or 24 h post-IR, to recapitulate the early, mid and late stages of DDR (Fig. 4a). Cells were fixed at the indicated times and immunostained for  $\gamma$ H2AX followed by hybridization with Alu/L1/satellite III probes. The analysis was performed as previously described, with the exception that the first image mask was built from the  $\gamma$ H2AX signal (Additional file 1: Fig. S2b). This allowed us to directly compare the repetitive element and the  $\gamma$ H2AX fluorescent signals. The focal  $\gamma$ H2AX pattern was segmented, and the fraction of each repetitive element within the segmented  $\gamma$ H2AX space was calculated for all time points. This was done by taking the sum of repetitive element intensity values within the segmented  $\gamma$ H2AX focal structures divided by the total nuclear intensity of the repetitive element. Data were further normalized to the median of the 0.5 h time point to represent the fold change in the damaged regions and in view of the strong differences in  $\gamma$ H2AX levels at the different time points. In the absence of IR,  $\gamma$ H2AX signal was low and not sufficient to efficiently run the analysis, and, therefore, it was

omitted. Images of unirradiated cells are shown in Additional file 1: Fig. S10.

The highest fraction of damaged DNA is expected at 0.5 h post-IR, with a decrease over the time, as the damage is repaired and  $\gamma$ H2AX signature is removed. This behavior was observed for all the repetitive elements investigated. The total fraction of Alu or L1 in  $\gamma$ H2AX foci was highest upon exposure to IR (0.5 h) (Fig. 4b, c) and decreased to about 50% at 3 h post-IR. 24 h post-IR, the repetitive element fraction in  $\gamma$ H2AX foci was about 14–39% of the original (0.5 h post-IR) (Fig. 4b, c). For satellite III DNA, we observed a delayed kinetics, whereby the fraction at 3 and 24 h post-IR was about 65% and 23%, respectively (Fig. 4d, filled boxes). Because satellite DNA elements present a focal structure, we also segmented the satellite regions and determined the fraction of total  $\gamma$ H2AX signal within the segmented regions (Fig. 4d, empty boxes). As Alu and L1 spread over the whole genome/nucleus and thus do not allow efficient segmentation, this reciprocal analysis was not performed. In satellites, the fraction of  $\gamma$ H2AX remained unvaried up to 3 h post-IR and decreased to 55% at 24 h. Visual inspection of the images revealed that in many cells, the satellite III regions contain only a few  $\gamma$ H2AX foci (with partial overlap due to DNA decondensation) or, vice versa, the majority of  $\gamma$ H2AX foci contained only a few satellite III regions with partial overlap. We obtained similar results when we probed the major satellite in C2C12 mouse cells (Additional file 1: Fig. S11a, b).

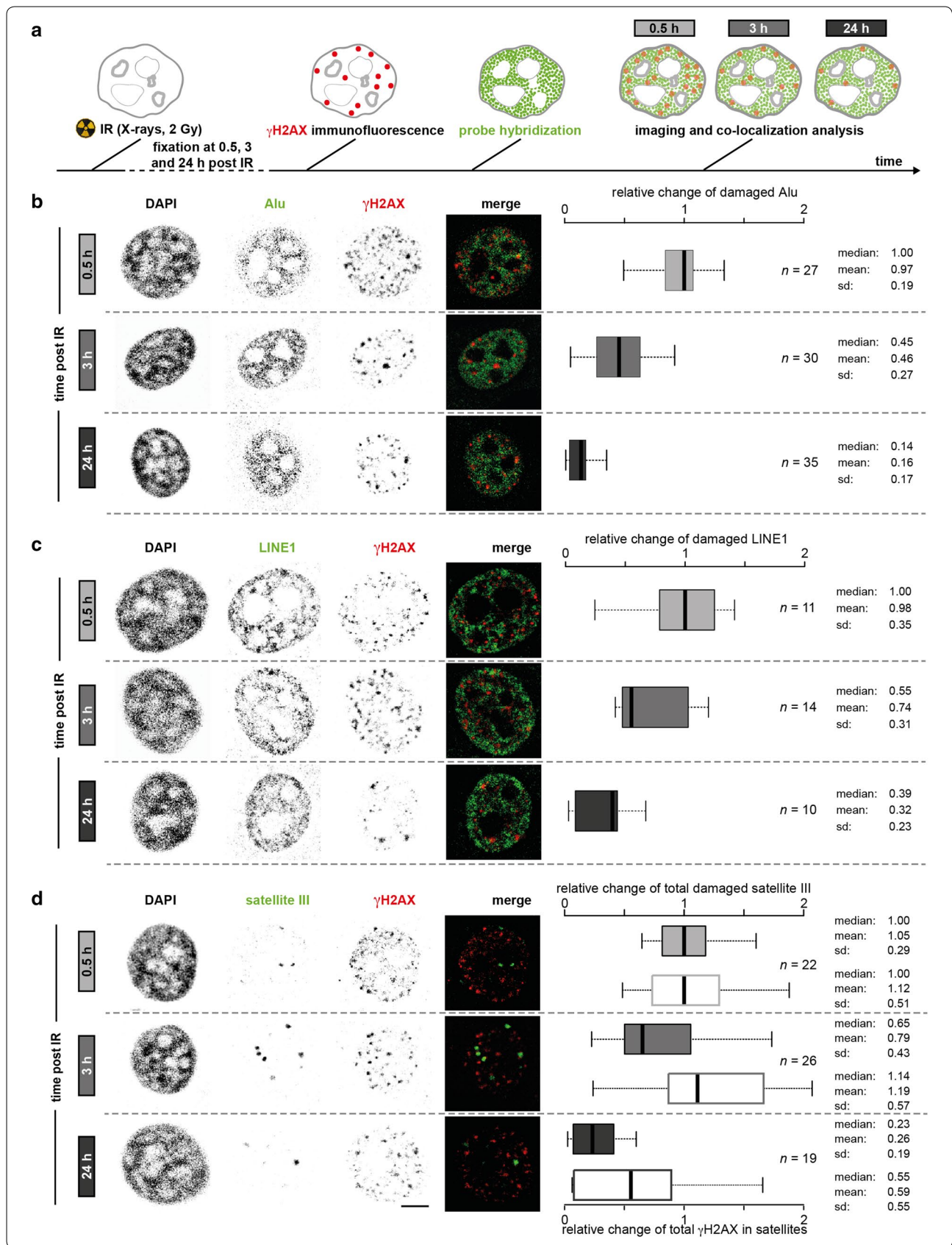
#### Discussion

Taken together, repetitive elements appear to be well integrated into chromatin and are preserved by DNA replication and repair processes with the same fidelity as the rest of the genome. All three repetitive elements examined by immuno-FISH follow the general trend of early replication and repair of euchromatin and later replication and repair of heterochromatin. Such observation is consistent with the trend that euchromatin is repaired faster than heterochromatin [40].

Previous studies employing genome-wide approaches showed a correlation between early replicating regions and enrichment for the interspersed Alu elements, while

(See figure on next page.)

**Fig. 4** DNA repair kinetics of repetitive DNA elements assessed by FISH. **a** Schematics of the experiment. HeLa cells were sham-irradiated or irradiated with 2 Gy X-rays and incubated for 0.5, 3 and 24 h.  $\gamma$ H2AX immunofluorescence and probe hybridization were performed before confocal micrographs acquisition and deconvolution. **b** (left) Representative confocal micrographs of HeLa cells depicting the DAPI, Alu elements and EdU as inverted gray channels, at the three time points post-IR. Merge is shown in pseudo-colors. Scale bar: 5  $\mu$ m. (right) Relative change of Alu fraction in  $\gamma$ H2AX foci. Data are normalized to the median of the 0.5 h time point. Boxes represent median, 2nd and 3rd quartile. Whiskers indicate three times the interquartile distance. Data are from three independent experiments. *n* combined total number of cells analyzed. *sd* standard deviation. **c, d** Represent the same as in **b** for LINE1 and satellite III, respectively. In **d**, the empty boxes represent the relative change of  $\gamma$ H2AX intensity in segmented satellite III regions



mid and late replicating regions were enriched in L1 [41]. Our combined genomics and immuno-FISH analyses further refines this conclusion in that L1 elements are found to be enriched from early throughout mid S-phase and, to a lesser degree, late S-phase. Recent findings indicate that L1 elements can modulate replication timing of mammalian chromosomes [42]. In fact, we also find that the presence of L1 elements—and possibly their transcriptional context—perturbs the replication program. Furthermore, we extended the analysis to tandem satellite repeats and showed that they are replicated in mid-to-late S-phase. The megabase-long tandem satellite repeats have been shown to contain replication initiation sites allowing their replication, which would be challenging from single forks emanating from the flanking non-repeated DNA regions [43]. As late replicating/repairing (peri-)centromeric genomic regions are rich in tandem repeats (in the order of hundreds of kilobases), these are not well represented in genome-wide studies. This may lead to a general lack of information on this last stage of the replication/repair processes and, thus, mislabeling as late replicating/repairing regions that are instead mid replicating/repairing. Therefore, immuno-FISH data and genome-wide studies complement each other.

Inverted repeats may form stem-loop structures that are often acknowledged to mediate genome instability through excision of the repeat-associated regions [44]. The same is the case for tandem repeats from satellites, where this chromatin mark and condensed structure has been proposed to play a role in avoiding spurious recombination events as discussed below.

We found that tandem repeats diverged from the global genome repair kinetics and were not repaired until late time points and were replicated in late S-phase. The minor overlap of  $\gamma$ H2AX with pericentromeric satellite DNA raised the question if the non-canonical histone H2AX is at all located at these regions. Our correlation analysis indicated that the H2AX signal was not correlated with satellite III, leaving the question open. Nonetheless, previous studies showed how  $\gamma$ H2AX signal was “bent” around the heterochromatic regions in human and mouse cells upon irradiation with accelerated charged particles [45], thus relocating the chromatin outside of the original lesion site, toward the interface between heterochromatin and euchromatin. This may explain the low colocalization of  $\gamma$ H2AX and satellite DNA signals observed under our experimental conditions. Due to their abundance, satellites may be erroneously utilized as repair templates during homologous recombination. The relocation of the damaged DNA outside of condensed heterochromatin has been proposed to avoid the utilization of the wrong chromosome as a template [46].

Altogether, the present study reemphasizes the need for complementary approaches (such as ChIP-Seq and immuno-FISH) to achieve an integrated and comprehensive investigation of any genomic processes.

## Conclusions

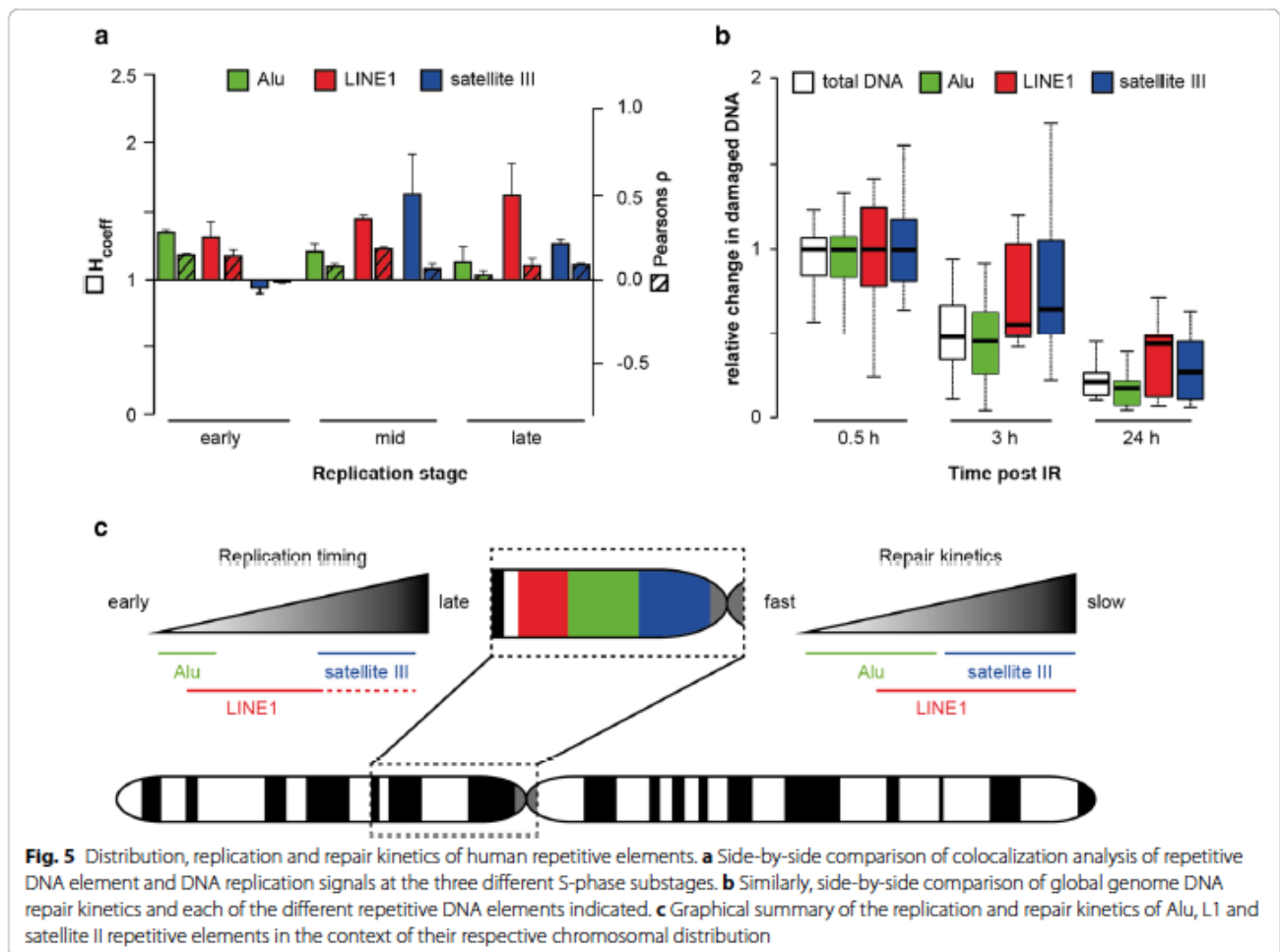
The aim of this work was to gain insight into how chromatin and its structural organization influences the genome maintenance processes of DNA replication and repair of repetitive elements. We employed an immuno-FISH approach to simultaneously label replication/repair and three different repetitive DNA elements. We were able to show that (1) the euchromatic Alu element is replicated during early S-phase; (2) L1, which is associated with AT-rich genomic regions, is replicated throughout S-phase, according to the repeat’s particular histone marks; (3) satellite III, which constitutes pericentromeric heterochromatin, is replicated exclusively at the mid-to-late S-phase. These data are summarized in Fig. 5a, c. As for the DNA repair process, we observe that Alu elements are repaired similarly to the total DNA, as observed by the concomitant decrease in the  $\gamma$ H2AX signal in Alu chromatin. Furthermore, this mirrors the global genome repair kinetics (Fig. 5b white boxes). Differently, satellite III and L1 elements showed slower repair kinetics, as their  $\gamma$ H2AX mark was retained longer (Fig. 5b, c). While the  $\gamma$ H2AX response in L1, Alu and satellite elements follows the corresponding GC equivalent of the total genome, this was not the case for LTR, indicating that their  $\gamma$ H2AX response may be further affected by other factors.

Based on the repeat’s specific histone marks, we conclude that the histone modifications in the specific repeat element predominantly determine its replication and repair timing. Thus, Alu elements, which are characterized by euchromatic chromatin features, are repaired and replicated the earliest, followed by LINE-1 elements, including more variegated eu/heterochromatic features and, lastly, satellite tandem repeats, which are homogeneously characterized by heterochromatic features and extend over megabase-long genomic regions.

## Methods

### Cell culture and exposure to ionizing radiation

C2C12 mouse myoblasts (ATCC CRL-1772) and HeLa cells (ATCC CCL-2) were grown at 37 °C and 5% CO<sub>2</sub>, in Dulbecco’s modified Eagle’s medium supplemented with 50  $\mu$ g/mL gentamycin, 20 mM L-glutamine and 10 or 20% fetal calf serum for HeLa and C2C12, respectively. For microscopy-based experiments, cells were grown on glass coverslips. Irradiation was performed with an ISO-VOLT Titan E X-ray machine (GE). Cells were exposed to doses of 2–10 Gy (90 kV, 33.7 mA).



### Chromatin immunoprecipitation

HeLa cells were fixed with 1% formaldehyde for 10 min at room temperature, and the crosslink was quenched with 125 mM glycine (5 min at room temperature). Nuclei were isolated after mild lysis in hypotonic buffer (10 mM HEPES pH 8, 1.5 mM MgCl<sub>2</sub>, 60 mM KCl) and 20 strokes in a tight dounce homogenizer. Chromatin was sheared in sonication buffer (0.5% SDS 10 mM EDTA, 50 mM Tris-HCl pH 8.1). Fragmentation of chromatin was carried out by ultrasound treatment (Bioruptor UCD200) so that fragments of 200–300 bp length were obtained. Chromatin from 1–2 × 10<sup>6</sup> cells was immunoprecipitated with 3 μg mouse anti-γH2AX (Clone JBW301, Upstate) or mouse anti-H2AX (Bethyl Laboratories, A300-83A) antibody. Chromatin was then incubated overnight at 4 °C with protein G-coated magnetic beads (ChIP-IT Express, Active Motif). The chromatin collected (ChIP sample) was then reverse-crosslinked in the presence of 200 mM NaCl at 65 °C for at least five hours, followed by RNase A (50 μg ml<sup>-1</sup>) treatment for 30 min at 37 °C and proteinase K (100 μg ml<sup>-1</sup>) treatment for 3 h at

50 °C. DNA elution was carried out in 1% SDS, 100 mM NaHCO<sub>3</sub>, in a rotary shaker at room temperature for 15 min. Pure DNA was isolated using the Qiagen PCR purification kit, and 15–30 ng of size-selected DNA fragments (Qubit fluorometric quantification) were used to produce ChIP-Seq libraries (Illumina ChIP-Seq DNA sample Prep Kit). Input sample was essentially prepared following the same protocol, but the immunoprecipitation step was skipped.

### Next-generation sequencing (NGS) and data analysis

γH2AX ChIP-Seq libraries were generated and processed as described in [33]. The corresponding datasets are from [33] and can be found at the Gene Expression Omnibus database (GEO accession number: GSE60526). Briefly, reads were mapped to the human genome (University of California, Santa Cruz (UCSC) hg19 assembly, based on the National Center for Biotechnology Information (NCBI) build 37.1) by means of SOAP2 software [47], allowing up to two mismatches for each 36 bp read. Data for genomic features were retrieved from publicly

available databases (UCSC/non-B DB) (Table 1). Accession numbers for histone modification ChIP-Seq data are given in Table 2.

Data that were originally generated in the hg18 assembly were transposed to hg19 using LiftOver (<http://genome.ucsc.edu/cgi-bin/hgLiftOver>). Reads per kilobase per million reads (RPKM) [50] were calculated for non-overlapping 10 kb genomic intervals for all sequence tracks. Correlation with  $\gamma$ H2AX, Repli-Seq data and genomic features was performed by Spearman's rho correlation coefficient, with  $P < 2.2 \times 10^{-16}$  in all cases.

Since the majority of reads containing repetitive elements cannot be mapped uniquely, they are usually underrepresented in NGS analysis. To measure the total  $\gamma$ H2AX signal mapped to Alu, LINEs, satellites and LTRs repetitive elements in an unbiased fashion, first, we mapped the quality filtered raw reads to human genome,

then the reads which uniquely mapped to the corresponding repetitive elements as annotated in RepeatMasker were counted into the corresponding repetitive elements. For the multiple mapped reads, if all mapped genomic positions were annotated as the same class of repetitive element, these reads were still counted into that single repetitive element. If multiple mapped positions were annotated as different type of repetitive elements, these reads were discarded, instead. Finally, the number of reads in each class was normalized to the total number of repetitive elements reads, containing only signatures of a single repetitive element type, so that the resulting fraction represents the genome-wide  $\gamma$ H2AX coverage in a given repetitive element, which we deemed "metarepetitive element." In ChIP-seq data analysis, which covers a minor proportion of the genome, the probability of reading the same sequence twice is higher than in whole-genome sequencing. Hence, de-duplication of PCR artifacts is less critical [51].

**Table 2 ENCODE datasets used for histone modifications and replication timing**

Cell line	Dataset	Data source [48, 49]
HeLa-S3	Repli-seq data	GSM923449
HeLa-S3	H3K4me1	GSM798322
HeLa-S3	H3K4me2	GSM733734
HeLa-S3	H3K4me3	GSM73368
HeLa-S3	H3K36me3	GSM733711
HeLa-S3	H3K79me2	GSM733669
HeLa-S3	H3K9ac	GSM733756
HeLa-S3	H3K27ac	GSM733684
HeLa-S3	H3K27me3	GSM733696
HeLa-S3	H3K9me3	GSM1003480
GM12878	Repli-seq data	GSM923451
GM12878	H3K4me1	GSM733772
GM12878	H3K4me2	GSM733769
GM12878	H3K4me3	GSM733708
GM12878	H3K36me3	GSM733679
GM12878	H3K79me2	GSM733736
GM12878	H3K9ac	GSM733677
GM12878	H3K27ac	GSM733771
GM12878	H3K27me3	GSM733758
GM12878	H3K9me3	GSM733664
HepG2	Repli-seq data	GSM923446
HepG2	H3K4me1	GSM798321
HepG2	H3K4me2	GSM733693
HepG2	H3K4me3	GSM733737
HepG2	H3K36me3	GSM733685
HepG2	H3K79me2	GSM733641
HepG2	H3K9ac	GSM733638
HepG2	H3K27ac	GSM733743
HepG2	H3K27me3	GSM733754
HepG2	H3K9me3	GSM1003519

#### Probe generation for fluorescence in situ hybridization (FISH)

Probes for Alu elements were generated by first amplifying Alu elements from HeLa genomic DNA (gDNA) via PCR using specific Alu primers (AluF: GGATTACAGGYRTGAGCCA; AluR: RCCAYTGCCTCCAGCCTG, [52]), followed by a labeling PCR with the same primers, biotin-labeled dUTP and the previous PCR product (diluted 1:50 in ddH<sub>2</sub>O) as DNA template. The product of the labeling PCR was then purified with QIAquick PCR purification kit (Qiagen). Mouse major satellite (MaSat) probes were generated by PCR using C2C12 gDNA and specific MaSat primers (MaSatF: AAAATGAGAAACATCCACTTG; MaSatR: CCATGATTTTCAGTTTTTC TT, [53]).

All PCRs and cycling conditions are listed in Tables 3 and 4.

Probes for L1 and for chromosome 1 specific satellite III were generated by nick-translation of plasmids containing the corresponding sequences: pLRE3-eGFP ([54] kind gift from John V. Moran) for LRE wild-type L1 element and pUC 1.77 [55] for chromosome 1 satellite III. Both probes were labeled with biotin-labeled dUTP via nick-translation. Conditions for nick-translation were as follows: 50 mM Tris-HCl pH 8, 5 mM MgCl<sub>2</sub>, 0.5 mg/mL BSA, 10 mM beta-mercapto-ethanol, 0.04 mM dUTP-biotin or dUTP-digoxigenin, 0.05 mM dATP/dGTP/dCTP each, 0.32 U DNase I (D5025, Sigma-Aldrich), 10 U Klenow fragment (M0210, NEB), 1  $\mu$ g plasmid DNA in a total volume of 100  $\mu$ L. The reaction was incubated for 90 min at 15 °C and stopped with 5  $\mu$ L 0.5 M EDTA.

**Table 3 PCR conditions for the generation of FISH probes**

PCRs	Alu (template)	Alu (labeling)	MaSat
PCR buffer*	1 ×	1 ×	1 ×
dATP/dGTP/dCTP	0.2 mM each	0.2 mM each	0.4 mM each
dTTP	0.2 mM	0.15 mM	–
dUTP-biotin	–	0.05 mM	0.08 mM
primer F/R	1 μM each	1 μM each	0.2 μM each
Taq polymerase	1.5 μL	1.5 μL	1 μL
DNA template	100 ng gDNA	1 μL 1:50 PCR product	100 ng gDNA
Final volume	to 50 μL	to 50 μL	to 50 μL

\* 10× PCR buffer: 100 mM Tris–HCl pH 8.3, 500 mM KCl, 15 mM MgCl<sub>2</sub>

**Table 4 PCR cycling conditions for the generation of FISH probes**

PCR cycling condition	Alu	MaSat
Initial denaturation	94 °C for 4 min	98 °C for 10 min
1. Denaturation	94 °C for 1 min	98 °C for 1 min
2. Annealing	57 °C/65 °C for 1 min*	56 °C for 1 min
3. Extension	72 °C for 3:30 min	72 °C for 2 min
No. of cycles steps 1.–3.	35	35
Final extension	72 °C for 10 min	72 °C for 5 min

\* Temperature for template PCR/temperature for labeling PCR

All probes were sheared with a Covaris S220 (Covaris Inc.) in microTUBEs (50–65 μL aliquots; 520,045, Covaris Inc.) to a final size of ~250 bp. The amount of probe required for FISH was then ethanol-precipitated in the presence of sodium acetate, washed with 70% ethanol, air-dried at room temperature, dissolved in hybridization solution: (1) with formamide (50–70% formamide, 2× SSC, 10% dextran sulfate, pH 7) for Alu and LINE1 FISH on metaphase spreads; (2) without formamide (10 mM Tris–HCl, 3 mM MgCl<sub>2</sub>, 50 mM KCl, 10 μg/mL gelatin, 2× SSC [56, 57] for satellite III and MaSat FISH on metaphase spreads and for interphase FISH. The amounts of probes were as follows: 250, 200 or 50 ng for L1, Alu and satellite III, respectively. 5 μL PCR were used for MaSat. For all probes (except the satellite III probe on interphase cells), 1 μg fish sperm DNA was added. For metaphase or interphase FISH, probes were dissolved in 30 and 15 μL hybridization solution, respectively. Probes were then denatured for five minutes at 80 °C.

#### Metaphase spreads preparation and FISH on metaphase chromosomes

HeLa and C2C12 cells were treated with 0.1 μg/mL colcemid for two hours. Cells were then harvested by

trypsinization and first incubated for 20 min with 75 mM KCl at room temperature. They were then fixed in dropwise added ice-cold methanol/acetic acid (3:1) for 30 min on ice. The fixation step was repeated twice. For chromosome spreading, the cell suspension was dropped onto a wet microscopy slide from a height of approximately 25 cm. The slide was then air-dried overnight. For metaphase FISH, the slides were rehydrated in ddH<sub>2</sub>O for 10 min, digested with 0.005% pepsin (165 U/mL, P6887, Sigma-Aldrich) in 0.01 M HCl for 10 min at 37 °C, then dehydrated in 70 and 100% ethanol for 5 min each. Finally, the slides were air-dried overnight.

Equilibration of metaphase spreads was performed with the respective hybridization solution (see above) at room temperature for 30 min. The solution was removed, and the probes were combined with the metaphase spreads in a humid chamber. Denaturation was performed at 70–80 °C in a water bath for 5 min and the hybridization followed at 37–42 °C overnight. Post-hybridization washing steps were done with 2× SSC and 0.1× SSC at 42 °C. Slides were blocked with 1% BSA/4× SSC for 30 min and the FISH probes detected with streptavidin Alexa Fluor 488 (S11223, Molecular Probes/Thermo Fisher Scientific, 1:800) or rabbit anti-digoxigenin (Cat#: 700772, Abfinity, 1:500) and anti-rabbit IgG Cy5 (711-175-152, Jackson ImmunoResearch, 1:400) in 1% BSA/4× SSC for 30 min. DNA counterstaining was performed with DAPI (1 μg/mL) for 10 min and the coverslips were mounted in Mowiol 4-88/2.5% DABCO.

#### Combination of replication staining (EdU Click reaction) or immunofluorescence staining of γH2AX with FISH

Cells were pulse-labeled with 10 μM EdU for 15 min or irradiated with 2 Gy X-rays. For replication staining, fixation with 3.7% formaldehyde/1× PBS followed directly after the pulse-labeling and for irradiated cells 0.5, 3 or 24 h post-IR. Cells were permeabilized and pre-denatured with 0.5% Triton X-100 in 1× PBS for 15 min, 0.1 M HCl for 15 min and 0.5% Triton X-100/1× PBS for 15 min.

EdU was detected with the EdU Click-594 ROTI kit (7776.1, Carl Roth), according to manufacturers' instructions. The dye azide was used in a final dilution of 1:2000.

For immuno-staining of γH2AX, irradiated cells were blocked with 1% BSA/1× PBS for 30 min, incubated with the primary antibody mouse anti-histone H2AX phospho-Ser139 (clone JBW301, 05-636, Upstate/Millipore, 1:200) in 1% BSA/1× PBS for 1 h and incubated with the secondary antibody donkey anti-mouse IgG Cy5 (715-175-150, Jackson ImmunoResearch, 1:250) in 1% BSA in 1× PBS for 1 h. Both stainings were post-fixed with 1%

formaldehyde/1 × PBS for 10 min before proceeding with FISH.

The cells were equilibrated with hybridization solution without formamide (composition as described above) at room temperature for 30 min. The solution was removed before combining the probes with the cells in a humid chamber; samples were denatured at 80 °C in a water bath for five minutes and hybridized overnight at 42 °C. Post-hybridization washing steps were done with 2 × SSC and 0.1 × SSC at 42 °C. FISH probes were detected with streptavidin Alexa Fluor 488 (S11223, Molecular Probes/Thermo Fisher Scientific, 1:800) in BSA/4 × SSC for 30 min. DNA counterstaining was performed with DAPI (1 µg/mL) for 10 min and the coverslips mounted in Mowiol 4-88/2.5% DABCO.

### Microscopy

Confocal imaging was performed using a Perkin Elmer VoX-1000 Spinning Disk microscope equipped with a 60 ×/1.4 NA/oil CFI Apochromat TIRF objective, four laser lines (405, 488, 561 and 635 nm) and a Hamamatsu EMCCD camera (C9100-50). The following filter sets were used: excitation: quad-bandpass 405/488/568/640 nm with the matching emission filters for DAPI (445/30 nm), Alexa Fluor 488 (500–548 nm), TRITC (526–623 nm) and Cy5 (664–750 nm). For higher special resolution, images were acquired using a Leica SP5 II laser scanning microscope using a 100 × 1.44 NA HCX PL APO Objective with a pixel size of 86.6 nm and a z-spacing of 125 nm for subsequent deconvolution. For imaging the 405, 488, 561 and 633 nm laser line and spectral detection windows of 425–465 nm (DAPI), 495–558 nm (Alexa 488), 600–660 nm (Alexa 594) and 640–705 nm (Cy5) were used. Images were then deconvolved with wavelength specific point spread functions using ImageJ and the Iterative Deconvolution 3D plugin [58]. In addition, a Zeiss Axiovert 200 with a 100 ×/1.4 NA/oil Plan-Apochromat objective was used to image metaphase spreads. Images were recorded using a Zeiss Axio-cam mRM, and the following filters were used: DAPI; ex: 350/50 nm; bs: 400 nm; em: 460/50 nm and Alexa Fluor 488: ex: 482/18 nm; bs: 495 nm; 520/28 nm.

### Image analysis for repair kinetics of repetitive elements

Image analysis was performed using the image analysis software Perkin Elmer Volocity 6.3. The following steps in the measurements tab were used to segment γH2AX foci and satellite regions (Additional file 1: Fig. S2): “Find object” (“nucleus”) using the DAPI channel (method “automatic,” minimum object size: 400 µm<sup>3</sup>), fill holes in object, dilate with two numbers of iterations, fill holes in object. “Find object” (“repair foci”) using the Cy5

channel, method “SD” (lower limit: set to “optimal value for all cells within one condition,” minimum object size: 0.3 µm<sup>3</sup>), remove noise from objects with fine filter, separate touching objects (object size guide: 1 µm, filter population: volume > 0.3 µm<sup>3</sup>), exclude “satellite” not touching “nucleus.”

### Colocalization analysis for replication timing of repetitive elements

To analyze at which S-phase stage any given repetitive elements are replicated, the cells were categorized into early, mid, or late replication patterns based on EdU signal [35]. The degree of colocalization was scored by the Pearson’s correlation coefficient and the  $H_{\text{coefficient}}$  [34]. First, a nuclear mask was derived from the DAPI channel using ImageJ (Gaussian blur with sigma = 1). Then, a local mean filter was applied (using the platform for image analysis Priithon) to the channels that are to be compared. This removes the background. Next, the  $H_{\text{coefficient}}$  and the Pearson’s coefficient  $r$  were calculated for each plane. For the plots, a mid-nuclear section was selected from each image as having the best signal quality. The method is schematically summarized in Additional file 1: Fig. S2.

### Additional file

**Additional file 1: Figure S1.** Genomic DNA repetitive elements, DNA replication and histone modifications distributions. **Figure S2.** Genome-wide correlation of DNA replication and histone modifications distributions in multiple cell lines. **Figure S3.** FISH probes and correlation analysis validation. **Figure S4.** Image analysis flowchart. **Figure S5.** Replication timing of murine major satellite DNA elements by FISH and S-phase sub-stages classification. **Figure S6.** Genome-wide correlation of DNA repetitive elements and histone γH2AX in HeLa cells. **Figure S7.** Genomic repetitive and non-B DNA elements, and γH2AX histone distributions. **Figure S8.** Relation of repetitive DNA elements to non-B DNA elements. **Figure S9.** Correlation of histone H2AX and repetitive DNA elements before and during the DDR by FISH. **Figure S10.** Complete DNA repair kinetics of repetitive DNA elements analyzed by FISH. **Figure S11.** DNA repair kinetics of murine major satellite DNA elements analyzed by FISH.

### Abbreviations

ChIP-Seq: chromatin immunoprecipitation-massively parallel DNA sequencing; DAPI: 4',6-diamidino-2-phenylindole; DDR: DNA damage response; FISH: fluorescence in situ hybridization; GEO: Gene Expression Omnibus Database; HR: homologous recombination; IR: ionizing radiation; LINE: long interspersed nuclear element; MaSat: mouse major satellite; NGS: next-generation sequencing; PCR: polymerase chain reaction; RPKM: reads per kilobase per million reads; SINE: short interspersed nuclear element; UV: ultraviolet light.

### Authors’ contributions

FN and AR performed the ChIP-Seq experiment; FN, AR, and WY performed genomic data analysis; AS performed and analyzed FISH experiments; AR and CR performed image analysis. FN, AS, AR and MCC wrote the manuscript. MCC conceived the project. All authors read and approved the final manuscript.

**Author details**

<sup>1</sup> Department of Biology, Technische Universität Darmstadt, 64287 Darmstadt, Germany. <sup>2</sup> Present Address: Biology Unit, IRBM Science Park S. p. A., 80131 Naples, Italy. <sup>3</sup> Present Address: G5 Lymphocyte Development and Oncogenesis, Immunology Department, Pasteur Institute, 75724 Paris Cedex 15, France.

**Acknowledgements**

We thank Anne Lehmkuhl and Stephanie Meyer for excellent technical assistance and John V. Moran for the kind gift of L1-containing plasmid.

**Competing interests**

The author declares that she has no competing interests.

**Availability of data and materials**

The microscopy datasets generated and analyzed during the current study are available from the corresponding author on reasonable request. All genomics datasets are publicly available, and the accession numbers are listed in the methods.

**Consent for publication**

Not applicable.

**Ethics approvals and consent to participate**

Not applicable.

**Funding**

This work was supported by grants of the Deutsche Forschungsgemeinschaft (GRK1657/TP1B to MCC; DFG CA198/8-1 and 2 to MCC) and the Bundesministerium für Bildung und Forschung (BMBF 02NUK036D to AR). The funding bodies had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 15 May 2018 Accepted: 25 September 2018

Published online: 23 October 2018

**References**

- Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet.* 2007;8:272–85.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. Initial sequencing and analysis of the human genome. *Nature.* 2001;409:860–921.
- Mouse Genome Sequencing C, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature.* 2002;420:520–62.
- Beck CR, Garcia-Perez JL, Badge RM, Moran JV. LINE-1 elements in structural variation and disease. *Annu Rev Genom Hum Genet.* 2011;12:187–215.
- Feng Q, Moran JV, Kazazian HH Jr, Boeke JD. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell.* 1996;87:905–16.
- Khazina E, Truffault V, Buttner R, Schmidt S, Coles M, Weichenrieder O. Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. *Nat Struct Mol Biol.* 2011;18:1006–14.
- Martin SL, Branciforte D, Keller D, Bain DL. Trimeric structure for an essential protein in L1 retrotransposition. *Proc Natl Acad Sci U S A.* 2003;100:13815–20.
- Mathias SL, Scott AF, Kazazian HH Jr, Boeke JD, Gabriel A. Reverse transcriptase encoded by a human transposable element. *Science.* 1991;254:1808–10.
- Hohjoh H, Singer MF. Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *EMBO J.* 1996;15:630–9.
- Ostertag EM, Kazazian HH Jr. Biology of mammalian L1 retrotransposons. *Annu Rev Genet.* 2001;35:501–38.
- Korenberg JR, Rykowski MC. Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. *Cell.* 1988;53:391–400.
- Quentin Y. A master sequence related to a free left Alu monomer (FLAM) at the origin of the B1 family in rodent genomes. *Nucleic Acids Res.* 1994;22:2222–7.
- Deininger P. Alu elements: know the SINEs. *Genome Biol.* 2011;12:236.
- Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet.* 2003;35:41–8.
- Burwinkel B, Killmann MW. Unequal homologous recombination between LINE-1 elements as a mutational mechanism in human genetic disease. *J Mol Biol.* 1998;277:513–7.
- White TB, Morales ME, Deininger PL. Alu elements and DNA double-strand break repair. *Mob Genet Elements.* 2015;5:81–5.
- Hancks DC, Kazazian HH Jr. Roles for retrotransposon insertions in human disease. *Mob DNA.* 2016;7:9.
- Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette LJ 3rd, Lohr JG, Harris CC, Ding L, Wilson RK, et al. Landscape of somatic retrotransposition in human cancers. *Science.* 2012;337:967–71.
- Scott EC, Gardner EJ, Masood A, Chuang NT, Vertino PM, Devine SE. A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. *Genome Res.* 2016;26:745–55.
- Guenatri M, Bailly D, Maison C, Almouzni G. Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *J Cell Biol.* 2004;166:493–505.
- Jarmuz M, Glotzbach CD, Bailey KA, Bandyopadhyay R, Shaffer LG. The Evolution of satellite III DNA subfamilies among primates. *Am J Hum Genet.* 2007;80:495–501.
- Waye JS, Willard HF. Structure, organization, and sequence of alpha satellite DNA from human chromosome 17: evidence for evolution by unequal crossing-over and an ancestral pentamer repeat shared with the human X chromosome. *Mol Cell Biol.* 1986;6:3156–65.
- Wong AK, Rattner JB. Sequence organization and cytological localization of the minor satellite of mouse. *Nucleic Acids Res.* 1988;16:11645–61.
- Wu JC, Manuelidis L. Sequence definition and organization of a human repeated DNA. *J Mol Biol.* 1980;142:363–86.
- Jolly C, Metz A, Govin J, Vigneron M, Turner BM, Khochbin S, Vourc'h C. Stress-induced transcription of satellite III repeats. *J Cell Biol.* 2004;164:25–33.
- Saksouk N, Simboeck E, Dejardin J. Constitutive heterochromatin formation and transcription in mammals. *Epigenetics Chromatin.* 2015;8:3.
- Valgardsdottir R, Chiodi I, Giordano M, Rossi A, Bazzini S, Ghigna C, Riva S, Biamonti G. Transcription of Satellite III non-coding RNAs is a general stress response in human cells. *Nucleic Acids Res.* 2008;36:423–34.
- Bacolla A, Wells RD. Non-B DNA conformations, genomic rearrangements, and human disease. *J Biol Chem.* 2004;279:47411–4.
- Cer RZ, Bruce KH, Mudunuri US, Yi M, Volfovsky N, Luke BT, Bacolla A, Collins JR, Stephens RM. Non-B DB: a database of predicted non-B DNA-forming motifs in mammalian genomes. *Nucleic Acids Res.* 2011;39:D383–91.
- RepeatMasker Open-4.0 [http://www.repeatmasker.org/].
- Hansen RS, Thomas S, Sandstrom R, Canfield TK, Thurman RE, Weaver M, Dorschner MO, Gartler SM, Stamatoyannopoulos JA. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc Natl Acad Sci U S A.* 2010;107:139–44.
- Chagin VO, Stear JH, Cardoso MC. Organization of DNA replication. *Cold Spring Harb Perspect Biol.* 2010;2:a000737.
- Natale F, Rapp A, Yu W, Maiser A, Harz H, Scholl A, Grulich S, Anton T, Horl D, Chen W, et al. Identification of the elementary structural units of the DNA damage response. *Nat Commun.* 2017;8:15760.
- Herce HD, Casas-Delucchi CS, Cardoso MC. New image colocalization coefficient for fluorescence microscopy to quantify (bio-)molecular interactions. *J Microsc.* 2013;249:184–94.
- Chagin VO, Casas-Delucchi CS, Reinhart M, Schermelleh L, Markaki Y, Maiser A, Bolius JJ, Bensimon A, Fillies M, Domaing P, et al. 4D Visualization of replication foci in mammalian cells corresponding to individual replicons. *Nat Commun.* 2016;7:11231.
- Casas-Delucchi CS, van Bommel JG, Haase S, Herce HD, Nowak D, Meilinger D, Stear JH, Leonhardt H, Cardoso MC. Histone hypoacetylation

- is required to maintain late replication timing of constitutive heterochromatin. *Nucleic Acids Res.* 2012;40:159–69.
37. Jackson DA, Pombo A. Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. *J Cell Biol.* 1998;140:1285–95.
  38. Jansen A, van der Zande E, Meert W, Fink GR, Verstrepen KJ. Distal chromatin structure influences local nucleosome positions and gene expression. *Nucleic Acids Res.* 2012;40:3870–85.
  39. Koo HS, Wu HM, Crothers DM. DNA bending at adenine-thymine tracts. *Nature.* 1986;320:501–6.
  40. Goodarzi AA, Noon AT, Deckbar D, Ziv Y, Shiloh Y, Lobrich M, Jeggo PA. ATM signaling facilitates repair of DNA double-strand breaks associated with heterochromatin. *Mol Cell.* 2008;31:167–77.
  41. Woodfine K, Fiegler H, Beare DM, Collins JE, McCann OT, Young BD, Debernardi S, Mott R, Dunham I, Carter NP. Replication timing of the human genome. *Hum Mol Genet.* 2004;13:191–202.
  42. Platt EJ, Smith L, Thayer MJ. L1 retrotransposon antisense RNA within ASAR lncRNAs controls chromosome-wide replication timing. *J Cell Biol.* 2018;217:541–53.
  43. Erliandri I, Fu H, Nakano M, Kim JH, Miga KH, Liskovych M, Earnshaw WC, Masumoto H, Kouprina N, Aladjem MI, Larionov V. Replication of alpha-satellite DNA arrays in endogenous human centromeric regions and in human artificial chromosome. *Nucleic Acids Res.* 2014;42:11502–16.
  44. Gordenin DA, Lobachev KS, Degtyareva NP, Malkova AL, Perkins E, Resnick MA. Inverted DNA repeats: a source of eukaryotic genomic instability. *Mol Cell Biol.* 1993;13:5315–22.
  45. Jakob B, Splinter J, Conrad S, Voss KO, Zink D, Durante M, Lobrich M, Taucher-Scholz G. DNA double-strand breaks in heterochromatin elicit fast repair protein recruitment, histone H2AX phosphorylation and relocation to euchromatin. *Nucleic Acids Res.* 2011;39:6489–99.
  46. Tsouroula K, Furst A, Rogier M, Heyer V, Maglott-Roth A, Ferrand A, Reina-San-Martin B, Soutoglou E. Temporal and spatial uncoupling of DNA double strand break repair pathways within mammalian heterochromatin. *Mol Cell.* 2016;63:293–305.
  47. Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics.* 2009;25:1966–7.
  48. Pope BD, Ryba T, Dileep V, Yue F, Wu W, Denas O, Vera DL, Wang Y, Hansen RS, Canfield TK, et al. Topologically associating domains are stable units of replication-timing regulation. *Nature.* 2014;515:402–5.
  49. Ernst J, Kheradpour P, Mikkelson TS, Shores N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature.* 2011;473:43–9.
  50. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 2008;5:621–8.
  51. Ebbert MT, Wadsworth ME, Staley LA, Hoyt KL, Pickett B, Miller J, Duce J. Alzheimer's Disease Neuroimaging I, Kauwe JS, Ridge PG: evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches. *BMC Bioinform.* 2016;17(Suppl 7):239.
  52. Liu P, Siciliano J, Seong D, Craig J, Zhao Y, de Jong PJ, Siciliano MJ. Dual Alu polymerase chain reaction primers and conditions for isolation of human chromosome painting probes from hybrid cells. *Cancer Genet Cytogenet.* 1993;65:93–9.
  53. Frauer C, Rottach A, Meilinger D, Bultmann S, Fellinger K, Hasenoder S, Wang M, Qin W, Soding J, Spada F, Leonhardt H. Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. *PLoS ONE.* 2011;6:e16627.
  54. Garcia-Perez JL, Morell M, Scheys JO, Kulpa DA, Morell S, Carter CC, Hammer GD, Collins KL, O'Shea KS, Menendez P, Moran JV. Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. *Nature.* 2010;466:769–73.
  55. Cooke HJ, Hindley J. Cloning of human satellite III DNA: different components are on different chromosomes. *Nucleic Acids Res.* 1979;6:3177–97.
  56. Celeda D, Aldinger K, Haar FM, Hausmann M, Durm M, Ludwig H, Cremer C. Rapid fluorescence in situ hybridization with repetitive DNA probes: quantification by digital image analysis. *Cytometry.* 1994;17:13–25.
  57. Celeda D, Bettag U, Cremer C. A simplified combination of DNA probe preparation and fluorescence in situ hybridization. *Z Naturforsch C.* 1992;47:739–47.
  58. Dougherty RP. Extensions of DAMAS and benefits and limitations of deconvolution in beamforming. In: AIAA 2005-2961; 2005.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

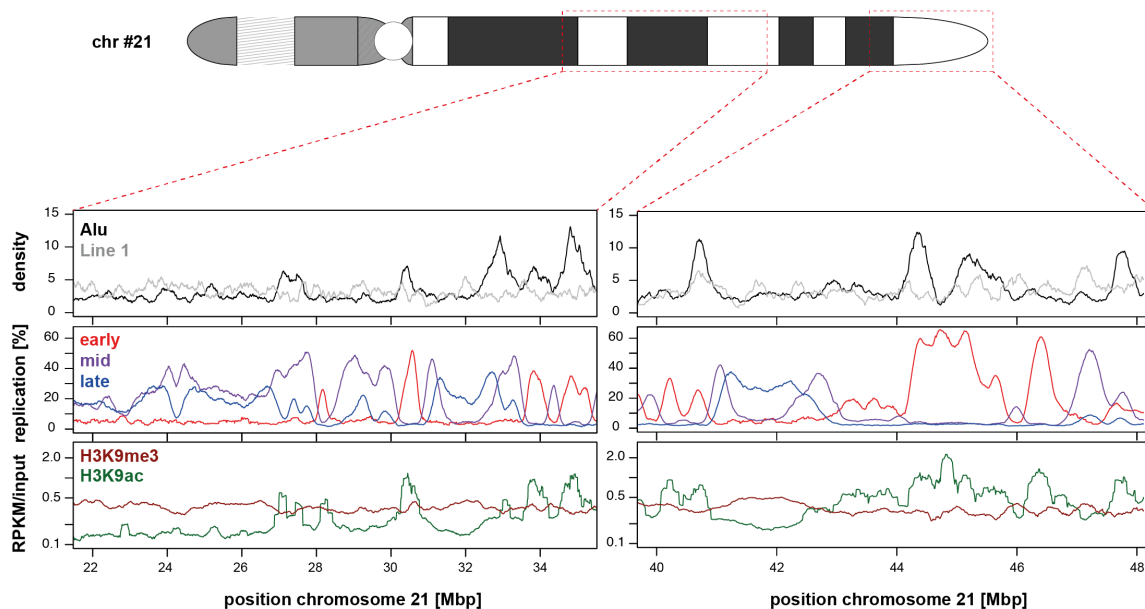
At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)



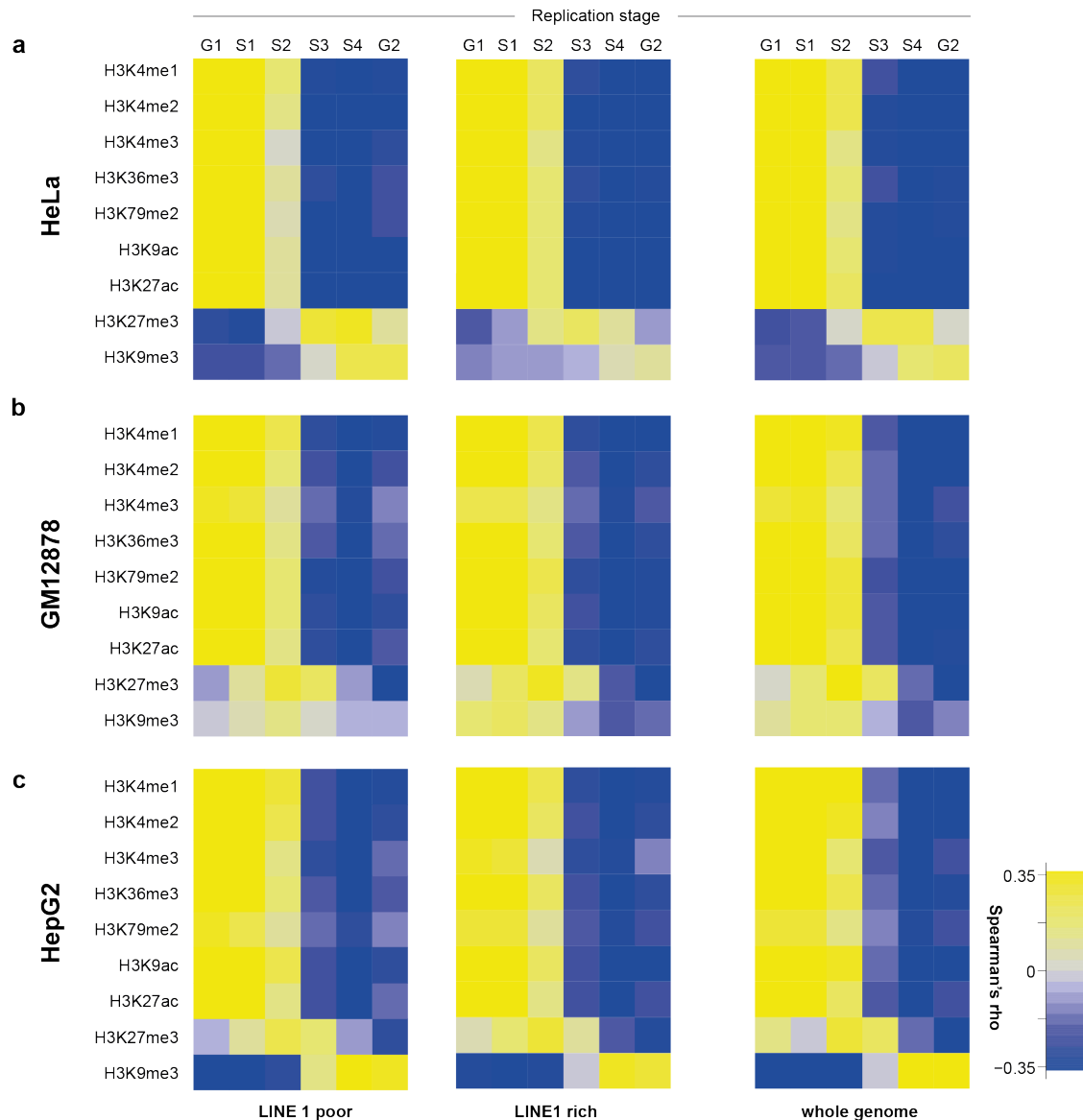
## SUPPLEMENTARY MATERIAL

Figure S1



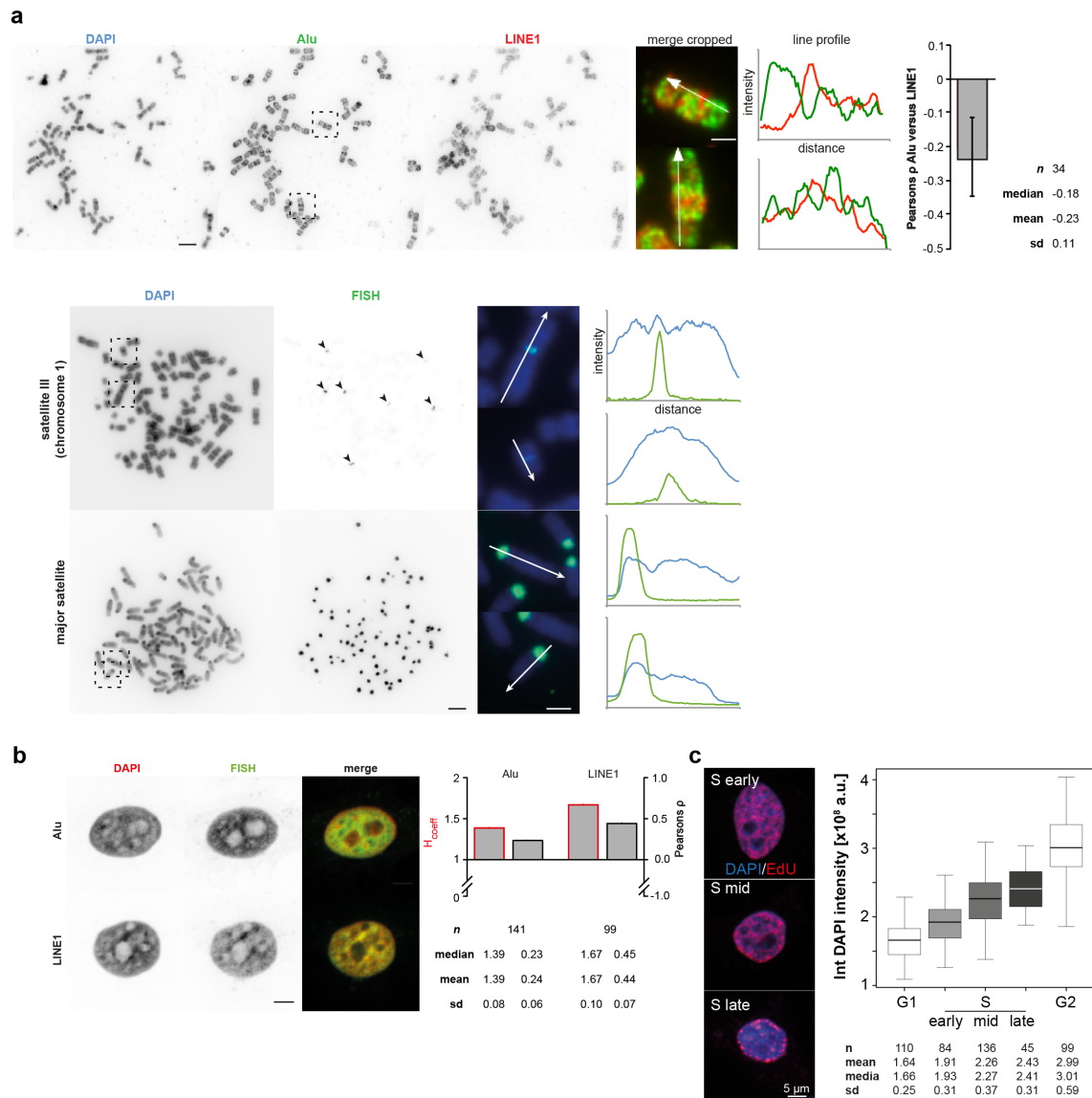
**Figure S1. Genomic DNA repetitive elements, DNA replication and histone modifications distributions.** Exemplary genomic (top) Alu and LINE1, (mid) Repli-Seq and (bottom) H3K9me3/ac distributions in 10 kbp intervals from HeLa cells. For Alu and LINE1 elements, the number of elements per interval is shown. All tracks were smoothed using a moving average of 10 intervals.

Figure S2



**Figure S2. Genome-wide correlation of DNA replication and histone modifications distributions in multiple cell lines.** Spearman's rho correlation matrix between replication stage (Repli-Seq) and histone modifications (ChIP-Seq) in uterine cervix cancer (HeLa, a), lymphoblastoid (GM12878, b) and hepatocarcinoma (HepG2, c) cell lines. Comparison between LINE1-poor (<1 count per genomic interval, left), LINE1-rich (>10 counts per genomic interval, middle) and whole-genome (right) are shown. For whole-genome correlation, data are from >290,000 genomic intervals with  $P < 2.2 \times 10^{-16}$ . For LINE1-rich/poor correlation matrices, data are from >10,000 genomic intervals with  $P < 10^{-9}$ .

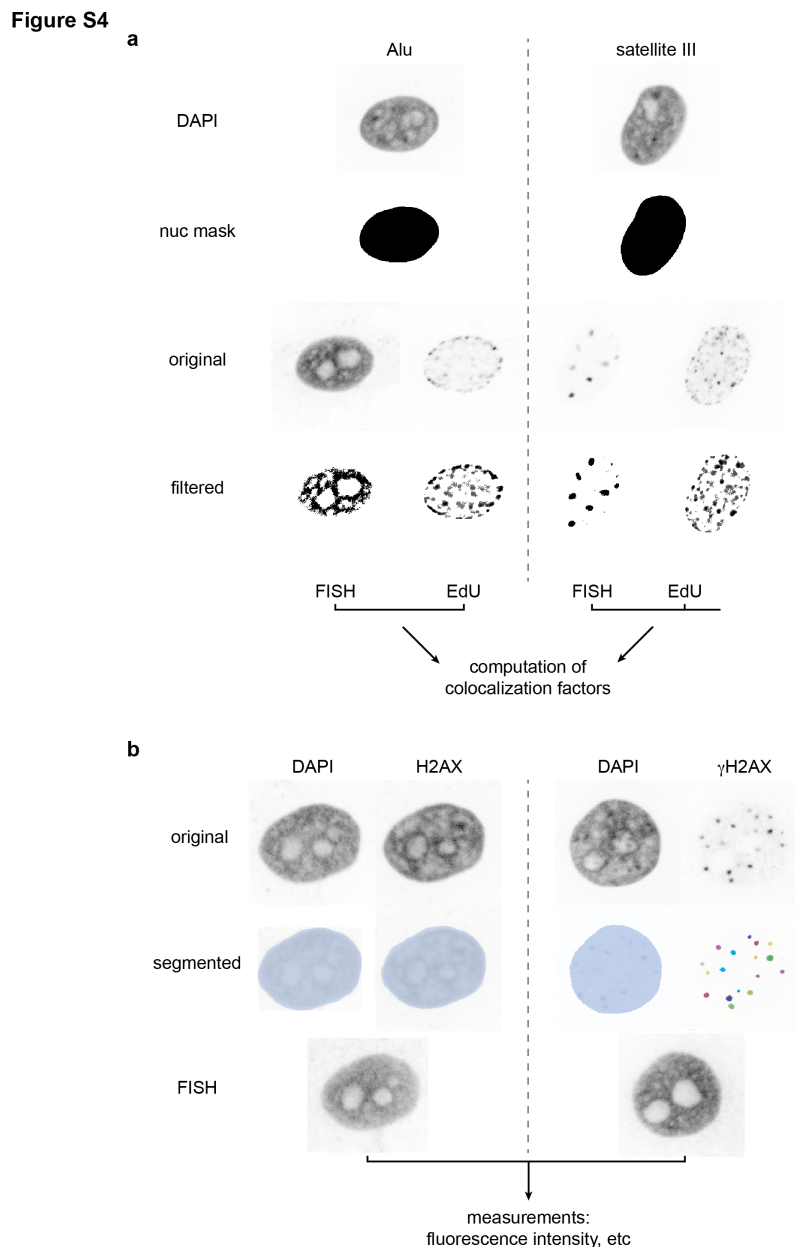
Figure S3



**Figure S3: FISH probes and correlation analysis validation.** (a) Double metaphase FISH for Alu and L1 and metaphase FISH for satellite III (chromosome 1) in HeLa cells and major satellite in C2C12 cells. Cell cultures were enriched for mitotic cells by colcemid treatment and after metaphase preparation spread onto glass slides. FISH for Alu and L1 was performed under stringent conditions (with formamide) whereas FISH for satellite III and major satellite was performed under non-stringent conditions (without formamide). Arrowheads point to the hybridized regions of satellite III FISH. Dashed frames are magnified in the merged cropped images. Arrows in the merged images show the position and direction of the line profiles (blue: DAPI; green and red: FISH). Pearson's correlation factor was calculated for 34 double hybridized chromosomes. Scale bar: 5  $\mu\text{m}$  and 2  $\mu\text{m}$  in

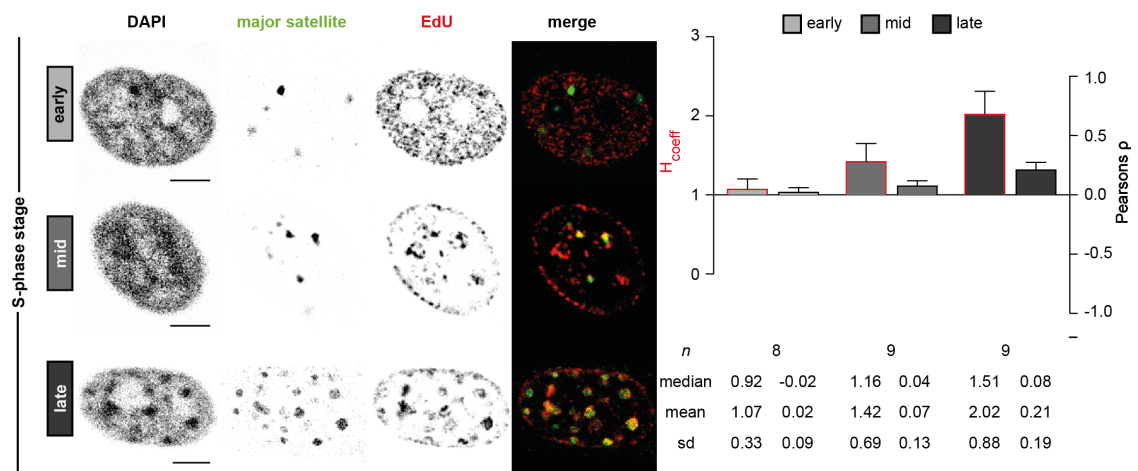
full and cropped images, respectively. (b) Colocalization of DAPI with Alu and L1 elements in HeLa cells. (left) Representative images of DAPI and Alu/L1 channels. Merged images are shown in pseudo-colors. Scale bar: 5  $\mu\text{m}$ . (right) Colocalization analysis of FISH and DAPI signal at the three different S-phase sub-stages using  $H_{\text{coefficient}}$  and Pearson's correlation coefficient as indicated. Error bars show the standard error of the mean. Data are from three independent experiments.  $n$ : combined total number of cells analyzed.  $sd$ : standard deviation. (c) Verification of S-phase classification using the integrated nuclear DAPI intensity. Left: exemplary images of early, mid and late S-phase cells classified based on immunofluorescence staining of replication patterns where EdU (active replication sites) is depicted in red and DAPI (DNA) in blue. The total nuclear DAPI intensity of cells that were thus classified into early, mid or late S-phase was measured and compared to EdU negative cells with small nuclei ( $<500 \mu\text{m}^3$ ) (G1 cells) and large nuclei ( $>750 \mu\text{m}^3$ ) (G2 cells).

---



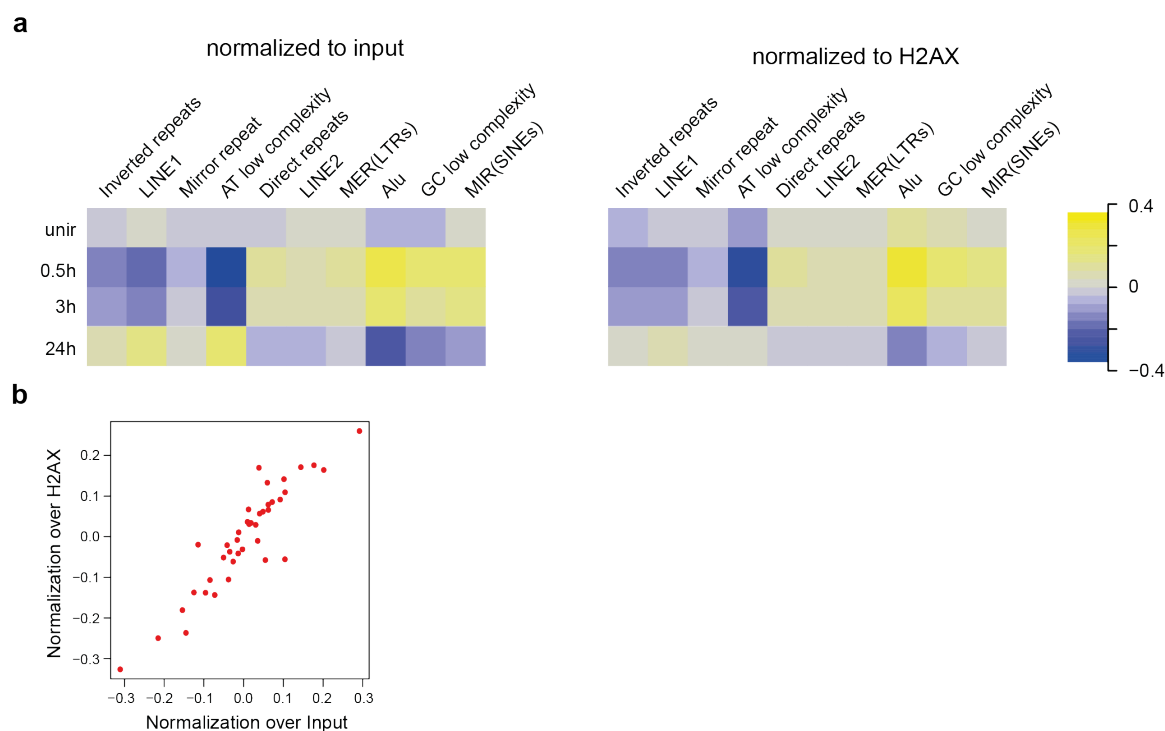
**Figure S4: Image analysis flowchart.** (a) Segmentation of DNA repetitive elements and colocalization analysis. First, the DAPI channel is used to generate the nuclear mask (nuc). The other channels are cropped to the “nuc mask” while maintaining the original intensity values. All intensity values outside the “nuc mask” are set to zero. Colocalization factors are then calculated from the filtered and masked images. (b) Segmentation of ( $\gamma$ )H2AX signal. The DAPI channel is used to segment the nucleus. Within the nucleus, H2AX fluorescence was used or  $\gamma$ H2AX foci were segmented (different colors refer to different objects). Objects outside the nucleus are excluded. Sum intensity values from all channels are then measured within the segmented focus/nucleus.

Figure S5



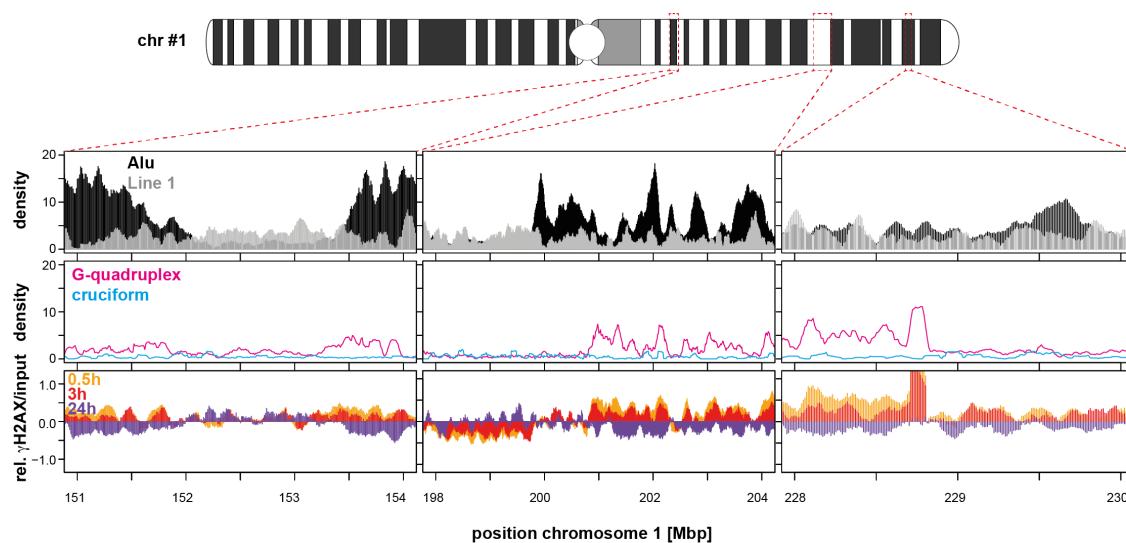
**Figure S5: Replication timing of murine major satellite DNA elements by FISH and S-phase sub-stages classification.** (left) Representative confocal and deconvolved micrographs of C2C12 cells depicting the DAPI, major satellite elements and EdU as inverted grey channels, at the three different S-phase sub-stages. Merge is shown in pseudo-colors. Scale bar: 5  $\mu$ m. (right) Colocalization analysis of FISH and EdU signal at the three different S-phase sub-stages by  $H_{\text{coefficient}}$  and Pearson's correlation coefficient as indicated. Error bars show the standard error of the mean. Data are from three independent experiments.  $n$ : combined total number of cells analyzed. sd: standard deviation.

Figure S6

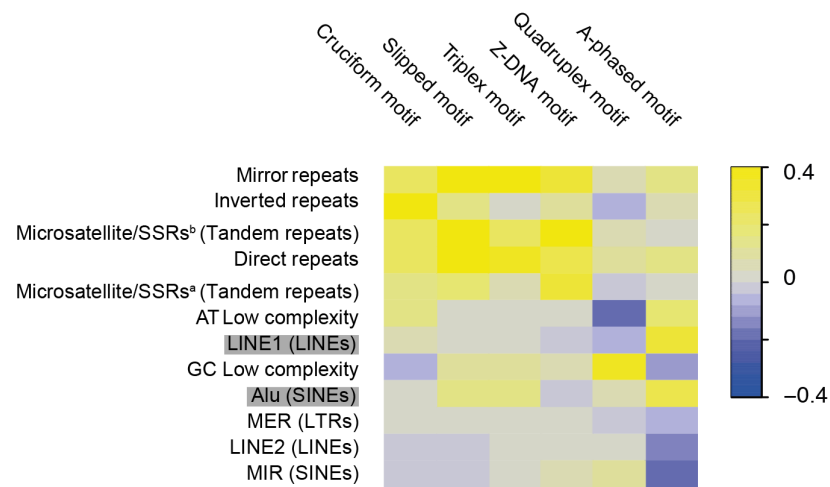


**Figure S6. Genome-wide correlation of DNA repetitive elements and histone  $\gamma$ H2AX in HeLa cells.** (a) Spearman's rho correlation matrices between DNA repetitive elements and  $\gamma$ H2AX abundance before and after ionizing radiation. The matrices have been produced by normalizing the  $\gamma$ H2AX CHIP sequencing abundance over the input (left) or the corresponding non-phosphorylated H2AX histone (right). Data are from >290,000 genomic intervals with  $P < 2.2 \times 10^{-16}$ . (b) Spearman's rho correlation coefficients from data normalized over H2AX as a function of their matched coefficient normalized over the input. Normalization over H2AX or over input shows only minor differences (Pearson's coefficient: 0.91).

Figure S7



**Figure S7. Genomic repetitive and non-B DNA elements, and  $\gamma$ H2AX histone distributions.** Exemplary genomic (top) Alu and LINE1, (mid) G-quadruplex and cruciform motifs and (bottom)  $\gamma$ H2AX distributions in 10 kbp intervals from HeLa cells. For the latter, the histone modification abundance at 0.5, 3 and 24 hours post ionizing radiation is shown. For Alu and LINE1 elements as well as for G-quadruplex and cruciform motifs, the number of elements per interval is shown. All tracks were smoothed using a moving average of 10 intervals.

**Figure S8**

**Figure S8: Relation of repetitive DNA elements to non-B DNA elements.** Spearman's rho correlation matrix. The number of each repetitive or non-B DNA element copies is computed in each 10 kb genomic interval. The correlation coefficient is then calculated. Data are from >290,000 genomic intervals. For each correlation,  $P < 2.2 \times 10^{-16}$ . Highlighted Alu and L1 repetitive elements are arbitrarily chosen to define chromatin compartments with opposing chromatin features, and are further investigated in FISH experiments.

Figure S9

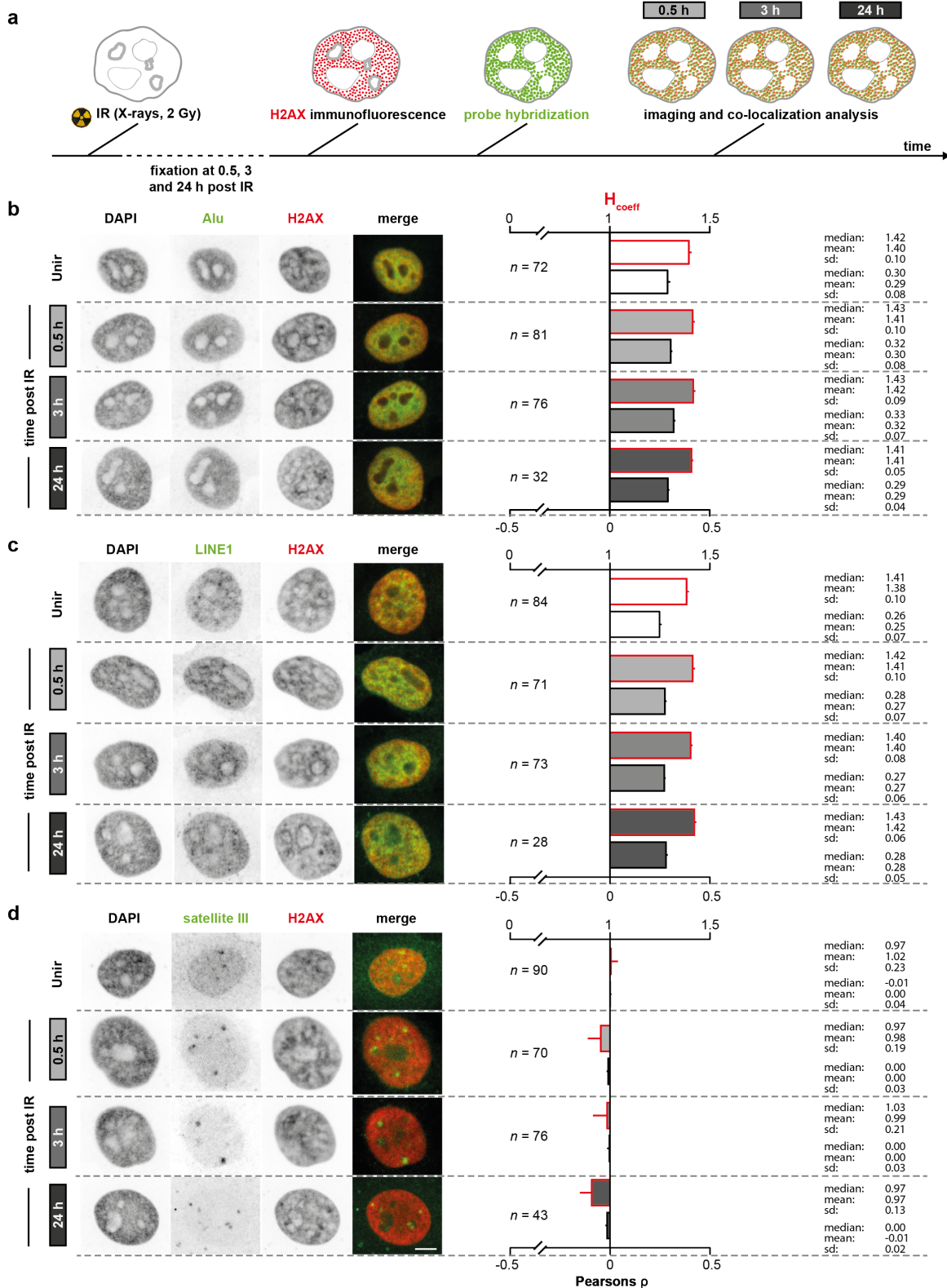
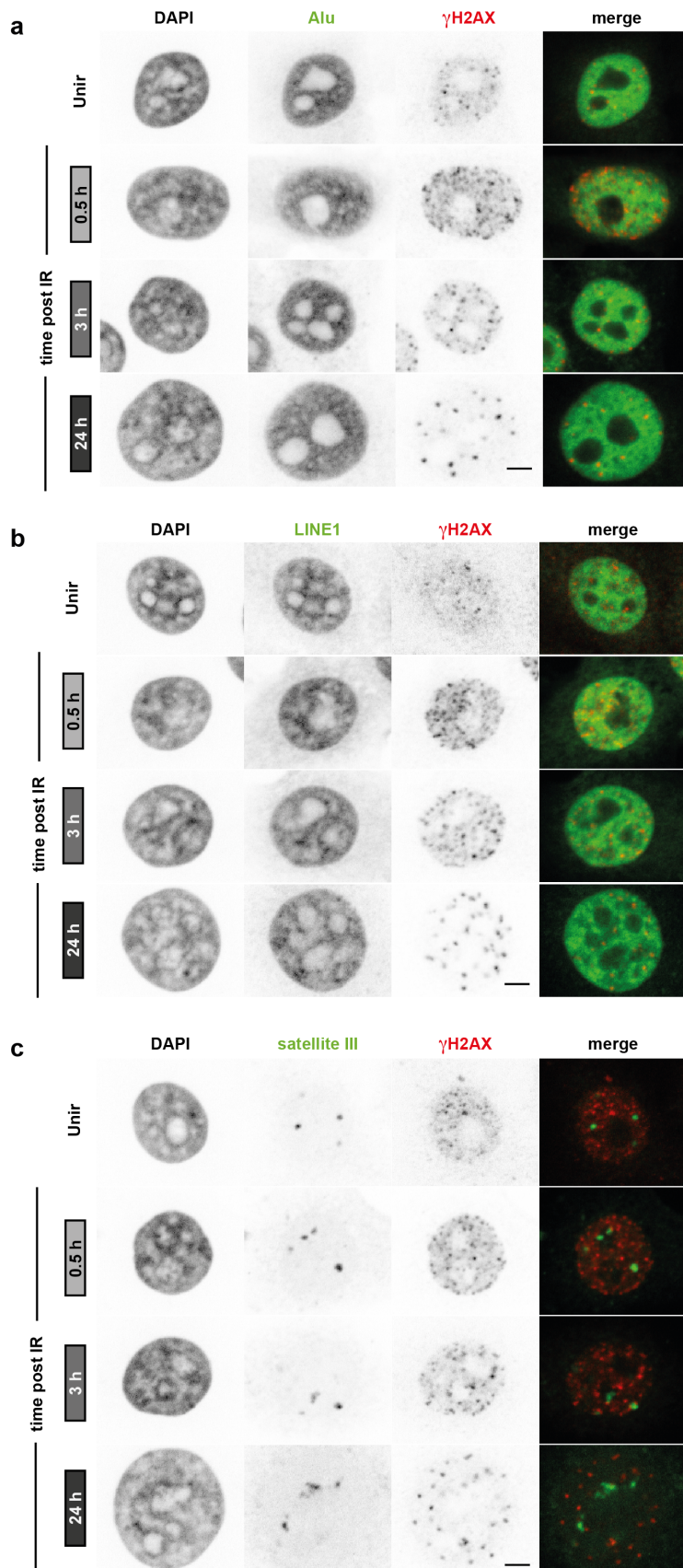


Figure S9: Correlation of histone H2AX and repetitive DNA elements before and during the DDR by FISH. (a) Schematics of the experiment. HeLa cells were sham-

irradiated or irradiated with 2 Gy X-rays and incubated for 0.5, 3 and 24 hours. H2AX immunofluorescence and probe hybridization were performed before confocal imaging. (b) (left) Representative confocal micrographs of HeLa cells depicting the DAPI, Alu elements and H2AX inverted grey channels, at the indicated times. Merge is shown in pseudo-colors. Scale bar: 5  $\mu\text{m}$ . (right) Colocalization analysis of FISH and H2AX signals via  $H_{\text{coefficient}}$  and Pearson's correlation coefficient as indicated. Error bars show the standard error of the mean. Data are from three independent experiments.  $n$ : combined total number of cells analyzed. sd: standard deviation. (c) and (d) represent the same as in (b) for L1 and satellite III, respectively.

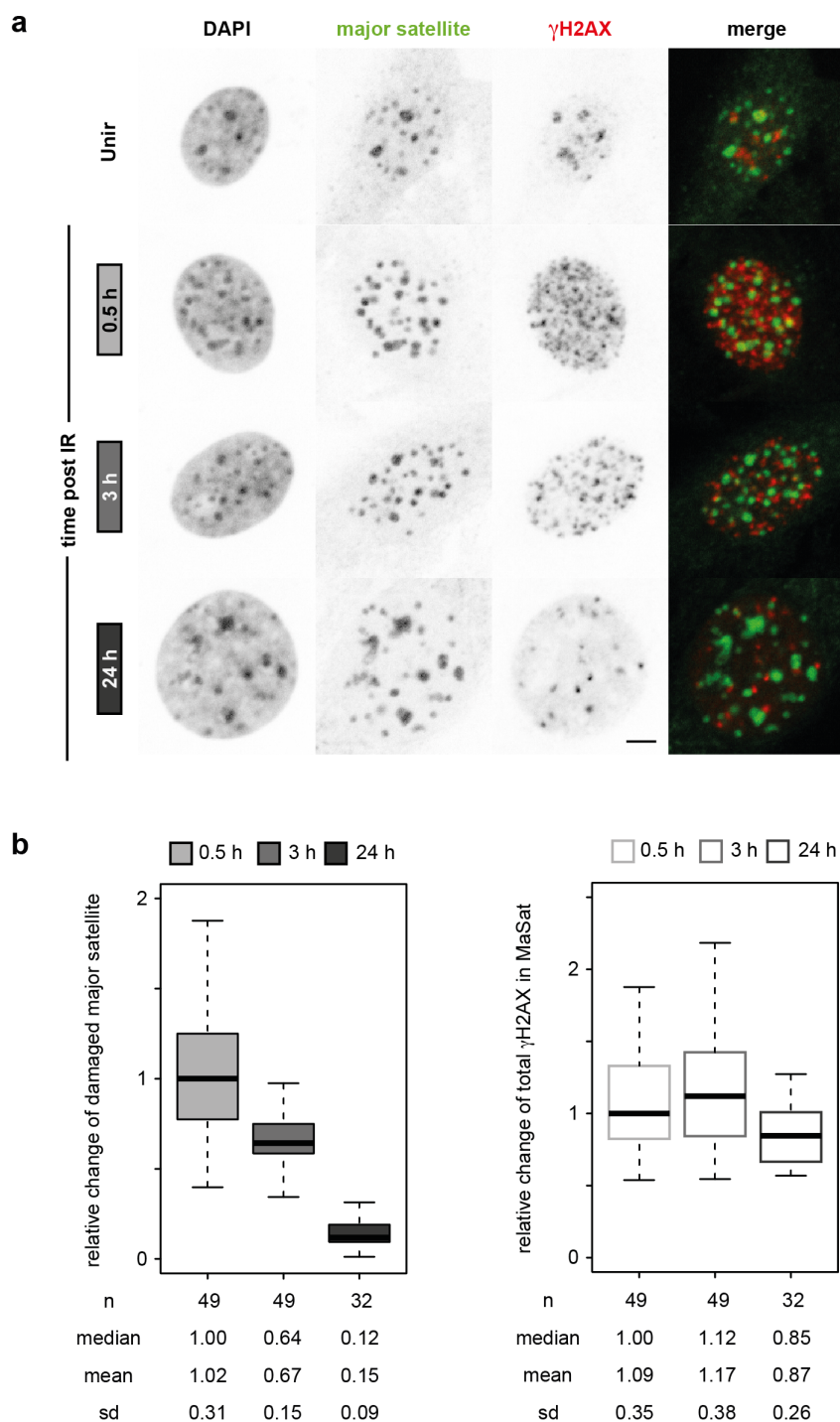
---

Figure S10



**Figure S10. Complete DNA repair kinetics of repetitive DNA elements analyzed by FISH.** (a) Representative confocal images of HeLa cells depicting the DAPI, Alu elements and  $\gamma$ H2AX as inverted grey channels, before irradiation (top) and at the three time-points post IR. Merge is also shown in pseudo-colors. Scale bar: 5  $\mu$ m. (b) and (c) represent the same as in (a) for L1 and satellite III, respectively.

---

**Figure S11**

**Figure S11: DNA repair kinetics of murine major satellite DNA elements analyzed by FISH.** (a) Representative confocal images of C2C12 mouse myoblast cells depicting the DAPI, major satellite elements and  $\gamma$ H2AX as inverted grey channels, before and at the three time-points post IR. Merge is shown in pseudo-colors. Scale bar: 5  $\mu$ m. (b) (left) Relative change of major satellite fraction in  $\gamma$ H2AX foci. (right) Relative change of  $\gamma$ H2AX

intensity in segmented major satellite regions. Data are normalized to the median of the 0.5 h time-point. Boxes represent median, 2<sup>nd</sup> and 3<sup>rd</sup> quartile. Whiskers are three times the interquartile distance. Data are from three independent experiments. *n*: combined total number of cells analyzed. sd: standard deviation.

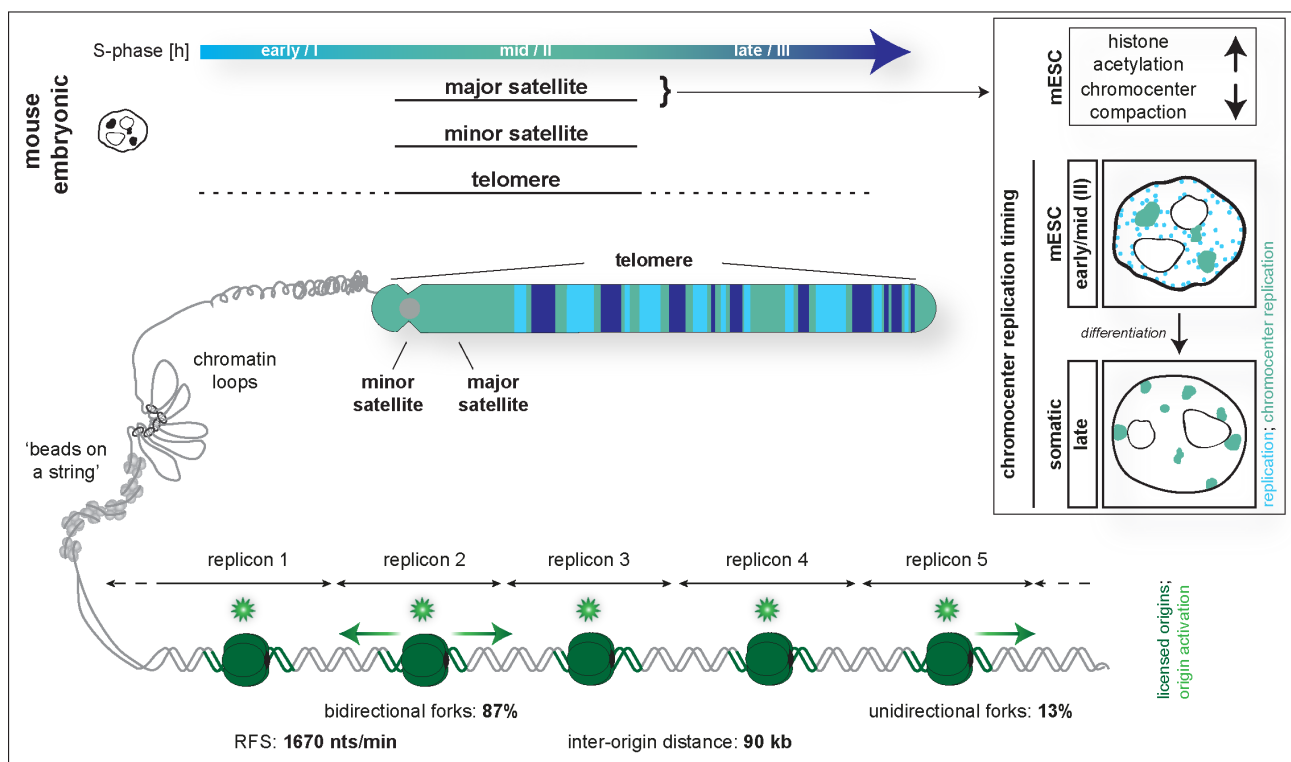
---



### 5.3 CHAPTER III: Developmental Differences in Genome Replication Program and Origin Activation

C. Rausch prepared all figures and tables and wrote the manuscript together with M. C. Cardoso. C. Rausch performed all experiments except for figures 3, 5B, 6, 7, S6, S11, S16 and S17. Experiments from figures 1, 2B-D, 5, S8C, S15 and S19 were performed in collaboration with P. Weber. C. Rausch analyzed all data except from figures 1E, 7D-E, S12, S13, S20, S21 and S22.

#### Graphical Abstract





Published online 2 December 2020

Nucleic Acids Research, 2020, Vol. 48, No. 22 12751–12777  
doi: 10.1093/nar/gkaa1124

# Developmental differences in genome replication program and origin activation

Cathia Rausch<sup>1,†</sup>, Patrick Weber<sup>1,†</sup>, Paulina Prorok<sup>1</sup>, David Hörl<sup>2</sup>, Andreas Maiser<sup>2</sup>, Anne Lehmkuhl<sup>1</sup>, Vadim O. Chagin<sup>1,3</sup>, Corella S. Casas-Delucchi<sup>1</sup>, Heinrich Leonhardt<sup>2</sup> and M. Cristina Cardoso<sup>1,\*</sup>

<sup>1</sup>Department of Biology, Technical University of Darmstadt, 64287 Darmstadt, Germany, <sup>2</sup>Department of Biology II, LMU Munich, 81377 Munich, Germany and <sup>3</sup>Institute of Cytology, Russian Academy of Sciences, St. Petersburg, Russia

Received June 30, 2020; Revised October 09, 2020; Editorial Decision November 01, 2020; Accepted November 04, 2020

## ABSTRACT

**To ensure error-free duplication of all (epi)genetic information once per cell cycle, DNA replication follows a cell type and developmental stage specific spatio-temporal program. Here, we analyze the spatio-temporal DNA replication progression in (un)differentiated mouse embryonic stem (mES) cells. Whereas telomeres replicate throughout S-phase, we observe mid S-phase replication of (peri)centromeric heterochromatin in mES cells, which switches to late S-phase replication upon differentiation. This replication timing reversal correlates with and depends on an increase in condensation and a decrease in acetylation of chromatin. We further find synchronous duplication of the Y chromosome, marking the end of S-phase, irrespectively of the pluripotency state. Using a combination of single-molecule and super-resolution microscopy, we measure molecular properties of the mES cell replicon, the number of replication foci active in parallel and their spatial clustering. We conclude that each replication nanofocus in mES cells corresponds to an individual replicon, with up to one quarter representing unidirectional forks. Furthermore, with molecular combing and genome-wide origin mapping analyses, we find that mES cells activate twice as many origins spaced at half the distance than somatic cells. Altogether, our results highlight fundamental developmental differences on progression of genome replication and origin activation in pluripotent cells.**

## INTRODUCTION

DNA replication, together with DNA transcription and repair, is a fundamental nuclear metabolic process. Complete and error-free genome duplication once every cell cycle is essential for genome integrity and maintenance. In eukaryotic cells, DNA replication can be subdivided in two main stages: recognition and subsequent licensing of origins of replication (ORIs) at the transition from mitosis (M-phase) to the gap 1 (G1) phase (1,2), and the activation of only a subset of these origins at the beginning of the synthesis (S) phase. The latter is eventually followed by the duplication of the (epi)genetic information by the multi-protein DNA synthesis complex (replisome) (3,4). After the initial unwinding of the DNA replication bubble at the origin of replication, the replisome ensures the semi-conservative duplication of the underlying DNA template (reviewed in (5)). Many features of DNA replication organization share high similarities between different species, including yeast, fruit flies, mice and humans (6–10), and homologues for the key factors involved have been identified in most of these species (3). Genome duplication follows a spatio-temporal program generally correlating with transcriptional activity, specific epigenetic marks and 3D genome architecture (11). Cytological methods relying on the detection of components of the replisome or nascent DNA via incorporation of modified nucleotides, allow the *in situ* visualization of newly synthesized DNA and sites of ongoing DNA replication (replication foci (RFi)). These replication foci form distinct spatial patterns, characterized by the dynamic intra-nuclear distribution of the replication signals during S-phase progression (5–7,12–14). In metazoan cells, three main patterns are observed, at early (I), mid (II) and late (III) S-phase, corresponding to the replication of euchromatin, facultative and constitutive heterochromatin, respectively (10,15,16).

\*To whom correspondence should be addressed. Tel: +49 6151 16 21882; Fax: +49 6151 16 21880; Email: cardoso@bio.tu-darmstadt.de

<sup>†</sup>The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Present address: Corella S. Casas-Delucchi, The Institute of Cancer Research | Chester Beatty Laboratories, London SW3 6JB, UK.

© The Author(s) 2020. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Fluorescence recovery after photobleaching (FRAP) experiments showed the *de novo* assembly of replisomes adjacent to previously activated sites (17), suggesting that, instead of persisting as permanent factories throughout S-phase (18), replisomes are activated in a ‘next in-line’ manner (domino model). Hence, complete DNA replication depends on the initial stochastic activation/spontaneous firing of a few origins with high firing probability within euchromatic regions of each chromosome and the subsequent domino-like activation of adjacent origins with decreasing firing probability during S-phase progression (19–22).

In early autoradiographic fiber studies of single DNA molecules it was observed that replicons resulted from individual initiation events at origins of replication, which are organized and activated in clusters of, on average, 1 Mb in size and consisting of 2–9 smaller replicons of 100–200 kb (6,23,24). DNA halo analysis showed that these replicon sizes are in good agreement with measured sizes of chromatin loops. Hence, loop structures, potentially mediated by cohesins or functionally related proteins (25,26), represent the DNA element that defines replicons as functional unit in the DNA replication context (reviewed in (27)). Labeling cells with modified nucleotides revealed that the replicon clusters observed on DNA fibers become visible as the before-mentioned replication foci in interphase nuclei (15). With higher optical resolution levels, the number of replication foci measured in cells increased and each replication nanofocus in somatic mammalian cells was shown to be equivalent to a replicon unit (28,29). Besides loop structures, chromatin signatures and the associated changes in chromatin structure and accessibility, influence licensing and activation of origins of replication and, thus, replication timing programs in mammalian cells (30). In that regard, major changes in DNA replication timing have been correlated with changes in histone acetylation levels, as histone hyperacetylation was shown to advance origin firing and DNA replication timing (31–34).

DNA replication studies in early developmental stages of *Drosophila* and *Xenopus* embryos revealed very rapid cell divisions with no gap phases and short S-phase duration (35–37). The latter is based on high origin activation levels, short inter-origin distances and concomitant differences in replicon sizes (36–38). While early developing mouse cells do not exhibit such fast cell division rates, they are characterized by short gap phases and the (almost complete) absence of transcription in the first zygotic cleavage stage (3,35–37,39–41). Additionally, specific spatial replication patterns already exist at the one-cell stage in mouse embryos (42) and replication programs of differentiating cells undergo large rearrangements during lineage commitment (43,44).

Here, we analyzed the replication dynamics in pluripotent mouse embryonic stem (mES) cells by characterizing the replication timing program and the replicon organization, and ultimately comparing it to known features of mouse somatic cells. We demonstrate that mES cells exhibit a distinct replication timing in comparison to differentiated and somatic cells, marked by early/mid replication of pericentromeric heterochromatin. We further show that this changes during mES cell differentiation when pericentromeric heterochromatin becomes late replicating. This major change correlates with differences

in chromatin compaction and histone acetylation levels and can be emulated by targeting histone deacetylases to pericentromeric heterochromatin. Analysis of the replication timing of (sub)chromosomal elements, revealed in addition a synchronous replication of the Y chromosome that concomitantly marks the end of S-phase in mES cells, as well as in differentiated and somatic cells. Using a combination of molecular and super-resolution imaging techniques, we characterized the mES cell replicon, the essential functional unit of DNA replication. We found by DNA combing and genome-wide origin profiling data analyses that mES cells activate more origins of replication compared to somatic cells resulting in shorter inter-origin distances, which in turn leads to smaller replicon sizes. Furthermore, in contrast to human somatic cells, duplication of the mES cell genome relies on a substantial amount of single (unidirectional) replication forks.

## MATERIALS AND METHODS

### Expression constructs

To generate the expression vector containing a fusion protein of eGFP and human HDAC1 (peGFP-hHDAC1, pc2447), hHDAC1 was amplified from human cDNA with the following primers containing BglII and Eco47III restriction enzyme sites: fw: 5'-AAAGATCTAGCAAGATGGCGCAGACGCAG-3' and rev: 5'-AAAGCGCTGGGCCAACTTGACCTCCTCC-3'. PCR amplicon and pCR2.1-TOPO vector (Thermo Fisher Scientific, Walham, MA, USA) were double digested with BglII and Eco47III and ligated. The final peGFP-hHDAC1 was generated by double digesting pCR2.1-TOPO-hHDAC1 with BglII and XhoI and hHDAC1 was inserted into the BglII/SalI digested backbone vector pEGFP-C1 (pc0592, Clontech Laboratories, Mountain View, CA, USA). SalI and XhoI form compatible ends that allow ligation. The final plasmid eGFP-hHDAC1 was verified by restriction enzyme digest, sequencing and enzymatic activity was analyzed via histone acetylation stainings (H4K5ac and H4K8ac, Figure 3B) in transfected cells.

All plasmid characteristics are summarized in Supplementary Table S1.

### Cell culture, differentiation and transfection

All cells used were mycoplasma free. J1 (45) and E14 (46) mouse embryonic stem cell lines were cultured in Dulbecco's modified Eagle's medium (DMEM) high glucose (Cat. No.: D6429, Sigma-Aldrich Chemie GmbH, Steinheim, Germany) supplemented with 15% fetal calf serum (FCS), 1× non-essential amino acids (Cat. No.: M7145, Sigma-Aldrich Chemie GmbH, Steinheim, Germany), 1× penicillin/streptomycin (Pen/Strep) (Cat. No.: P4333, Sigma-Aldrich Chemie GmbH, Steinheim, Germany), 1× L-glutamine (Cat. No.: G7513, Sigma-Aldrich Chemie GmbH, Steinheim, Germany), 0.1 mM beta-mercaptoethanol (Cat. No.: 4227, Carl Roth, Karlsruhe, Germany), 1000 U/ml recombinant mouse LIF (Millipore) and 2i (1 μM PD032591 and 3 μM CHIR99021 (Cat. Nos.: 1408 and 1386 respectively, Axon Medchem, Netherlands)) on gelatin-coated culture dishes (0.2% gelatin; Cat. No.:

G2500, Sigma-Aldrich Chemie GmbH, Steinheim, Germany). Culture medium was changed every day and cells were split every 2 days.

Mouse embryonic fibroblasts (MEF W8 (47)) and mouse myoblasts (C2C12 (48)) were cultured in DMEM high glucose supplemented with  $1 \times$  L-glutamine,  $1 \mu\text{M}$  gentamicin (Cat. No.: G1397, Sigma-Aldrich Chemie GmbH, Steinheim, Germany) and 10% and 20% FCS, respectively.

Mouse primary fibroblasts were isolated from ear tissue of adult mice (C57BL/6) sacrificed according to the animal care and use regulations (Government of Hesse, Germany). The tissue was shaken at  $37^\circ\text{C}$  with  $2 \text{ mg/ml}$  collagenase NB8 (Serva Electrophoresis, Heidelberg, Germany) in DMEM supplemented with  $1 \times$  Pen/Strep and vortexed repeatedly. Dissociated ear tissue was pipetted through a cell strainer (Cat. No.: 352235, Corning, NY, USA), single cells were plated on a gelatin-coated (0.2%) dish and cultivated in DMEM supplemented with 10% FCS and  $1 \mu\text{M}$  gentamicin at  $37^\circ\text{C}$  and 5%  $\text{CO}_2$ .

Differentiation of naïve pluripotent stem cells was performed as previously described (49) with the exception that the differentiation medium contained  $10 \mu\text{M}$  retinoic acid (Cat. No.: R2625, Sigma-Aldrich Chemie GmbH, Steinheim, Germany).

For live cell experiments, J1 mES cells were co-transfected with plasmids encoding for mRFP-tagged hPCNA (50) and MaSat-GFP (51) using AMAXA nucleofection (B016, Amaxa Nucleofactor II, Lonza Ltd., Basel, Switzerland) and plated on gelatin-coated p35 dishes containing a glass bottom that allowed for optical imaging.

For HDAC1 targeting to chromocenters, J1 mES cells were co-transfected with plasmids encoding for GBP-MaSat (52) and eGFP-hHDAC1/eGFP-C1 using AMAXA as described above. Twenty-four hours after transfection, cells were pulse chased and EdU, PCNA and histone modifications were detected as described below.

All cell line and plasmid characteristics are summarized in Supplementary Tables S1 and S2, respectively.

#### **DNA replication labeling and (immuno)fluorescent visualization**

For immunostaining experiments, cells were grown on glass coverslips coated with 0.2% gelatin.

To visualize and analyze progression of DNA replication, cells were labeled with  $10 \mu\text{M}$  5-ethynyl-2'-deoxyuridine (EdU) for 12 min, chased for varying times with medium supplemented with  $50 \mu\text{M}$  thymidine, fixed with 3.7% formaldehyde/ $1 \times$  phosphate-buffered saline (PBS) (Cat. No.: F8775, Sigma-Aldrich Chemie GmbH, Steinheim, Germany) for 10 min and permeabilized with 0.5% TritonX100 in  $1 \times$  PBS for 20 min. All washing steps were performed with PBS-T ( $1 \times$  PBS/0.02% Tween-20). For detection of PCNA, cells were further incubated for 5 min in ice-cold methanol on ice for antigen retrieval. Blocking (1% bovine serum albumin, 0.5% fish skin gelatin and 0.02% Tween-20 in  $1 \times$  PBS) was performed for 30 min at room temperature. EdU was detected using the Click-IT assay as described by the manufacturer (1:200 3-azido-

7-hydroxycoumarin, 1:1000 6-carboxyfluorescein (6-FAM azide) or 1:2000 5/6-Sulforhodamine azide; Cat. No.: 7811, 7806 and 7776 respectively, Carl Roth, Karlsruhe, Germany). Primary mouse anti-PCNA and secondary donkey anti-mouse IgG conjugated with Cy3 were diluted in blocking buffer and incubated for 1 h at room temperature. DNA was counterstained with DAPI (4',6-diamidino-2-phenylindole,  $10 \mu\text{g/ml}$ , Cat. No.: D27802, Sigma-Aldrich Chemie GmbH, Steinheim, Germany) for 10 min and samples were mounted in Mowiol4-88 (Cat. No.: 81381, Sigma-Aldrich Chemie GmbH, Steinheim, Germany) containing 2.5% DABCO (1,4-diazabicyclo[2.2.2]octane, Cat. No.: D27802, Sigma-Aldrich Chemie GmbH, Steinheim, Germany).

To analyze replication foci in 3D super-resolution microscopy (3D-SIM), samples were prepared as described (28,53). Briefly, cells were seeded on gelatin-coated high precision coverslips (Cat. No.: LH22.1, Carl Roth, Karlsruhe, Germany), grown for 6–8 h to ensure attachment but avoid 3D colony formation, and labeled with  $100 \mu\text{M}$  5-bromo-2'-deoxyuridine (BrdU) for 25 min followed by a thymidine chase ( $50 \mu\text{M}$ ) of 0, 30 or 60 min. Fixation and prestaining treatments were performed as described above. BrdU was detected with a rat anti-BrdU antibody diluted in buffer consisting of a 1:1 mixture of blocking and  $2 \times$  DNase I reaction buffer (60 mM Tris/HCl pH 8.1, 0.66 mM  $\text{MgCl}_2$ , 1 mM beta-mercaptoethanol) and 25 U/ml DNase I (Cat. No.: D5025, Sigma-Aldrich Chemie GmbH, Steinheim, Germany). Samples were incubated for 1 h at  $37^\circ\text{C}$  and DNase I digestion was stopped by washing with PBS-TE (PBS-T with 1 mM EDTA). PCNA detection was done as described above. After secondary antibody incubation with donkey anti-rat IgG AlexaFluor 488 and donkey anti-mouse IgG AlexaFluor 594, samples were mounted in Vectashield (Invitrogen, Carlsbad, CA, USA).

All nucleotide and antibody characteristics are summarized in Supplementary Tables S3 and S4, respectively.

#### **Immunofluorescence**

For pluripotency marker detection, cells were grown, fixed, permeabilized and blocked as described for DNA replication visualization. Primary mouse anti-Oct3/4, rabbit anti-Sox2 and secondary goat anti-rabbit IgG Alexa647 and goat anti-mouse IgG Alexa647 were diluted in blocking buffer and applied for 1 h at room temperature.

For histone modification analysis, cells were grown, EdU labeled, fixed, permeabilized and blocked as described above. EdU was detected as described by the manufacturer (1:1000 6-FAM azide). The following primary and secondary antibodies were diluted in blocking buffer: mouse anti-H3K9ac, rabbit anti-H3K9m3, rabbit anti-H4K5ac, rabbit anti-H4K8ac, donkey anti-mouse IgG Cy3 and donkey anti-rabbit IgG Cy3. Incubation was done for 1 h at room temperature.

DNA counterstaining and mounting using Mowiol was performed as described above.

All nucleotide and antibody characteristics are summarized in Supplementary Tables S3 and S4, respectively.

### Major satellite (MaSat) polydactyl zinc finger (PZF) fixation

Since transfected MaSat-GFP (pc1803 and (51)) is not fixable with standard formaldehyde or methanol fixation protocols, we made use of a gradient fixation protocol combined with a simultaneous mild permeabilization on ice. Twenty-four hours after transfection with MaSat-GFP, J1 mES cells were put on ice and 3.7% formaldehyde/1× PBS with 0.1% Nonidet™ P 40 Substitute (Cat. No.: 74385, Sigma-Aldrich Chemie GmbH, Steinheim, Germany) was added to the medium to achieve a final concentration of 0.1% formaldehyde. After 10 min, formaldehyde concentration was increased to 0.2% and incubated for 10 min. This procedure was repeated 6 times (0.5%, 1%, 1.5%, 2%, 2.5%, 3% formaldehyde). Increasing formaldehyde concentration to 1.5% was performed with a 0.1% Nonidet P 40 substitute containing formaldehyde stock solution. In a final step, the medium/formaldehyde mixture was replaced by 3.7% formaldehyde (10 min incubation) and exchanged with PBS-T.

### Molecular combing

Molecular combing experiments were performed using the FiberPrep® kit (Cat.No.: EXTR-001, Genomic Vision, Bagneux, France) and as described before (54). Briefly, J1 mES cells were pulse labeled for 15 min with 10 μM 5-chloro-2'-deoxyuridine (CldU), washed twice with pre-warmed PBS, labeled with 100 μM 5-iodo-2'-deoxyuridine (IdU) for 15 min, washed extensively and chased for 1 h with 50 μM thymidine. Cells were subsequently embedded in low-melting point agarose, genomic DNA was isolated by proteinase K (Cat. No.: 7528, Carl Roth, Karlsruhe, Germany) digestion and single high molecular weight DNA molecules were stretched on silanized glass coverslips (Cat. No.: COV-002-RUO, Genomic Vision, Bagneux, France), using the FiberComb®-Molecular Combing System (Cat. No.: MCS-001, Genomic Vision, Bagneux, France) as described by the manufacturer. Incorporated nucleotides and single stranded DNA were detected using mouse anti-BrdU/IdU, rat anti-BrdU/CldU and mouse anti-single stranded DNA (IgG2a) primary antibodies and goat anti-mouse IgG Chromeo 546, donkey anti-rat IgG AlexaFluor 488 and goat anti-mouse IgG2a AlexaFluor 647 secondary antibodies.

All nucleotide and antibody characteristics are summarized in Supplementary Tables S3 and S4, respectively.

### Probe generation, metaphase and (Repli-)FISH (fluorescence *in situ* hybridization)

Probes against major satellites, minor satellites and telomeres were generated as described in (55).

The Y chromosome probe was generated via DOP-PCR (degenerated oligonucleotide-primed-PCR). For template stock generation, PCR reactions contained mouse Y chromosome-specific template DNA (kind gift of Prof. Dr. Diane Krause, Yale University School of Medicine), 2 μM 6AI primer (5'-CCGACTCGAGNNNNNTACA CC-3'), 0.25 mM dNTPs and 2.5 U Taq polymerase in 1× PCR buffer (10 mM Tris/HCl pH 8.3, 50 mM KCl and 1.5

mM MgCl<sub>2</sub>) and cycling conditions were set to (45'' at 94°C, 45'' at 15°C, 12' at 37°C) × 1, (40'' at 94°C, 45'' at 37°C, 4' at 66°C) × 5 and (40'' at 94°C, 45'' at 54°C, 4' at 66°C) × 24. Y chromosome template DNA was labeled with biotinylated nucleotides (55) in the following reaction: template stock DNA was mixed with a nucleotide mixture containing unlabeled nucleotides (0.2 mM each dATP, dCTP and dGTP with 0.1 mM dTTP), biotinylated dUTPs (0.1 mM biotin-16-dUTPs), 2 μM 6AI primer, 2.5 U Taq polymerase and 1× PCR buffer. Cycles were set to (5' at 94°C) × 1, (30'' at 94°C, 30'' at 54°C, 90'' at 72°C) × 35 and (5' at 72°C) × 1.

Metaphase FISH and co-visualization of DNA replication and DNA probes were performed as previously described (55), using rabbit anti-digoxigenin, anti-rabbit IgG Cy3 antibodies and Streptavidin Alexa488 or Streptavidin Cy5 (1:500, Cat. No.: S11223 and SA1011, Thermo Fisher Scientific, Waltham, MA, USA).

Triple FISH was performed by a combination of biotinylated, Cy3 and Cy5 labeled probes.

To visualize MaSat repeat sequences in MaSat-GFP transfected cells, cells were fixed using a gradient fixation protocol described above and FISH was done as described above. Since heating GFP expressing cells to 80°C dramatically reduced GFP fluorescence, cells were stained with a mouse anti-GFP antibody to re-visualize MaSat-GFP. After probe annealing and washing, cells were first incubated with the anti-GFP primary antibody followed by the secondary donkey anti-mouse Alexa488 antibody and Streptavidin Cy5 for 1 h in blocking buffer (1% bovine serum albumin, 0.5% fish skin gelatin and 0.02% Tween-20 in 1× PBS).

All nucleotide and antibody characteristics are summarized in Supplementary Tables S3 and S4, respectively.

### Karyotype analysis

To prepare J1 mES cell metaphase spreads, cells were arrested in mitosis by adding 0.02 μg/ml colcemid (Cat. No.: 10 295 892 001, Roche Diagnostics GmbH, Basel, Switzerland) for 1.5 h at 37°C. The supernatant and harvested cells were pelleted by centrifugation for 5 min at 300 × g, cells were resuspended in 10 ml pre-warmed hypotonic solution (0.075 M KCl) and incubated for 6 min at 37°C. After centrifugation for 5 min at 300 × g, cells were fixed by dropwise addition of fixative solution (3:1 methanol:acetic acid) and incubation for 45 min on ice. Etched microscope slides were prepared by submerging the slides for 15–20 min in etching solution (0.1 N HCl in 95% ethanol) followed by cleaning steps in 95% EtOH and ddH<sub>2</sub>O (3 times each). Finally, spreads were generated by dropping fixed cells onto etched slides and air-drying. Individual metaphase spreads were imaged by phase-contrast microscopy and analyzed manually.

### Doubling time and S-phase duration

For growth curve analysis,  $2 \times 10^5$  J1 mES cells were seeded as technical quadruplicates at day 0 and cell numbers were counted with a Neubauer haemocytometer for four consecutive days. Population doubling times were derived with  $\log_2(n_x/n_0)/t$  (h) ( $n_x$ : cell number at day  $x$ ,  $n_0$ : cell number

at day 0,  $t$ : hours after seeding). To determine the percentage of cells in every cell cycle and S-phase substage, asynchronously growing J1 cell cultures were pulse labeled with 10  $\mu$ M EdU for 12 min, fixed and EdU was detected as described. Cells were manually grouped into S-phase substages (stage I to Y), non S-phase or mitosis, and percentages were calculated. S-phase (substage) duration was derived by multiplying the doubling time with the percentage of cells in the respective phase.

### Microscopy

All characteristics of the microscopy systems, including lasers, filters and objectives used, are summarized in Supplementary Table S5.

Molecular combing samples were imaged using a Zeiss Axiovert 200 widefield microscope.

Confocal z-stacks of live cells were acquired using the Ultra-View VoX spinning disk microscopy system. Time-lapse microscopy was carried out in a closed live-cell microscopy chamber at 37°C, with 5% CO<sub>2</sub> and 60% humidity. mRFP-PCNA and MaSat-GFP double transfected J1 mES cells were imaged every 30 min for 24 h to follow cell cycle progression.

Confocal z-stacks were acquired with a Leica TCS SP5 II confocal laser scanning microscope or the spinning disk microscope.

3D SIM images were acquired with a DeltaVision OMX V3 system (56).

### Image analysis

Image analysis was done using ImageJ (<http://rsb.info.nih.gov/ij/>, v1.51s and earlier), Volocity 6.3 (Perkin Elmer), ilastik (57) (<https://www.ilastik.org>, v1.3.3post3) and Python with the scipy-stack and scikit-image (58) (Anaconda distribution 2020.07).

**Quantitative analysis of replication foci features.** EdU labeled RFI in manually cropped confocal image stacks of mES cells were segmented by supervised pixel classification using ilastik (57). A subset of pixels was manually annotated as belonging to the background or to a replication focus (RF) and used to train a random forest classifier on pixel features to propagate the classification to the remaining pixels in all images. Similarly, a 3-class (background/nuclear border/nuclear interior) classifier was created for generation of nuclear marks. To separate touching nuclei, the center of each individual nucleus was manually labelled in the cropped images. Using Python and scipy/scikit-image (58), an instance segmentation was created via marker-controlled watershed using these manual markers and the probability of the 'border' class as the ridge image. For each crop, only the nucleus with centroid closest to the image center was used for the subsequent steps. The RFI mask was used to generate features for each RF in the EdU channel. Touching objects were separated by a watershed transform on the Euclidean distance transform (EDT) of the mask. For feature calculation, the stacks were scaled in z to achieve isotropic resolution. For each object overlapping more than 50% with the nuclear mask, a series of features were calculated (Supplementary Figure S1). RF objects smaller than 200 px<sup>3</sup> (~1

confocal PSF) were discarded. For each image, a single feature vector consisting of medians and standard deviation of the RFI features (see Supplementary Figure S1 for complete list of analyzed parameters), as well as the number of RFI and the chase duration was created.

Further analysis was performed using scikit-learn. Missing features were filled with the mean value of that feature in all images. The feature vectors were normalized to zero mean and unit variance and visualized in a 2D embedding via  $t$ -distributed stochastic neighbor embedding ( $t$ -SNE).

The analysis pipeline is summarized in Supplementary Figure S1 and the Python code for the pipeline is available at <https://doi.org/10.25534/tudatalib-220>.

### Replication signal and chromocenter colocalization analysis.

To determine the degree of colocalization of replication signals (EdU) and constitutive heterochromatin (chromocenters) in differentiated mES cells, cells were grouped into early, mid and late replicating cells ( $S_e$ ,  $S_m$  and  $S_l$ ) according to their EdU pattern. Chromocenters were segmented based on the DAPI signal, EdU signal intensities within the segmented regions were measured and EdU intensities in chromocenters were plotted for  $S_e$ ,  $S_m$  and  $S_l$  cells.

### Histone modification analysis.

To determine the histone modification accumulation at chromocenters, cells were segmented at single mid planes according to the DAPI channel and DAPI intensities were measured. S-phase cells were identified based on the EdU signal and cells were grouped into G1, S and G2 phases ( $\text{intensity[DAPI]}_{G1} < \text{intensity[DAPI]}_S < \text{intensity[DAPI]}_{G2}$ ) (20). For subsequent histone modification analysis, only G1 phase cells were considered. Within the DAPI channel, four circular regions of interest (ROI) were drawn inside DAPI intense chromocenters at mid planes, and four ROIs outside chromocenters. Histone modification levels were measured inside the eight ROIs. The analysis pipeline is summarized in Supplementary Figure S2. Mean intensity averages were calculated for the four regions inside chromocenters (chromocenter) and accordingly for regions outside chromocenters (nucleoplasm). Histone modification accumulation at chromocenters was determined as a ratio of chromocenter/nucleoplasm.

To determine histone modification changes after HDAC1 targeting to chromocenters, histone acetylation levels were measured in mid focal planes of individual transfected cell nuclei. Background in the histone channel was subtracted and fluorescence intensities were measured. Results are plotted as a ratio to GFP control cells.

**Chromocenter characteristic analysis.** Images of DAPI stained nuclei of (un)differentiated J1 mES cells and primary mouse ear fibroblasts were imported into Volocity 6.3 (Perkin Elmer). Pericentromeric heterochromatin (chromocenters) were segmented based on fluorescence intensity and the total number of chromocenters per cell as well as the volume, the degree of compaction and the shape factor of every chromocenter were calculated.

**Repli-FISH analysis.** To determine the replication timing of specific (sub)chromosomal elements, 3D masks of individual cell nuclei were manually generated based on the

DAPI counterstaining (single FISH) or the EdU staining (triple FISH). FISH signals were segmented independently after applying a Gaussian filter ( $\sigma = 1$ ) and subtracting background (rolling ball algorithm with radius = 10). Similarly, PCNA images were background corrected and basic PCNA signal from non replicating cells (non S-phase cells) was subtracted. The 3D nucleus mask was subsequently applied to the segmented FISH signals and the generated nuclear FISH ROIs were used to mask the PCNA signals. PCNA intensities within the mask were measured and plotted. The analysis pipeline is summarized in Supplementary Figure S3.

To analyze if the replication timing of telomeres is dependent on their 1D proximity to chromocenters, telomere signals (visualized by Repli-FISH in combination with PCNA) in stage II S-phase cells were segmented and grouped according to their location compared to chromocenters. Telomeres located on the short arm of acrocentric mouse chromosomes are in close 1D proximity of pericentromeric heterochromatin (referred to as telomeres in chromocenters), while telomeres capping the long arm of chromosomes are more distant from major satellite repeats (referred to as telomeres out of chromocenters). PCNA intensities within the segmented telomere signals were measured and plotted.

**Replication foci analysis.** Quantification of replication foci within individual cell nuclei was mainly performed as described in (28,53). In brief, 3D-SIM images were reconstructed, exported from the DeltaVision software (softWoRx 6.0 Beta 19, Applied Precision) and raw 3D-SIM images were converted to 16-bit images using a custom-written FIJI (59) macro. Individual cell nuclei were segmented using maximum intensity projections of the DAPI signal. Replication signals were segmented by auto-thresholding using the Triangle method. The resulting binary images were used to mask the original replication foci signals of interest and to discriminate them from background (set to '0'). These images and the corresponding DAPI images were imported to the image analysis software Volocity 6.3 (Perkin Elmer) and replication foci were quantified for individual nuclei. First, nuclear masks were generated based on the DAPI images and defined as regions of interest (ROIs). Next, 3D-SIM replication foci were detected by intensity excluding only black pixels (i.e. background with intensity '0'), touching foci were separated (object size guide =  $0 \mu\text{m}^3$ ) and signals smaller than  $0.0002 \mu\text{m}^3$  were excluded from the final counting as they represented unspecific background signal. Only foci within the nuclear ROI were counted. A detailed analysis pipeline is summarized in Supplementary Figure S4.

Pseudo wide-field (pseudoWF or pWF) replication signals were generated from the same datasets (C2C12 and J1 mES cells labeled with BrdU). Generation of the pseudo wide-field data was described in (60). For correlation analysis, the pseudoWF images were initially processed in ImageJ to match the image dimension of the 3D-SIM data. To achieve similar voxel sizes ( $40 \times 40 \times 125 \text{ nm}$ ), the images were scaled using a bicubic interpolation, doubling the number of pixels in  $x$  and  $y$ . Next, pseudoWF images were corrected for pixel shifts by translating the image stack  $-2$  pixels in  $x$  and  $y$ . For segmentation of the pseudoWF repli-

cation signals, the histogram was normalized and a background subtraction was performed using a rolling ball algorithm with radius = 10. Segmentation was performed by auto-thresholding using the Otsu algorithm. 3D-SIM replication signals were processed as described above. Finally, segmented and masked pseudoWF and 3D-SIM image stacks were merged and used for foci counting in Volocity. Detection of pWF RFI was based on intensity as for 3D-SIM images, separation of touching objects was based on object size (object size guide =  $0.02 \mu\text{m}^3$ ) and signals smaller than  $0.02 \mu\text{m}^3$  were excluded. Overlapping signals used for nanoRFI counting within pWF RFI were filtered by an additional compartmentalization step. The analysis pipeline is summarized in Supplementary Figure S5.

**DNA fiber analysis.** For replication signal analysis from molecular combing experiments, fluorescent DNA fiber tracks were selected according to their pattern. Only lengths of the second pulse (CldU) of progressing forks (CldU track preceded by a clear IdU signal on fibers with a ssDNA signal up- and downstream of the marked tracks) were considered. Replication fork speed (RFS) was calculated as a ratio of the track length (track length  $\times$  2000 due to the constant stretching factor resulting in  $1 \mu\text{m} \sim 2000 \text{ nts}$ ) and the time of nucleotide application. Inter-origin distance (IOD) was calculated in kb as the product of the measured track length and the conversion factor of 2 ( $1 \mu\text{m} \sim 2000 \text{ nts}$  or 2 kb). Bidirectional fork asymmetry was analyzed as the ratio of the long track and the short track and percentage of unidirectional forks was calculated by dividing the number of unidirectional forks by the total number of analyzed forks (uni- and bidirectional forks). A graphical summary of the selection of the fiber tracks and the corresponding calculations are depicted in Figure 7A and Supplementary Figure S6.

### Genome-wide replication origin profiling

The GEO (Gene Expression Omnibus, <https://www.ncbi.nlm.nih.gov/geo/>) samples GSM3602315, GSM3602316 and GSM3602317 from the dataset GSE126477 (61) and samples GSM2651111 and GSM2651112 from the dataset GSE99740 (62) were used for genome-wide replication origin profiling in mES cells. The samples GSM2651107 and GSM2651108 from the dataset GSE99740 (62) were used for genome-wide replication origin profiling in MEF cells. The above-mentioned datasets correspond to five replicates of origin mapping in mES cells and two replicates in MEF cells realized by sequencing of isolated small nascent DNA strands (SNS-seq). The analysis was performed using peaks reproducibly found in at least two replicates using the multiIntersectBed command in bedtools (63).

The GEO samples GSM3227970, GSM3227971 and GSM3227972 from the dataset GSE116321 (64) were used for genome-wide replication origin profiling in activated mouse B cells using Okazaki fragment sequencing method (OK-seq). The coordinates of OK-seq replication initiation zones were kindly provided by Andre Nussenzweig and Sridharan Sriram. The mm10 reference genome (<http://hgdownload.cse.ucsc.edu/goldenPath/mm10/bigZips/>) was

used in all data analysis. The operations on genomic intervals were performed using bedtools (63).

The IODs were calculated between the middle point of each replication origin zone identified and the distances between origins flanking chromosomal regions unmapped in sequencing analyses (centromeres, etc.) were omitted from the plots.

All genome-wide origin mapping datasets and samples are summarized in Supplementary Table S6.

### Data visualization and statistical analysis

Data visualization and statistical analysis (independent two-group student's *t*-tests and Mann–Whitney–Wilcoxon tests) were performed with *RStudio* (v1.0.143–v1.1.447, <https://rstudio.com/>).

Visualization of origin replication profiles was performed with IGV (Integrative Genomics Viewer, version 2.8.6, <https://software.broadinstitute.org/software/igv/>).

Statistical values (number (#) of cells (*N*), mean, median, standard deviation (SD), standard error of the mean (SEM), 95% confidence interval (CI) and *P*-values) are indicated in the plots or summarized in Supplementary Tables.

Boxplots and violin plots represent the median (center line) with the box depicting the 25–75 percentiles and the lines the upper and lower whiskers with 1.5 times the IQD (inter-quartile distance) (Supplementary Figure S7). Barplots show averaged values and error bars the respective standard deviation.

All cells analyzed (*N* numbers stated in Supplementary Tables) showed the reported behavior of the representative images shown in the respective figures.

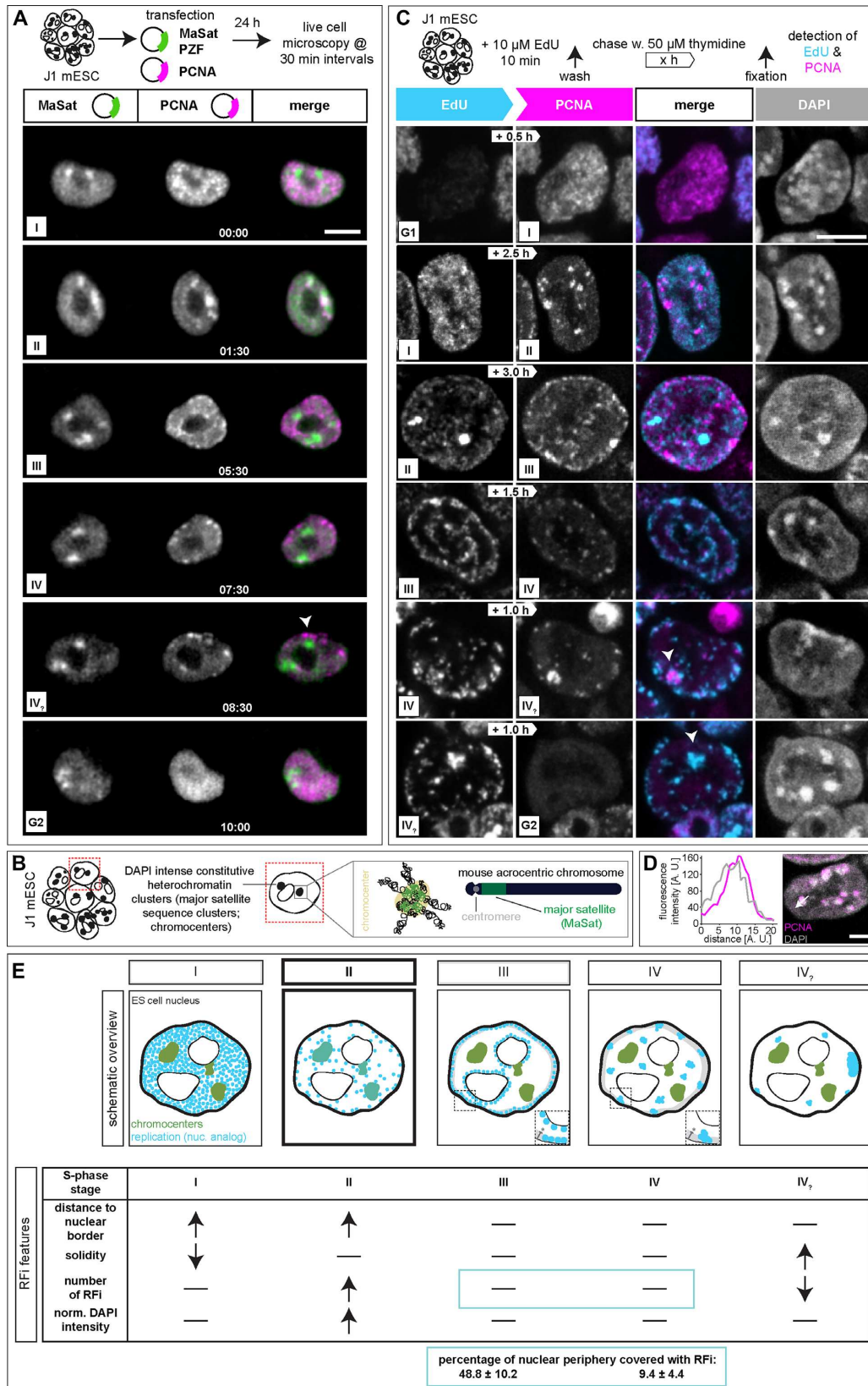
## RESULTS AND DISCUSSION

### Characterization of the spatio-temporal DNA replication patterns in mouse embryonic stem cells reveals differences to somatic cells

Replication patterns are a direct visual representation of the spatial organization and temporal order of DNA replication and, in somatic cells, have been shown to reflect the chromatin organization level (10,15,16). DNA replication timing profiles (RT-profiles) from large cell populations (43,65–68) revealed distinct replication domains (1.5–2.5 Mb), that exhibit sharp boundaries between neighboring domains with different replication timing, alternating along individual chromosomes (43). While RT-profiles can directly be linked to the underlying DNA sequence, they fall short on temporal resolution and do not provide 3D spatial information. Single cell microscopy analysis, however, allows a specific 4D analysis of DNA replication. Importantly, it allows to map replication timing of DNA repeat elements, which are largely not mappable by sequencing-based approaches and constitute a very large portion of mammalian genomes (69). To investigate and compare the spatio-temporal organization of DNA replication (5) of pluripotent cells, we first analyzed S-phase progression in mouse mES cells by live cell imaging experiments. Therefore, we transfected J1 mES cells with plasmids encoding

for mRFP-PCNA and a GFP tagged polydactyl zinc finger protein specifically binding to major satellite sequences (MaSat PZF) and imaged the cells every 30 min for 24 h (Figure 1A). To exclude artifacts introduced by transfection and overexpression of fluorescently tagged PCNA, we validated mRFP-PCNA localization during S-phase by immunostaining with a PCNA specific antibody on transfected mES cells (Supplementary Figure S8A). Additionally, we validated binding of the MaSat PZF to major satellite repeats. Due to its high mobility and fast binding kinetics, MaSat PZF is not fixable with standard formaldehyde or methanol fixation protocols. We, therefore, established a gradient formaldehyde fixation protocol with simultaneous permeabilization, and performed fluorescence *in situ* hybridization (FISH) with a probe specifically binding to MaSat repeats. MaSat-GFP colocalized with the major satellite probe signal, as well as with DAPI intense nuclear regions (Supplementary Figure S8B). The live cell microscopy approach provided a detailed spatio-temporal analysis of S-phase progression *in vivo* and revealed visually distinguishable spatial replication patterns in mouse J1 mES cells (Figure 1A, Supplementary Figure S8C, Movies 1 and 2). At the beginning of S-phase, replication foci distributed homogeneously in the nuclear interior (stage I), possibly reflecting duplication of the euchromatic portion of the genome, as seen in somatic cells. Next, and differing from somatic cells, MaSat PZF-labeled and condensed clusters of pericentromeric heterochromatin (chromocenters, Figure 1B) were replicated (stage II). Although this observation might be unexpected in view of the late replication timing of heterochromatin in somatic cells, it has been previously shown that in *Drosophila*, satellite sequences became increasingly heterochromatic and late replicating only with successive differentiation at later developmental cycles (70). After pericentromeric heterochromatin replication, duplication of chromatin located at the nuclear and nucleolar borders was observed in mES cell (stage III), which, in somatic cells, reflects duplication of facultative heterochromatin. Next, the cells displayed a pattern with a decreased number of foci which however increased in size, suggesting clustering of the underlying chromatin fiber. Replication signals were mostly, but not exclusively, located at the nuclear periphery (stage IV). The end of S-phase was marked by a strong accumulation of replication signals within one particular region of the mES cell nucleus (stage IV<sub>?</sub>). Additionally, we measured the duration of the individual S-phase substages from the live cell microscopy. While early and mid S-phase (stage I and II) lasted over 3.5 and 4.5 h, respectively, the subsequent phases were relatively short, with average durations of ~1 h each (stage III, IV and IV<sub>?</sub>, Supplementary Figure S8D and Supplementary Table S7).

Next, we validated the observed S-phase progression in mES cells by pulse-chase experiments. Asynchronously growing mouse J1 mES cells were pulse labeled for 12 min with the nucleotide analog 5-ethynyl-2'-deoxyuridine (EdU), washed to remove nucleotide excess, chased with thymidine for different periods of time and fixed. Subsequently, replication foci marked by the incorporated EdU and the replisome component PCNA were (immuno)fluorescently detected (Figure 1C and Supplemen-



**Figure 1.** DNA replication dynamics in mouse embryonic stem cells. (A) Experimental setup of a live cell experiment to determine the *in vivo* spatio-temporal progression of DNA replication in J1 mES cells. Cells were transfected with plasmids encoding mRFP-PCNA (magenta) and GFP-tagged

tary Table S11). EdU marked DNA replicated during the nucleotide pulse, while the PCNA pattern corresponded to the active replication sites at the time of fixation of the cells. The different chase times resulted in different degrees of replication foci separation and, with increasing chase times, eventually also to transition from one S-phase pattern to the next (Supplementary Figure 9A). To achieve visualization of S-phase substage progression and, concomitantly, separation of replication patterns of the different stages (Supplementary Figure S9B), chase durations were changed according to S-phase substage duration, as obtained from the time lapse movies (Supplementary Figure S8D and Supplementary Table S7). Live cell microscopy revealed substantially longer durations of stage I and II compared to stage III and IV (Supplementary Figures S8C, S8D and Supplementary Table S7). Hence, to reflect stage to stage transitions, chase times had to be adapted accordingly. To visualize stage I to stage II transitions in a significant number of S-phase cells, chase times had to be longer than to visualize transitions from stage III to stage IV (Supplementary Figure S9B). This approach allowed us to get a spatio-temporal resolution of DNA replication in fixed cells, underlining the domino-like DNA replication model (Supplementary Figure 9C). As in the live-cell data, replication of the large heterochromatin clusters (stage II), colocalizing with DAPI intense nuclear regions (Figure 1D), took place after the early S-phase stage I pattern and was followed by the nucle(ol)ar periphery stage III replication pattern. All other features described in the live cell experiments were also reproduced with the pulse-chase approach in fixed cells.

Since mouse ES cells grow in very specific microenvironments in cell culture, which include the presence of leukemia inhibitory factor (LIF) and two inhibitors (2i, PD032591 and CHIR99021) to preserve pluripotency and self-renewal capacities, as well as form 3D colonies, we tested the influence of this stem cell specific microenvironment on the temporal organization of DNA replication. S-phase progression from stage I to stage II was observed for single mES cells (Supplementary Figure S10A) and in cells grown in the absence of the 2i (Supplementary Figure S10B), similar to cells grown in 3D colonies in the presence of 2i and LIF. Additionally, we analyzed the replication foci pattern distribution over time via the above mentioned pulse-chase experiments in mouse E14 mES cells, and found similar spatio-

temporal progression as observed for J1 cells (Supplementary Figure S11) indicating a conservation of S-phase characteristics for mouse pluripotent cells.

In summary, the general characterization of the spatial distribution of replication signals during S-phase progression in mouse embryonic stem cells by pulse-chase and live cell experiments led to the identification and temporal classification of a sequence of replication patterns, which differ from somatic cells (Figure 1E).

### Quantitative features of mouse embryonic stem cell replication foci patterns

To characterize the different replication patterns observed in live cell and pulse chase experiments (Figure 1E), we determined quantitative features of the underlying replication foci (RFi) within the different S-phase substages (I–IV<sub>2</sub>). EdU labeled cells were classified according to their S-phase pattern, replication signals were segmented, separated using a watershed algorithm and location, shape and intensity features were determined (Figure 1E and Supplementary Figure S12). In a 2D embedding via *t*-distributed stochastic neighbor embedding (*t*-SNE) the RFi features did not form clearly separated clusters, but temporally adjacent stages lied next to each other in the embedding, hinting at a continuum in feature space from the beginning to the end of replication (Supplementary Figure S13). Stage I of mES cell S-phase was characterized by the combination of a significant increase in the distance of the RFi from the nuclear border (Figure 1E and Supplementary Figure S12) and a decrease in RFi solidity ('rougher'). In contrast, stage II was marked by an increase in RFi numbers with a concomitant increase in DAPI intensity. This is in agreement with our findings that DAPI intense nuclear regions, i.e. chromocenters, are replicated during this stage (Figure 1A–D). Moreover, we determined specific characteristic features for stage IV<sub>2</sub>. At the end of S-phase, the amount of RFi decreased dramatically, whereas solidity, i.e. smoothness, increased. As RFi volumes as well as other features in stages III and IV showed no significant differences, we additionally calculated the percentage of the nuclear periphery covered by replication foci in these two stages. The size of the mask used to segment the nuclear periphery was set to the diameter of the RFi located at the nuclear border in stage III (Figure

polydactyl zinc finger protein (PZF) specifically binding to major satellite repeats (MaSat-GFP, green) fusion constructs to mark ongoing DNA replication and pericentromeric heterochromatin (chromocenters), respectively. Imaging was performed for 24 h with 30 min intervals. Representative spinning disk confocal images show the cell cycle progression of a representative mES cell and S-phase was further subdivided into five main replication patterns (I–IV<sub>2</sub>). The arrowhead in the IV<sub>2</sub> stage marks a prominent accumulation of replication signals observed at the end of S-phase. (B) Schematic representation of acrocentric mouse chromosome clustering in mES cell nuclei. At the chromosomal level, constitutive heterochromatin major satellite repeats (green) flank the centromere (grey) and in interphase nuclei, pericentromeric DNA from different chromosomes clusters to chromocenters. (C) Experimental setup of a pulse-chase experiment to determine the spatio-temporal progression of DNA replication in mouse J1 ES cells. Asynchronously growing mES cell cultures were pulse labeled with the nucleotide analog EdU, followed by various thymidine chase periods (white arrows) and fixation. EdU, i.e. nascent DNA during the first pulse labeling (cyan), and endogenous PCNA, i.e. ongoing replication at the time point of fixation (magenta), were (immuno)fluorescently detected and allowed the identification and the temporal order classification of five main replication patterns in mouse J1 ES cells. Representative spinning disk confocal images of G1 to S-phase, S-phase substage transitions (I–IV<sub>2</sub>) and S-phase to G2 progression are shown. The arrowheads in the IV<sub>2</sub> stage mark a prominent accumulation of replication signals observed at the end of S-phase. (D) Line profile analysis of PCNA fluorescence intensities within one chromocenter in a stage II cell. (E) Schematic summary of the five replication patterns observed in mES cells. Replication signals are shown in cyan, pericentromeric heterochromatin in green and replicating chromocenters (stage II) are marked in dark cyan. Table summarizing the significant RFi features of the five S-phase substages. Grey areas in stage III and IV S-phase schematic cells represent the nuclear periphery segmented to determine the percentage of the latter covered with RFi. The width of this area (\*) corresponds to the diameter of stage III peripheral RFi ( $0.44 \pm 0.13 \mu\text{m}$ ). All experiments were done in at least three independent biological replicates. Detailed statistics are summarized in Supplementary Tables S7 and S11. Scale bars = 5  $\mu\text{m}$ . Dotted lines represent cell contours.

1E). While almost half of the nuclear border exhibited RFI in stage III cells, only 10% of the periphery contained RFI in stage IV cells. Since these observations suggest a difference in the clustering of the underlying chromatin, we retained the S-phase substage division into five substages. Taken together, we identified several distinct location, shape and intensity features of replication foci that characterize and distinguish mES cell S-phase substages.

### Constitutive heterochromatin shifts its replication timing during loss of pluripotency

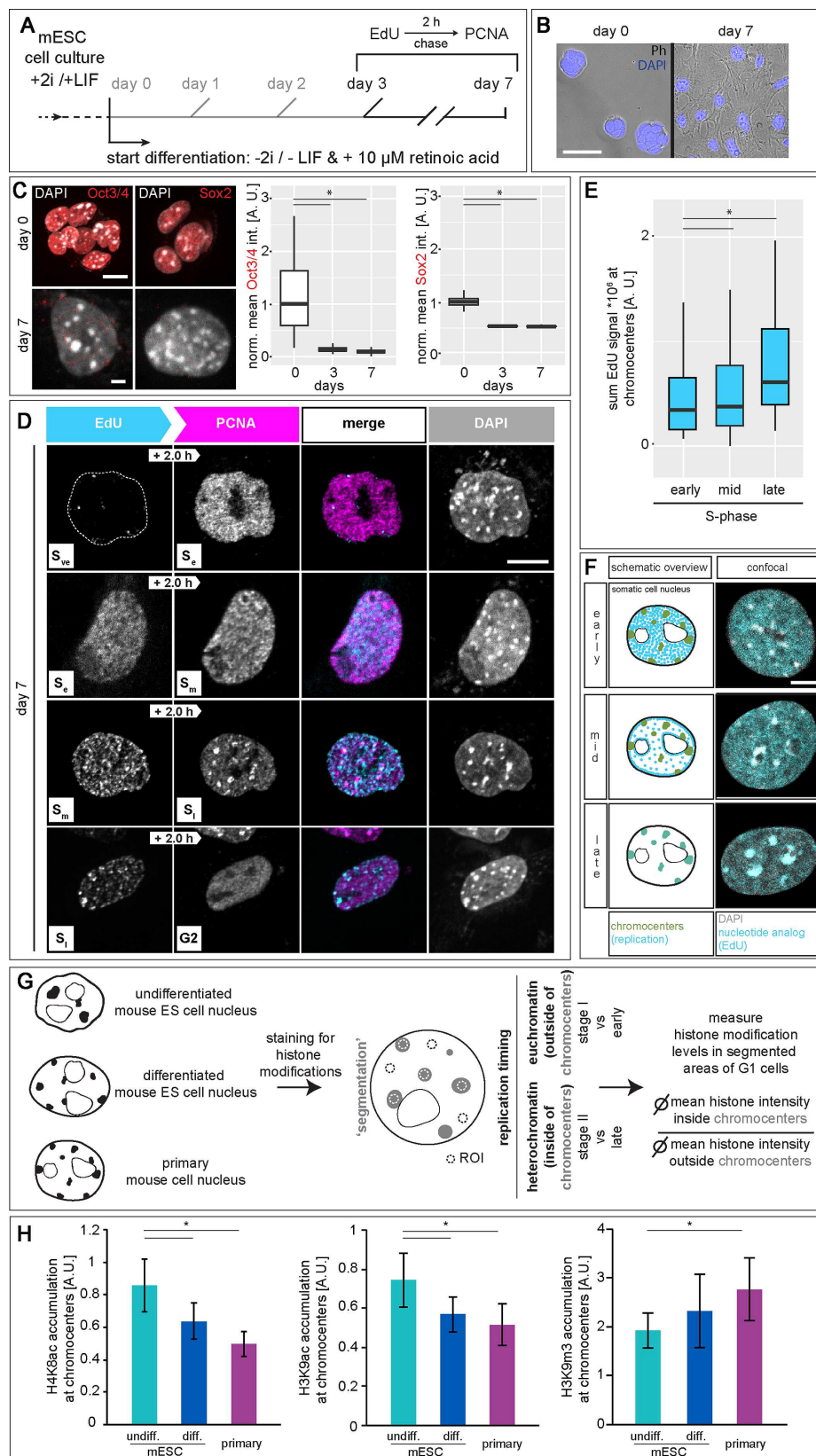
In somatic cells, S-phase progression follows chromatin compaction and is commonly subdivided into early ( $S_e$ ), mid ( $S_m$ ) and late ( $S_l$ ) when euchromatin, facultative and constitutive heterochromatin, respectively, are replicated (1,7,14). Since we observed replication of mES cell chromocenters within the first half of S-phase, we aimed to clarify if and when during mES cell differentiation, replication timing of chromocenters switches to late replication. Therefore, we differentiated mES cells in the absence of 2i and LIF and in the presence of retinoic acid (RA) and performed pulse-chase labeling from day 3 to day 7 of differentiation (Figure 2A). During differentiation, mES cell morphology changed from round and compact 3D colonies (day 0) to more flat and spread out cells growing in a monolayer (day 7, Figure 2B). Concomitantly, levels of the pluripotency markers Oct4 and Sox2 decreased dramatically already after 3 days of differentiation and were almost undetectable after day 7 of differentiation (Figure 2C and Supplementary Table S8). All together, these results indicated loss of the stem cell phenotype and pluripotency markers that are associated with exit from pluripotency and cellular differentiation.

At day 7 of differentiation, the beginning of S-phase was marked by very few replication foci (very early S-phase,  $S_{ve}$ ), followed by a homogeneous distribution of replication foci throughout the nucleus ( $S_e$ , Figure 2D). Interestingly, and in contrast to undifferentiated mES cells (Figure 1), replication signals were next observed at the nucle(ol)ar periphery ( $S_m$ ). Importantly, replication of chromocenters took place at the end of S-phase ( $S_l$ ) and was followed by transition to G2 phase. Colocalization analyses between DAPI intense nuclear regions and EdU signals showed increased overlaps of the two signals in cells showing a late S-phase pattern, further confirming the switch to late replicating chromocenters in differentiated mES cells (Figure 2E and Supplementary Table S8). In summary, the observed spatio-temporal order of replication pattern in differentiated mES cells reflects the subdivision into  $S_e$ ,  $S_m$  and  $S_l$  known from somatic cells (Figure 2F).

### Replication timing of constitutive heterochromatin depends on histone acetylation levels

Chromocenters are marked by the trimethylation of histone H3 at lysine 9 (H3K9m3, (47)), by histone hypoacetylation (71), exhibit increased levels of DNA methylation (72) and are bound by specific heterochromatin proteins (73). Previous studies showed that cell types of a different origin or developmental status have a different organization of pericentromeric heterochromatin (74–76). In view

of this and since a more open chromatin state was proposed to facilitate early replication onset, we characterized pericentromeric heterochromatin clusters in DAPI stained mES cells and compared them to chromocenters of differentiated mES cells and primary mouse fibroblasts (Supplementary Figure S14A and Supplementary Table S9). Average numbers of chromocenters doubled over cell differentiation (Supplementary Figure S14B and Supplementary Table S9), reflecting differences in clustering of the pericentromeric regions from multiple chromosomes. Cooperatively with the increase in chromocenter numbers, we observed a decrease in the volume of individual chromocenters in differentiated cells (Supplementary Figure S14C and Supplementary Table S9). In line with this, analysis of the compaction state of the chromocenter clusters in pluripotent cells showed a more decompacted chromatin (Supplementary Figure S14D and Supplementary Table S9) and a more irregular shape (Supplementary Figure S14E and Supplementary Table S9). Accordingly, super-resolved chromatin mobility assays demonstrated a more dynamic chromatin and less defined domain structures in mES cells (77). In addition, chromatin associated proteins were found to be more mobile in pluripotent cells (78,79). It is considered that such an open conformation represents a necessary prerequisite of pluripotent cells to remain responsive to the changes that occur during differentiation (80). Additionally, the differences in clustering of chromocenters reflect the chromatin reorganization that occurs during differentiation and development (73,81). The observed differences in heterochromatin morphology, volume, clustering and compaction may provide a mechanistic basis for the observed changes in the DNA replication program of mES cells. However, a general decompaction at the scale measured via DAPI staining of DNA may not be sufficient to promote early replication of pericentromeric heterochromatin. Additionally, maintenance of late replication timing of pericentromeric heterochromatin in somatic cells was shown to depend on histone hypoacetylation (32). We, therefore, compared histone acetylation levels of chromocenters in (un)differentiated and primary mouse fibroblasts (Figure 2G) and found higher H3K9ac and H4K8ac accumulation at chromocenters in undifferentiated mES cells (Figure 2H and Supplementary Table S8). Heterochromatin acetylation levels decreased significantly during differentiation, concomitantly with the switch of replication timing of pericentromeric heterochromatin from relatively early (stage II) to late replication ( $S_l$ ). Since H3K9ac and H3K9m3, a marker for constitutive heterochromatin, are found in a mutually exclusive way in cells, we also analyzed accumulation of the latter at chromocenters. In line with the increased acetylation levels, we found less H3K9m3 accumulation in pluripotent stem cells and an increase in H3 lysine 9 trimethylation in differentiated cells (Figure 2H and Supplementary Table S8). This result is in line with genome wide studies as well as single cell microscopy analysis showing that histone acetylation is an important regulator of the replication timing of DNA. While genomic loci with increased histone acetylation levels and high accessibility tend to replicate early during S-phase, loss of histone acetylation leads to replication timing switch to late S-phase (32–34,82–89). Taken together, mid S-phase replication of pericentromeric



**Figure 2.** DNA replication dynamics in differentiated mouse embryonic stem cells. (A) Experimental setup of pulse-chase experiments in differentiated mouse J1 ES cells. Naïve pluripotent mESC cells were cultured in 2i (two inhibitors (PD032591 and CHIR99021)) and LIF (leukemia inhibitory factor)

heterochromatin in pluripotent stem cells is changed to late replication in differentiated mES cells and this switch in replication timing is likely dependent on histone hypoacetylation and chromatin compaction occurring during cell differentiation.

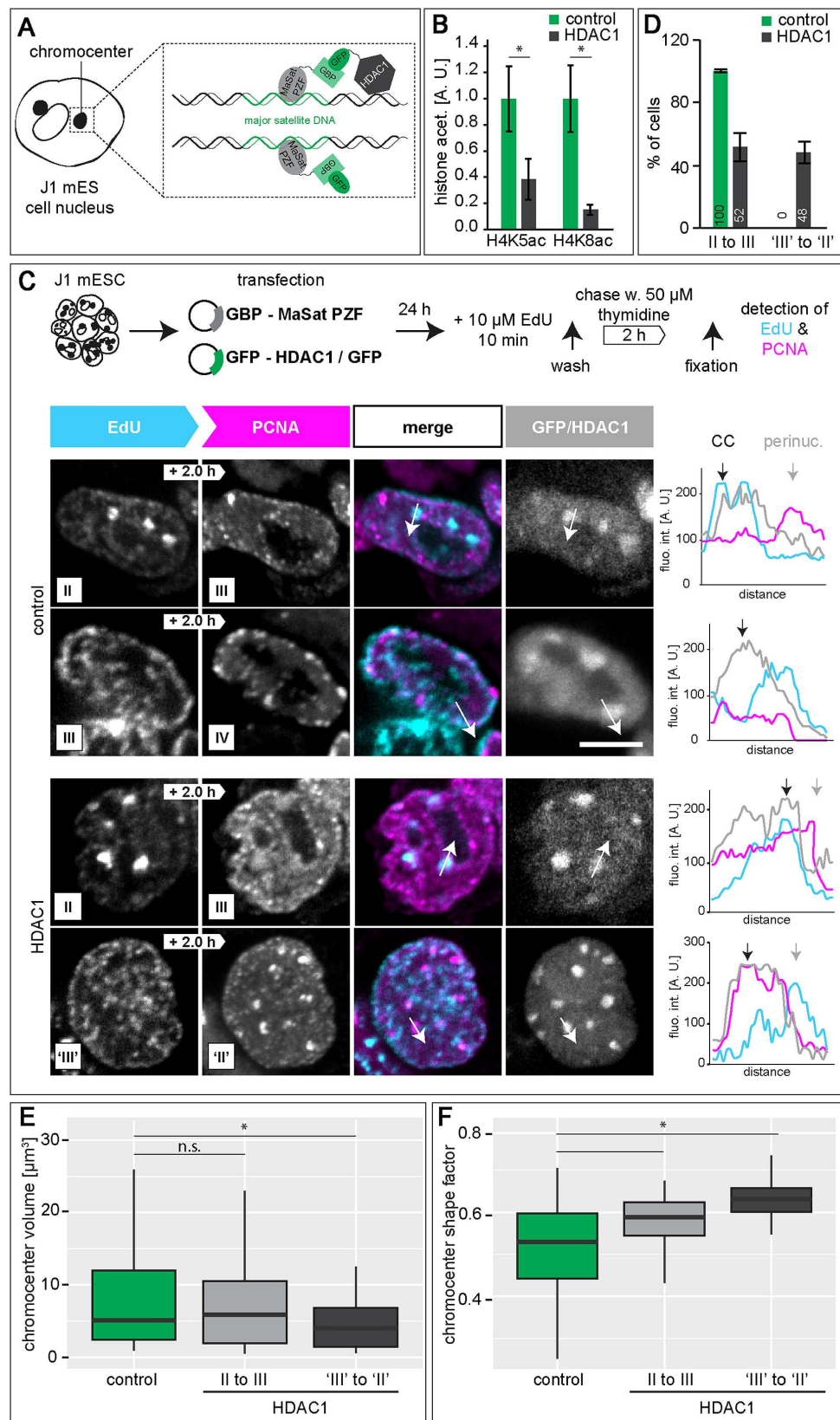
To validate the causality between the observed histone hyperacetylation and early/mid replication of pericentromeric heterochromatin in mES cells, we targeted a GFP tagged histone deacetylase to chromocenters in mES cells (90). Specific chromocentric targeting was achieved via co-transfection of GFP tagged HDAC1 and a GFP binding protein (GBP) tagged MaSat polydactyl zinc finger (PZF) protein (GBP-MaSat). MaSat PZF specifically binds to major satellite repeat DNA (chromocenters) and its GBP domain interacts with GFP-HDAC1, thereby recruiting the deacetylase to chromocenters (Figure 3A). As a control, we targeted GFP to chromocenters. With this setup, we first analyzed histone acetylation levels in transfected cells, and found decreased H4K5ac and H4K8ac levels in HDAC1 targeted cells compared to control GFP cells (Figure 3B and Supplementary Table S8). To analyze spatio-temporal S-phase progression in transfected cells with altered chromocentric histone acetylation levels, we performed pulse chase experiments. Therefore, 24 h after double transfection, cells were labeled with EdU, chased for 2 h and fixed. EdU and PCNA were (immuno)fluorescently detected and DNA replication progression was investigated by pattern order analysis as before. In GFP control targeted cells, we observed spatio-temporal DNA replication progression as previously described. Pericentromeric heterochromatin was replicated before perinucle(ol)ar chromatin and finally replication signals were observed as bigger replication foci throughout the nucleus (stage II–stage III–stage IV). In HDAC1 targeted cells, however, we observed two populations of cells. First, we found cells following the classical mES cell replication progression where S-phase moves from chromocenters to the perinucle(ol)ar border. Second, cells showed replication of the perinucle(ol)ar chromatin followed by replication signals in normally earlier replicating chromocenters (stage ‘III’ to stage ‘II’). This indicates an at least partial delayed chromocenter replication upon histone acetylation level decrease (Figure 3C–D and Supplementary Table S8). Quantitative analysis revealed almost equal distributions between the two substage orders

in HDAC1 targeted cells, while control cells only showed stage II to stage III transitions (Figure 3D and Supplementary Table S8). Interestingly, after HDAC1 targeting we observed massive rearrangements of pericentromeric heterochromatin. GFP control cells exhibited few but large chromocenters, while 24 h of histone deacetylase targeting resulted in an increased number of constitutive heterochromatin clusters with decreased size. Analyzing chromocenter characteristics, we measured similar volumes and shape factors for control cells than in untransfected mES cells (Figure 3D–E, Supplementary Figure S14C, S14E and Supplementary Tables S8 and S9). HDAC1 targeted cells were subdivided according to their S-phase progression pattern (II to III or ‘III’ to ‘II’). We detected significant changes in chromocenter volumes in stage ‘III’ to ‘II’ cells compared to control and stage II to III cells. Additionally, shape factors were also significantly different in targeted cells showing a replication pattern switch (Figure 3E–F and Supplementary Table S8). In conclusion, we show that a targeted histone deacetylation of mES chromocenters leads to a switch in replication timing of pericentromeric heterochromatin to later in S-phase, mimicking our findings in differentiated mES cells (Figure 2D).

#### Replication timing of (sub-)chromosomal elements in mouse embryonic stem cells

In human and mouse genomes, only minor parts (1.2–1.4%, respectively (91,92)) are protein-coding sequences, while the major portion is composed of non-coding DNA, including interspersed and tandem repeat sequences. There is growing evidence that the latter is more than ‘junk’ DNA, since it is thought to be involved in the establishment of distinct eu- and heterochromatin compartments (93–95). Additionally, genome function might not only be influenced by epigenetic factors, but also by the spatial organization of the genome within the cell nucleus (96). Hence, we analyzed the nuclear distribution and the replication timing of several sub-chromosomal tandem repeat elements in mouse ES cells. Due to their repetitive nature, repeat elements are normally under-represented in genome-wide sequencing studies. We, therefore, opted for a single cell microscopic approach, where we combined DNA replication visualization and marking of three major chromosomal tandem repeats

← containing medium. On day 0 of the differentiation, cells were seeded in 2i and LIF deficient medium containing retinoic acid. Cells were pulse labeled with EdU, chased with thymidine for 2 h and fixed at day 0 and from day 3 to day 7 of the differentiation. (B) Overlay of phase contrast (Ph) and DAPI channels, showing changes in cellular morphology during mES cell differentiation. (C) Immunofluorescent detection of the pluripotency markers Oct3/4 and Sox2 in (un)differentiated mES cells. Representative spinning disk confocal images of *in situ* stainings were imaged and the mean value of the fluorescence signal was plotted as a ratio to the undifferentiated (day 0) cells. (D) Representative spinning disk confocal images of pulse chased mES cells revealed replication timing switch to late replicating chromocenters at day 7 of differentiation. Very early ( $S_{ve}$ ) to early S-phase ( $S_e$ ), early to mid ( $S_m$ ), mid to late ( $S_l$ ) and late S-phase to G2 transitions are shown. (E) To analyze the replication timing switch of chromocenters in differentiated mES cells, the sum value of EdU fluorescence signal within chromocenters (masked according to DAPI channel) were measured in spinning disk confocal images of cells (from D) within early, mid and late S-phase for the EdU pulse. An increase in signal overlap of chromocenters and replication signal is observed during late S-phase in differentiated mES cells. (F) Schematic summary and corresponding confocal images of the three main replication patterns observed in differentiated mES cells. Replication signals are shown in cyan, pericentromeric heterochromatin in green (scheme) and replicating chromocenters (late) are marked in dark cyan. (G–H) Experimental setup for the analysis of histone modification accumulation at pericentromeric heterochromatin. (Un)differentiated mES cells and primary mouse ear fibroblasts were pulsed with EdU to identify S-phase cells and histone modifications were immunofluorescently detected. Regions of interest (ROI) were manually drawn in G1 phase cells and histone modification levels were measured. Shown are the accumulations of H4K8ac, H3K9ac and H3K9m3 at chromocenters (ratio of mean histone modification values at chromocenters and mean histone modification values in the nucleoplasm  $\pm$  StDev). Scale bars = 5  $\mu$ m. \* $P < 0.05$ . All experiments were done in at least two independent biological replicates. All boxes and whiskers represent 25–75 percentiles and 1.5 times the IQD (inter-quartile distance), respectively and the center line depicts the median (Supplementary Figure S7). Detailed statistics are summarized in Supplementary Table S8. Dotted lines represent cell contours.



**Figure 3.** HDAC1 deacetylase targeting to chromocenters in mouse embryonic stem cells. (A) Schematic representation of the targeting strategy to recruit HDAC1 to chromocenters. In a targeted state, the fusion protein composed of the major satellite binding MaSat PZF (MaSat) and GFP binding protein

(Figure 4A and Supplementary Figure S15A) via FISH (Repli-FISH, (55)).

To analyze the replication timing of the repeat elements, we quantified the overlap of the replisome factor PCNA or nascent DNA labeled by EdU with individually segmented repeat sequence specific FISH signals within 3D mES cell interphase nuclei. The specificity of the FISH probes for (peri)centromeric DNA (major and minor satellites, respectively) and telomeres was validated by their location on mitotic chromosomes (Supplementary Figure S15B and (55)). Major satellite pericentromeric repeats were visualized as large clusters of DNA, co-localizing with chromocenters, marked by bright DAPI counterstaining, and mainly associated with the nuclear periphery and nucleoli in the 3D nucleus (Supplementary Figures S16A, S17A and Supplementary Table S10). Strongest overlap with PCNA or EdU was observed in S-phase stage II (Figure 4B, Supplementary Figures S16A, S15A, S18A–B and Supplementary Table S10) in line with our results above from *in vivo* and pulse-chase experiments (Figure 1). Minor satellite centromeric repeat signals were seen as small focal structures located in close proximity to chromocenters (Figure 4B, Supplementary Figures S16B, S17B, S18B–C and Supplementary Table S10). Akin to major satellite repeats, centromeric DNA of mES cells was found to also replicate preferentially in early/mid S-phase (stage II). Centromeres of *Drosophila* as well as yeast, were found to replicate only early in S-phase suggesting that early replication timing is a conserved feature of centromeres (97,98). In mouse fibroblasts, however, centromere replication was also reported to occur throughout S-phase (99,100). Given their physical location along chromosomes and their 3D organization in interphase nuclei (Supplementary Figure S15A and (72)), it is likely that centromeric regions are also less compacted in mES cells so that partial overlap of the replication timing profiles of these two structures can be expected. Furthermore, and as shown in somatic cells (99), centromeric repeats likely replicate just before or after the directly adjacent pericentromeric chromosomal domains. In the case of telomeres, the FISH signals distributed as smaller individual foci throughout the cell nucleus (Figure 4B, Supplementary Figures S16C, S17C, S18A, S18C and Supplementary Table S10). Overlap of replication and telomere signals was observed at all S-phase stages, with an increase during stage II.

Although our previous observations (Figure 1A and C and Supplementary Figure S11) suggest a difference in chromatin clustering in S-phase stage III and IV, we could not identify any specific tandem repeat sequence underlying these stages. Since none of the RFi features quantitatively analyzed (Supplementary Figure S12) showed significant differences between these two stages, we consider this

part of S-phase as one substage and will refer to it, hereafter, as stage III.

Since the J1 mES cells were derived from the inner cell mass of a male blastocyst (45), we additionally analyzed the replication timing of the Y chromosome (Supplementary Figure S15C). The strong accumulation of replication signals observed in stage IV<sub>7</sub> of S-phase showed a significant overlap with FISH signals specific for the Y chromosome (Figure 4C, Supplementary Figures S16D, S17D, S18A–C and Supplementary Table S10). We, therefore, consider S-phase stage IV<sub>7</sub> as the male specific S-phase stage where the Y chromosome is replicated and will refer to it, hereafter, as stage Y.

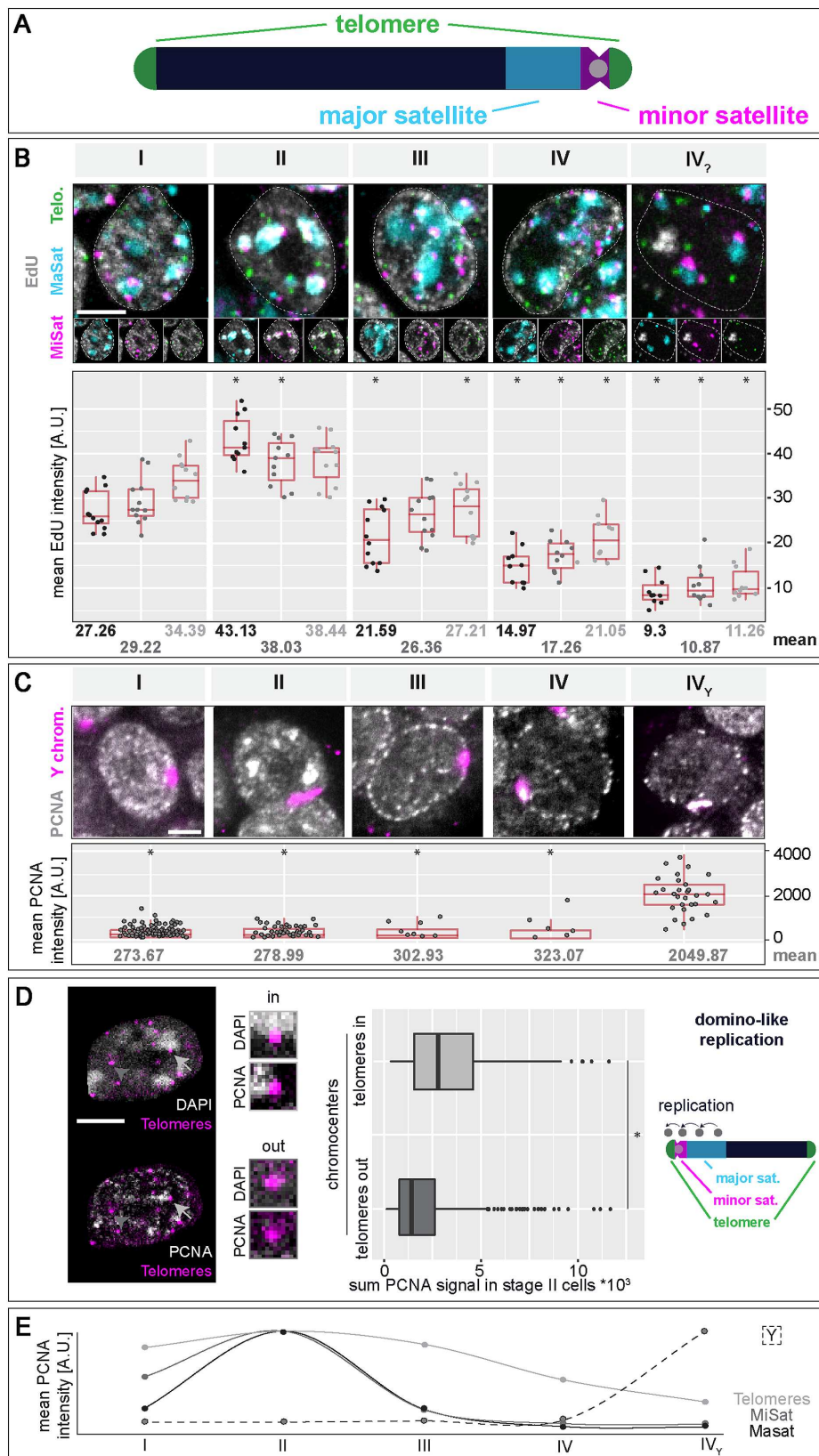
Since we measured an increase in PCNA-telomere signal overlap in S-phase stage II cells, we analyzed if telomere replication timing was dependent on their 1D proximity to pericentromeric heterochromatin, which is replicated during this stage of S-phase. We, therefore, segmented telomeres in stage II S-phase cells, grouped them according to their location with regards to chromocenters and measured PCNA signal intensities within the segmented regions. Telomeres on the short arm of acrocentric chromosomes (telomeres in chromocenters) showed higher PCNA signals in stage II cells than telomeres capping the long chromosomal arms (telomeres out of chromocenters, Figure 4D and Supplementary Table S10). This observation underlines a potential domino-like replication where the activation of origins of replication takes place in a next in-line manner, thereby spreading from the (peri-)centromeric repeats towards the telomeres located on the short arm of the chromosome (99). This is in line with earlier studies relating the replication timing of telomeres to nuclear position, with telomeres positioned towards the nuclear interior replicating earlier than the ones associated with the nuclear periphery (101).

In summary, we characterized the replication timing of the three main classes of tandem repeat sequences and of the Y chromosome via a single cell microscopic approach. (Peri)centromeric DNA was mainly replicated during stage II of S-phase, while telomeres were replicated over the complete duration of S-phase. Replication of the Y chromosome marked the end of S-phase and corresponded to the strong accumulation of replication signals in stage Y (Figure 4E).

### Synchronous replication of the Y chromosome marks the end of S-phase in pluripotent and differentiated cells

The Repli-FISH method allowed us to determine that the prominent structure that is replicated at the end of S-phase in stage Y in male J1 mES cells is the Y chromosome (Figure 4C). In female cells, the inactive X chromosome (Xi) replicates in a highly synchronous manner and, in contrast

(GBP) recruits the GFP-HDAC1 fusion protein to chromocenters via binding of the GBP to GFP. In the control experiments, only GFP was recruited to chromocenters. (B) Histone acetylation (H4K5ac and H4K8ac) levels in (HDAC1) targeted and GFP control cells (mean  $\pm$  StDev). (C) 24 h after double transfection of J1 mES cells, cells were pulse chased as described in Figure 1. Representative spinning disk confocal images from GFP control cells showing S-phase substage transitions from stage II to III and from stage III to IV and from HDAC1 targeted cells showing transitions from stage II to III and from stage 'III' to 'II'. Line profiles represent fluorescence intensities along the arrow marked in the images. CC = chromocenter, perinuc. = perinucle(ol)ar. (D) Percentages of cells representing stage II to III and stage 'III' to 'II' transitions in control and targeted cells (mean  $\pm$  StDev). (E-F) Chromocenter volume (E) and chromocenter shape factor (F) in control cells and cells showing stage II to III and stage 'III' to 'II' transitions in HDAC1 targeted cells. \* $P < 0.05$  and n.s. = non-significant. Detailed statistics are summarized in Supplementary Table S8. Scale bar = 5  $\mu$ m.



**Figure 4.** Replication timing of (sub)-chromosomal structures in mouse embryonic stem cells. (A) Schematic representation of a mouse acrocentric chromosome. Centromeric satellite regions (MiSat) and flanking pericentromeric DNA (MaSat) are depicted in magenta and cyan, respectively and telomeres

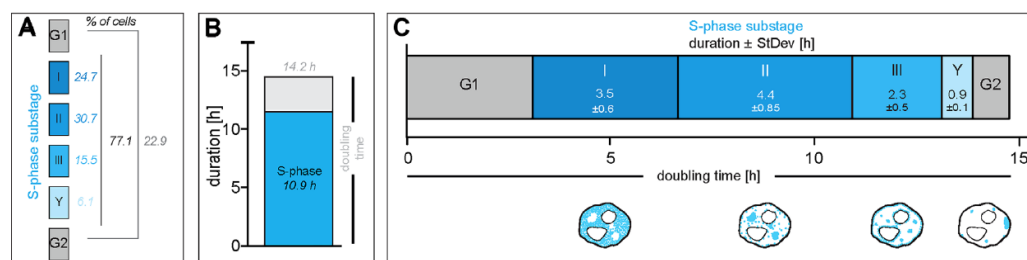
to the active homologue, within a short time interval during early-mid S-phase (31). Similarly, we did not detect any major replication signal at the Y chromosome in any of the other four S-phase stages and, therefore, conclude that the Y chromosome is synchronously replicated within this short period of S-phase stage Y (Supplementary Figure S8D). To further determine if this *en bloc* replication of the Y chromosome at the end of S-phase is a characteristic of pluripotent mES cells or a general feature of male cells, we analyzed the PCNA and Y chromosome hybridization signal overlap in replicating male mouse embryonic fibroblasts (MEF W8). In addition to the three well characterized S-phase patterns known from somatic cells ( $S_e$ ,  $S_m$  and  $S_i$ ), MEF cells showed a fourth pattern with a clear accumulation of replication sites at the Y chromosome (Supplementary Figure S19A and Supplementary Table S10). Pulse chase experiments in differentiated mES cells and mouse embryonic fibroblasts showed a similar pattern after chromocenter replication during late S-phase ( $S_i$ ) and before transition to G2 phase (Supplementary Figure S19B). The Y chromosome of male cells, one of the smallest chromosomes in mice (~92 Mb), is mostly studied in the context of evolution, clinics and forensics (102) and mainly believed to consist of non-functional DNA (103). At the genomic level, the Y chromosome is marked by a very low gene density (1.7 genes/Mb for mice, (74,104,105)) of which only a fraction appears to be potentially protein coding in humans and that are mostly required for testis development and sex determination, or have a X encoded paralog (105,106). Additionally, the Y chromosome is composed of large heterochromatic DNA blocks (107,108). Therefore, gene expression and transcriptional activity on the Y chromosome are low and a connection between low gene density and transcriptional activity could explain the late replication timing of the male sex chromosome. Accordingly, a correlation between gene expression and replication timing was demonstrated in *Drosophila* (109) and, similarly, between replication timing, GC content, gene density and transcriptional activity in human cells (68,110–112). Besides this, the Y chromosome is frequently lost in most male cell lines during prolonged cell culture and, thus, neglected in most studies. Our microscopic data provide evidence for a synchronous replication of the Y chromosome marking the end of S-phase independently of the pluripotency state as a general characteristic of mouse male cells. Interestingly, mid S-phase replication of the silenced copy of the X chromosome in somatic female cells, also occurs in a synchronous manner (31), highlighting a common repli-

cation mode for transcriptionally inactive chromosomes. This replication mechanism bears resemblance to observations in early *Drosophila* and *Xenopus* embryos, where genome duplication is performed in extraordinarily short time frames and in the complete absence of transcription (35,113–116).

#### Cell cycle and S-phase stage kinetics in mouse embryonic stem cells

Embryonic stem cells have the unique characteristics to replicate indefinitely in cell culture while maintaining their self-renewal capacity and to differentiate into cells of all three germ layers. Their cell cycle also differs significantly from somatic cell types (117) and upon differentiation, cell cycle dynamics undergo massive reorganization (118–120). To study DNA replication kinetics in mouse J1 ES cells, we analyzed total S-phase and S-phase substage lengths, as well as the population doubling time in asynchronously growing J1 mES cells. Cell cycle distribution was analyzed by counting S-phase and non S-phase cells (G1, G2 or mitotic cells) in EdU pulsed cell populations. On average, around 77% of J1 mES cells were in S-phase at any given time, while 23% were not actively replicating DNA (Figure 5A and Supplementary Table S11). Furthermore, around 31% of S-phase cells were in stage II, replicating pericentromeric heterochromatin, while 25% showed the stage I pattern. The other stages were represented by 20% of the cells. Cell proliferation rates were calculated over five consecutive days and J1 mES cells exhibited a doubling time of around 14 h (Figure 5B), which is in agreement with doubling times measured for other mouse ES cell lines (121). Together with the percentage of cells in S-phase (Figure 5A and Supplementary Table S11), an average S-phase duration of almost 11 h was calculated (Figure 5B). Similarly, S-phase substage durations were derived (Figure 5C). During almost two-thirds of S-phase length, mES cells were present in stage I or II, replicating euchromatic and constitutive heterochromatin. Stage III and Y accounted for only 1/3 of S-phase length. These results are comparable to the S-phase substage durations we measured via live cell microscopy (Supplementary Figure S8D). All in all, J1 mES cells exhibit a similar S-phase duration to somatic cells, however, their doubling time is significantly shorter (28,120). This observation is in line with previous reports that stem cells have shorter gap phases and are devoid of a G1-S transition regulation (118). Strikingly, although mES cells exhibit a different temporal organization of DNA replication, around 72% of S-phase

are shown in green. (B) Tandem repeat elements were co-visualized with EdU (labeling of nascent DNA, grey) in mES cell interphase nuclei by triple FISH hybridization. Cells were classified into S-phase stages I to IV<sub>7</sub> according to their EdU pattern, mean EdU intensities within the marked elements were measured as described in Supplementary Figure S3 and plotted. Mean values are indicated below each plot. (C) Analysis of Y chromosome FISH in combination with PCNA staining was performed as in (B). (D) Analysis of telomere replication timing in J1 mES cells co-stained for PCNA and telomeres. Chromocenters of S-phase stage II cells were segmented according to the DAPI staining, and sum values of PCNA fluorescent intensity were measured within segmented telomeres located in close proximity to chromocenters (light grey arrows, 'telomeres in chromocenters') and within telomeres located on the long arm of the chromosome (not in proximity of chromocenters, dark grey arrow heads, 'telomeres out of chromocenters'). Replication of chromocenter near telomeres within S-phase stage II hints towards a domino-like replication model with a sequential order of replication of adjacent chromosomal regions (MaSat/MiSat) to telomeres on the short chromosome arm). (E) Summary of the replication timing of tandem repeat elements and the Y chromosome in mES cells. (Peri)centromeric DNA regions (MaSat and MiSat) are mainly replicated within the first half of S-phase (stage II), telomeres are replicated throughout S-phase and the Y chromosome marks the end of S-phase (stage Y). Dotted lines represent cell contours. Boxplots are as in Figure 2 and Supplementary Figure S7. Scale bar = 5  $\mu$ m. Detailed statistics are summarized in Supplementary Table S10. \* $P < 0.05$  (calculated among each elements against the respective stage I value (B) or against stage Y in C).



**Figure 5.** Cell cycle characteristics of mouse embryonic stem cells. (A) The cell cycle distribution within an asynchronous mES cell population was analyzed by labeling the cells with EdU and counting the number of S-phase cells (EdU positive). Around 77% of the cells are in S-phase and additionally, the percentage of cells within S-phase substages (I–Y) is detailed. (B) Growth curve analysis over 5 days revealed a population doubling time of around 14 h for mouse J1 ES cells. Together with the percentage of replicating cells from (A), an approximate S-phase duration for mES cells of about 10.9 h was calculated. (C) From the fraction of cells within every S-phase substage (I–Y, A) and the total S-phase duration (B), approximate durations of the individual substages were calculated (mean  $\pm$  StDev). All experiments were done in at least three independent biological replicates. Detailed statistics are summarized in Supplementary Table S11.

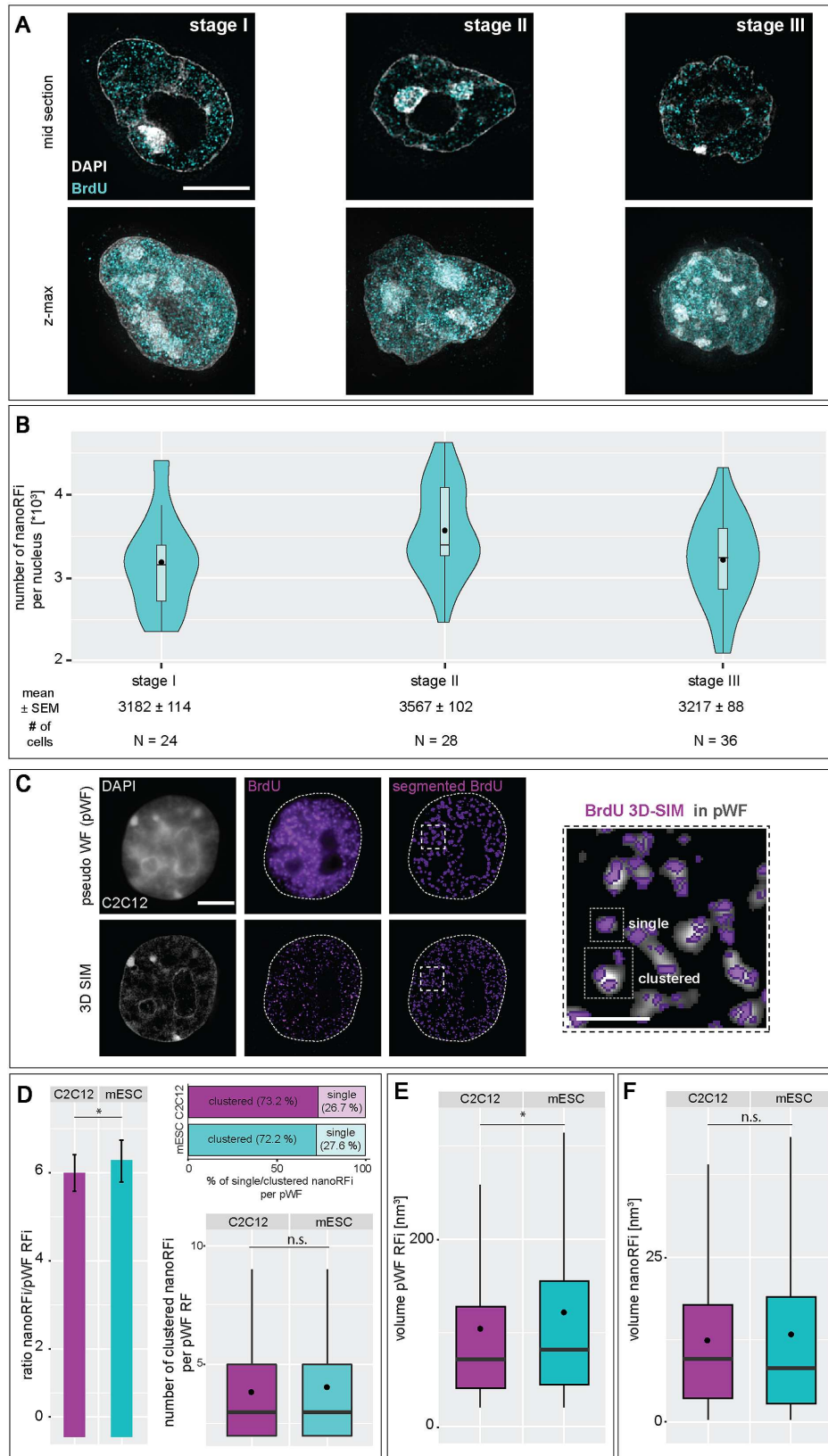
are necessary to replicate active euchromatic chromatin and silenced constitutive heterochromatin (stage I and II), which is similar to the time mouse myoblasts and embryonic fibroblasts spend in  $S_c$  and  $S_i$  (20)

#### Superresolved replication nanofoci in mouse embryonic stem cells are activated in spatial clusters similar to somatic cells in number but larger in volume

Advanced optical microscopy techniques allow imaging beyond the resolution limit inherent to light microscopy. By using multicolor 3D structured illumination microscopy (3D-SIM), it was recently shown that individual replication foci are resolved down to single replicons and, to some extent, even to individual replication forks (28). Chromatin organization and chromosomal interactions are known to change during embryonic stem cell differentiation (122). Moreover, differences in the organization of the underlying chromatin fiber, e.g. a more open chromatin conformation, may have an impact on the organization and activation of replication origins, and, in turn, on the regulation of replication timing of different chromatin classes. We therefore analyzed replication foci at the different S-phase substages in mouse ES and mouse myoblast (C2C12) cells imaged with 3D-SIM (nano replication foci or nanoRFi). To quantify the number of super-resolved nanoRFi, mES cells were labeled with 5-bromo-2'-deoxyuridine (BrdU), the nucleotide analog was immunofluorescently detected and S-phase cells were imaged. Replication foci were thresholded, masked and counted as described (53). Comparable numbers of nano replication foci for cells from the first three S-phase substages were obtained, with on average  $3320 \pm 60$  (mean  $\pm$  SEM) nanoRFi at any given time during S-phase (Figure 6A, B and Supplementary Table S12). Since the number of nano replication foci dropped dramatically at the end of S-phase (stage Y:  $1881 \pm 214$ ), we excluded this stage from all subsequent calculations. A similar trend was found in the number of RFi measured from confocal microscopy images (Supplementary Figure S12). NanoRFi measurements in somatic cells revealed a decreased number of super-resolved foci in late S-phase cells, when highly compacted constitutive heterochromatin is replicated (28). Interestingly, we detected comparable numbers of nanoRFi in

mES stage II S-phase cells and all other S-phase stages. This discrepancy could be explained by the more decompacted heterochromatin in mES cells (Supplementary Figure S14), facilitating individual foci segmentation. Additionally, this result supports the idea that differences in chromatin organization influence DNA replication (78,79,123).

Since origins of replication are proposed to be located at loop anchors (26) and based on the hypothesis that a different chromatin compaction and (loop) organization in mES cells could result in a spatially different organization of nano replication foci and replicons, we compared the numbers of nanoRFi in a given volume within the cell nucleus of mouse myoblast and embryonic stem cells. The 3D-SIM system simultaneously allows the generation of reconstructed super-resolved 3D image sets and the corresponding (pseudo)wide-field (pWF) images (Figure 6C). This allowed us a direct comparison of the total number of RFi per cell from different imaging resolution conditions from the same cells. We, therefore, segmented 3D-SIM nanoRFi and pWF RFi from early and stage I C2C12 and mES cells, respectively, and calculated the population ratio of RFi. The two cell types showed a similar ratio indicating that every RFi imaged at conventional light microscopy resolution corresponds on average to 5.9 and 6.5 super-resolved nanoRFi for C2C12 and mES cells, respectively (Figure 6D and Supplementary Table S12). In addition to the ratiometric approach, we correlatively analyzed the connection of pWF and nanoRFi by determining the number of super-resolved RFi per underlying pWF focus in both cell types. Hence, we first aimed to determine the amount of clustered ( $> 1$  nanoRFi per pWF) versus single (exactly 1 nanoRFi per pWF) nanoRFi within the given volume of the pseudo-widefield focus. Interestingly, in both cell lines, around 72–73% of the nanoRFi formed clusters, while 26–27% of the pWF contained only one nanoRFi. While the above mentioned ratiometric approach includes all detected and segmented pWF, independent of the number of underlying nanoRFi, the correlative analysis takes into consideration only pWF foci containing more than one nanoRFi (clustered nanoRFi). These pWF replication foci comprised clusters of on average 3.82 and 4.02 nanoRFi for C2C12 and mES cells, respectively (Figure 6D and Supplementary Table S12). Notably, stem cell pWF RFi are significantly



**Figure 6.** 3D quantification and analysis of replication foci throughout S-phase in mouse embryonic stem cells. (A) Mid sections and maximum intensity  $z$ -projections ( $z$ -max) of 3D structured illumination microscopy (3D-SIM) images of mouse ES cells representative of the first three S-phase patterns (I–III)

larger than myoblast pWF foci while no difference in size of nanoRFi could be detected, suggesting that they correspond to conserved (elementary) structural units of the mammalian genome (28) (Figure 6E and F and Supplementary Table S12). In summary, the observed difference of mES cell replicons from those of mouse myoblast cells, suggests that the underlying chromatin and loop organization is different in mES cells (122,124) and that this difference is reflected in the organization and dynamics of DNA replication (78,79,123). Chromatin loop structures are considered distinct units in the complex hierarchy of genome organization within the cell nucleus. At this level, individual loops may harbor the DNA elements that act as templates or regulators in different molecular processes (reviewed in (27)). The most prominent examples are enhancer elements and promoters of genes that coordinate and regulate transcription via long ranging cis-interactions (94,125,126). Notably, in somatic cells, measured sizes for chromatin loops, replication forks and also nano-repair foci are highly consistent and rely on the same structural DNA unit of about 90 kb (28,60,127). Coherently, a replicon consisting of bidirectional replication forks, duplicates a DNA segment the size of a pair of loops, i.e. around 180 kb (28). Thus, differences in loop conformation and (local) chromatin density are potent features that can modulate interactions between genetic elements that are required for DNA-dependent metabolic processes such as DNA replication. In this regard, it is appealing to propose that if replication origins are defined by loop anchoring sites, then organization of replicons will be directly linked to loop structures.

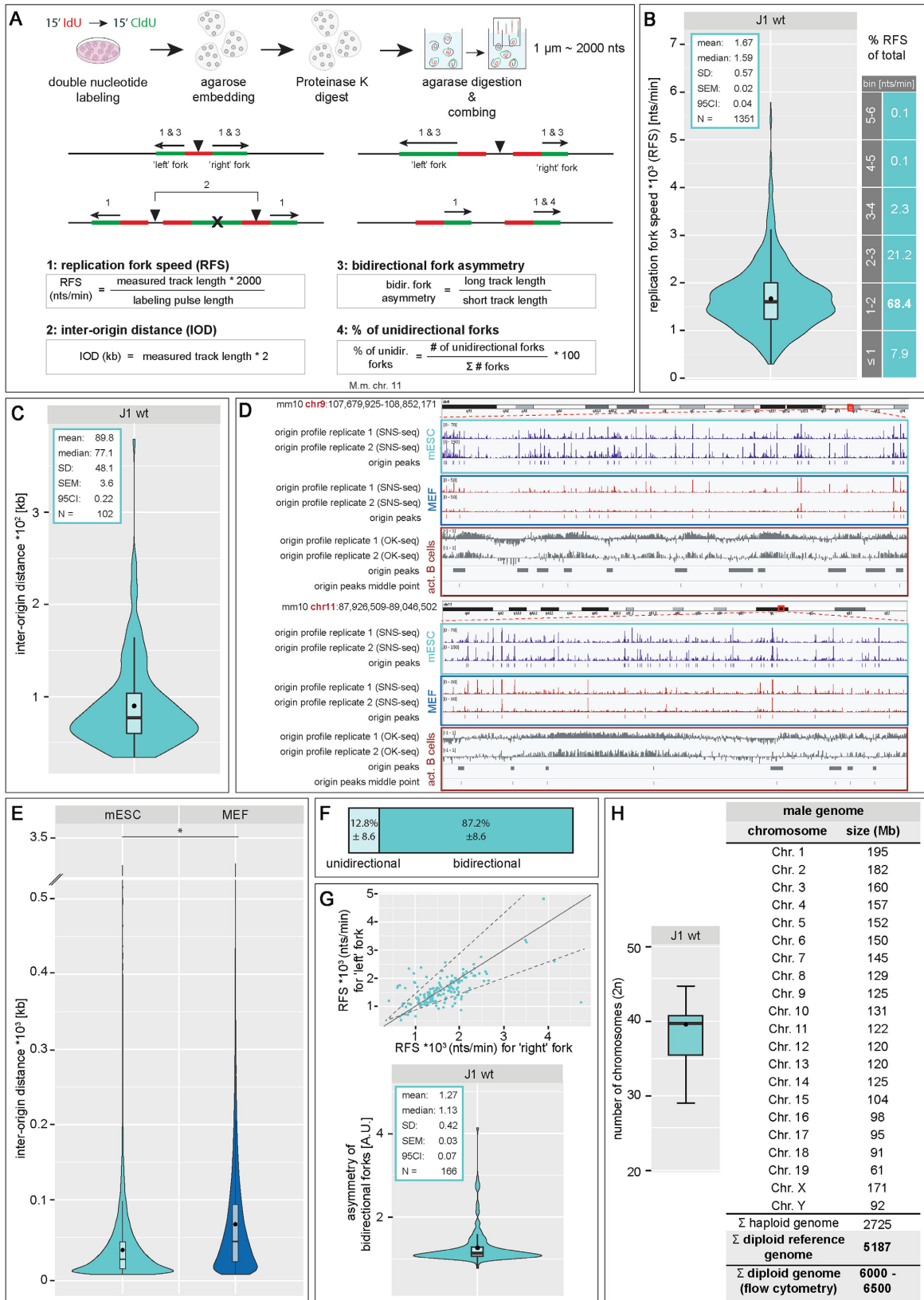
#### Molecular characteristics of the replicon in mouse embryonic stem cells reveal increased numbers of activated origin and unidirectional forks

Complete genome duplication once per cell cycle and in the restricted time frame of S-phase depends on two determinants: (i) the number and distribution of initiation sites (i.e. replication origins) along the genome (inter-origin distance) and (ii) the processivity rate (nucleotides/time) of replication forks (replication fork speed) emanating from these sites. With regard to the differences in the spatio-temporal replication dynamics of mouse ES cells described above and in view of the highly dynamic spacing of adjacent origins within replicon clusters in *Xenopus* and *Drosophila* development (36,37,113), we aimed to clarify if mES cells exhibit further replication related differences/adaptations compared to somatic cells. To analyze the molecular char-

acteristics of replicons in mES cells, i.e. the replication fork (elongation) speed (RFS) and the inter-origin distance (IOD), we consecutively labeled asynchronous J1 mES cell populations with two halogenated thymidine analogs, IdU and CldU and performed molecular combing assays with isolated high molecular weight DNA that led to uniformly stretched DNA fibres (1  $\mu\text{m} \sim 2$  kb, Figure 7A). The replication fork speed obtained in mES cells of  $1.67 \pm 0.02$  kb/min (mean  $\pm$  SEM, Figure 7B) is within the range of fork rates measured by the same single molecule DNA fiber analysis for other cell lines (28,128). This suggests that replication fork speeds are similar between pluripotent and somatic cells. On the other hand, the average inter origin distance of  $\sim 90$  kb (Figure 7C) is smaller in mES cells compared to mouse myoblasts and human cells (28). This result indicates that the organization of replicons or replicon clusters is different between mouse pluripotent and somatic cells. Although the observed differences in IOD lengths between mES and differentiated somatic cells are not as dramatic as those occurring during development of *Xenopus* embryos, they suggest that the modulation of inter-origin distances, and concomitantly the resulting replicon sizes, represent a mechanism that is similar between the two species and highlights further developmental differences of the replication timing program of murine ES cells. Of note, while the first cell divisions in the *Xenopus* zygote occur in the absence of DNA transcription, transcription initiation is observed in the pronuclei of the zygote and at the two-cell stage during mouse development (35–37,39,41,129).

To get a deeper insight in the distribution of genome-wide replication initiation events, we compared the results of our microscopic analysis with the available datasets of sequencing-based genome-wide origin mapping in mES cells (61,62), mouse embryonic fibroblast (MEF) cells (62) and activated mouse B cells (64). This reflects the origin distribution in three different stages of differentiation, from stem cells through differentiated but still reprogrammable MEF cells, to a terminally differentiated B cell line. The origin mapping in mES and MEF cells was realized using short-nascent-strand (SNS-seq) isolation and sequencing (130) and origins in B cells were mapped by Okazaki fragment sequencing (OK-seq) (131). The genome-wide origin mapping was realized using asynchronous cell populations that permit cumulative mapping of all replication origins independently of their timing of firing. A total number of 78 238 and 71 257 DNA replication origins were identified in two independent datasets using mES cells (61,62). In MEF, 34 196 DNA replication origins were identified, i.e.,

are shown. (B) Numbers (mean  $\pm$  SEM) of nano replication foci (nanoRFi) quantified as described in Supplementary Figure S4 are plotted separately for each of the three S-phase patterns. N indicates the number of cells analyzed. (C) At lower optical resolution, replication signals appear as larger foci (pseudo wide-field (pWF) foci). These can be resolved to a number of smaller foci when imaged by super-resolution microscopy. Shown are representative pWF (upper row) and the respective 3D SIM images (lower row) of the cell nucleus of a mouse myoblast, the unsegmented (middle column) and segmented (right column) BrdU replication signals. The pWF replication foci were segmented as described in Supplementary Figure S5 and used to demarcate a distinct volume of DNA in which the number of nano replication foci (nanoRFi) was quantified (magnified inset). nanoRFi are considered as 'clustered' if one pWF focus contains more than one nanoRFi. (D) Results of cluster analysis of nanoRFi within the distinct volume of a pWF replication focus are shown. RFi ratios from super-resolution versus pseudo-widefield microscopy (barplot  $\pm$  Stdev) and analysis of the number of clustered nanoRFi in individually segmented pWF foci (boxplot) in early S-phase mouse ES and myoblast (C2C12) cells are shown. Percentages of single and clustered nanoRFi are depicted. (E-F) Volumes of the segmented pWF (E) and 3D-SIM (F) nano replication foci (nanoRFi) are shown. Detailed statistics are summarized in Supplementary Table S12. Boxplots are as in Figure 2. \*  $P < 0.05$  and n.s. = non-significant. Black dots within violin/box plots represent mean values. Scale bar = 5 and 2.5  $\mu\text{m}$  for main graphs and magnified regions, respectively. Brightness and contrast of 3D-SIM images were adjusted for every image depicted. Dotted lines represent cell contours.



**Figure 7.** DNA replication fiber and genome size analysis in mouse embryonic stem cells. (A) Schematic outline of the experimental setup for DNA fiber analysis. mES cells were sequentially labeled with IdU and CldU for 15 min, harvested and embedded in agarose. After a proteinase K digestion step,

two times less origins than in mES cells. The two times lower number of origins led to a two times larger inter-origin distance in MEF cells in comparison to the mES cells (Figure 7D and Supplementary Figure S20). We determined average IODs of 33.6 and 69.93 kb for mES and MEF cells, respectively (Figure 7E and Supplementary Table S13). Interestingly, the IOD calculated for the somatic chromosomes was two times smaller than for the X chromosome (62 in mES and 82 kb in MEF) (Supplementary Figure S21A-C). This suggests a different spatial origin activation on the two parental homologue chromosomes leading to an apparently smaller IOD calculation than the IOD measured for a single stretched DNA molecule in the combing experiments. The decrease in origin number and increase in the IOD became even more prominent in the terminally differentiated B cell line activating around 9000 origins. The SNS-seq and OK-seq differ in the origin-mapping precision, the SNS-seq method offers a very good mapping resolution and identifies all possible initiation sites (ISs). With an average peak size of 0.5 kb in the datasets analyzed (Supplementary Figure S22A and Supplementary Table S13), most of the ISs are situated in close proximity and represent alternative origin firing patterns within a cell population. The OK-seq identifies large initiation zones (IZs) with an average resolution of 22 kb (Supplementary Figure S22A) without distinguishing single initiation sites. To equalize the resolution of the two methods for comparison purposes, all SNS-seq initiation sites were clustered in the distance of the average resolution of OK-seq origin mapping (22 kb). As a result of this operation, all origins found in mES and MEF cells were clustered in 33 765 and 23 862 border IZs, situated at an average distance of 70.8 and 100.2 kb, respectively. In contrast the 9000 IZs identified in the activated B cells were spaced at the distance of 288 kb (Supplementary Figures S20 and S22B). This result suggests a gradual decrease in origin firing and increase in origin spacing during differentiation and supports our finding of smaller inter-origin spacing in mES cells compared to somatic cells obtained from single molecule DNA combing data.

Next, we analyzed the portion of single/unidirectional forks (replication forks without counterpart/opposite direction fork) per total forks present in mES cells and observed that around 13% of all forks migrate away from the origin of replication in only one direction (Figure 7F and Supplementary Figure S6). The presence of unidirectional

**Table 1.** mES cell replicon characteristics

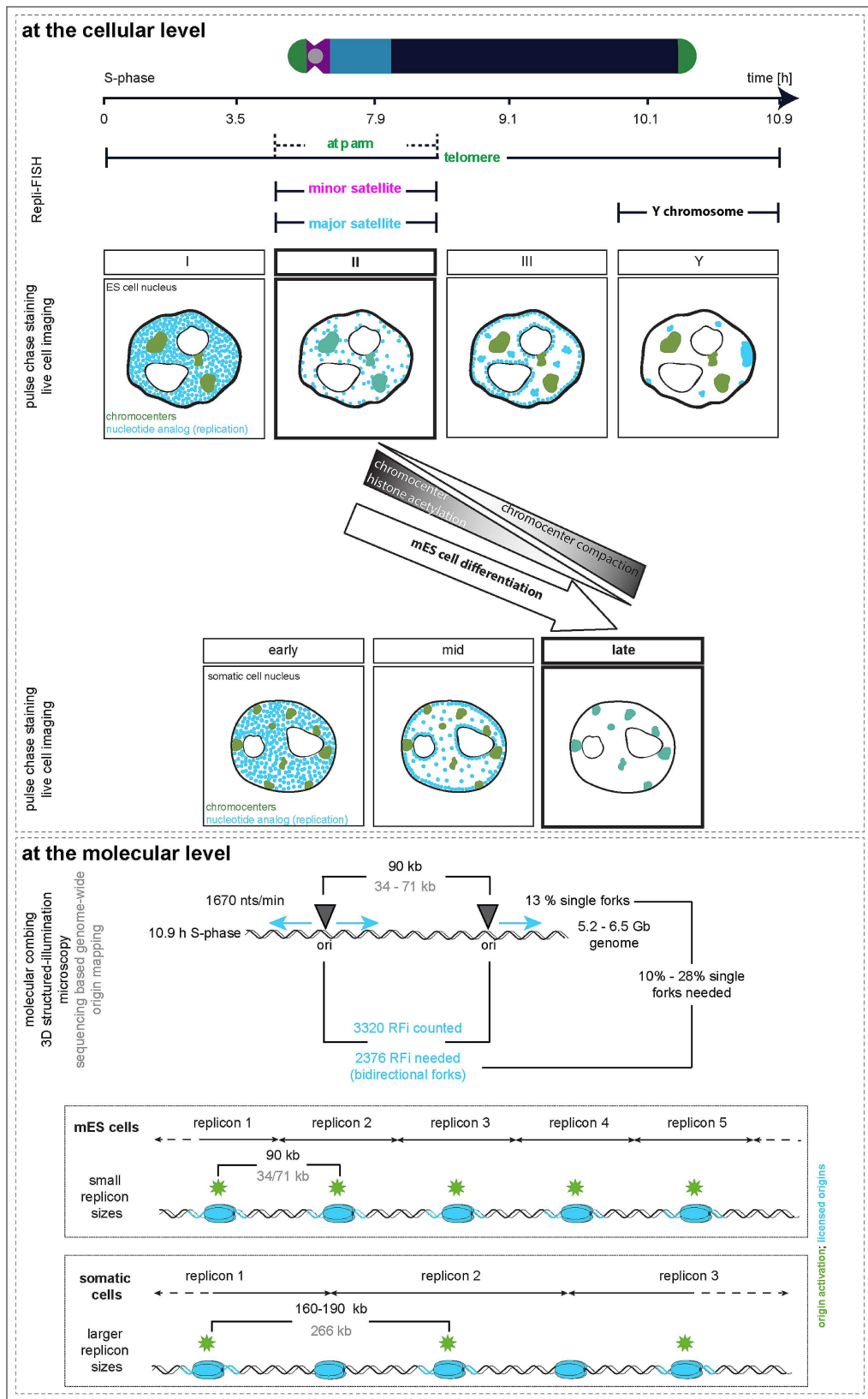
Experimental data	Mean $\pm$ SEM
RFS, 1000 nts per min	1.67 $\pm$ 0.02
IOD, kb	90 $\pm$ 3.6
Genome size (GS), 1000 Mb	5.19*–6.5**
Active RFi at any given point during S-phase	3 320 $\pm$ 20
Total S-phase duration, minutes	654 $\pm$ 21
Single forks, %	12.81 $\pm$ 8.6
Calculations	Mean
Time to replicate the genome with one fork, (GS/RFS), hours	51 796–64 870
Replication forks active in parallel, (GS/RFS/S-phase duration)	4 752–5 951
Replicons active in parallel, (active forks/2)	2 376–2 976
Replicons per RF, (calculated replicons active in parallel/counted RFi)	0.72–0.9
Calculated % of single forks	10–28

RFS, replication fork speed; nts, nucleotides; IOD, inter-origin distance; GS, genome size (based on published genome data); RF, replication focus; RFi, replication foci; SEM, standard error of the mean; \* GS according to mouse reference genome (GRCm38); \*\* GS according to flow cytometry measurements.

forks has also been found in sequencing-based genome-wide origin mapping techniques in human cells. In the latter, genome duplication relies on 4.1% to 7.3% of unidirectional forks (131). We obtained similar numbers upon extracting the amount of unidirectional forks from the DNA combing analysis datasets of Chagin *et al.* (28), where we measured  $5.5 \pm 1.3\%$  of forks travelling in only one direction from the origin of replication. In addition, we also analyzed the (a)symmetry of bidirectional forks and found that the two forks proceeding in opposite directions from the same origin mostly travel at similar rates (Figure 7G). This suggests that the replication forks of mES cells do not experience a high level of stalling events due to exogenous or endogenous factors, albeit mES cells are transcriptionally hyperactive (132).

In view of the above mentioned essential parameters for genome duplication once every cell cycle, the karyotype and the resulting genome size of a cell line is important for the determination of replicon characteristics. Most established cell lines used for *in vivo* studies have initially been transformed in order to immortalize them and to revive

agarose was digested and high molecular weight naked DNA was stretched on silanized glass coverslips. Nucleotide analogs and single stranded DNA (ssDNA) were immunofluorescently detected. (B–E) The length of fluorescent tracks of the second pulse (CldU) were measured (1  $\mu\text{m} \sim 2000$  nucleotides, Supplementary Figure S6) and the mean replication fork speed (RFS, (B)), inter-origin distance (IOD, (C)), percentage of unidirectional forks (F) and asymmetry of bidirectional forks (G) were calculated as indicated in (A). Additionally, the percentages of forks within a given range of RFS are indicated in (B). For comparison of the RFS of the 'left' and 'right' fork of a bidirectional fork, RFS values were plotted in a scatterplot. The solid grey line represents the linear relation  $x = y$  and dotted lines represent thresholds allowing for a 35% ( $\pm$  StDev calculated for the asymmetry factor) difference between lengths of the two forks. (D) Visual representation of origin mapping in two arbitrarily selected regions (mouse (*Mus musculus*, mm10) chromosomes 9 and 11). SNS-seq origin profiles and identified origins in mES and MEF cells are shown. OK-seq origin profiles in activated (act.) B cells, called peaks along with the middle point of each peak are represented. Replication profile scale is indicated in the upper left corner. The comparison between identified and clustered origin peaks is shown in Supplementary Figure S20. (E) IOD distributions based on the genome-wide origin maps in mES cells and mouse embryonic fibroblast (MEF) are shown (SNS-seq). The sequencing datasets used for the analysis include two independent replicates of origin mapping for each condition. (H) Ploidy of J1 mES cells was determined via karyotype analysis of metaphase spreads. For genome size calculation, the sizes of individual mouse chromosomes (19 autosomes + X and Y chromosomes) were retrieved from the Genome Reference Consortium database. Additionally, genome sizes measured by flow cytometry are indicated. Boxplots/violinplots are as in Figure 2 and Supplementary Figure S7. Statistical details are depicted in the plots or summarized in Supplementary Table S13. All experiments were done in at least two independent biological replicates. Black dots within box/violin plots represent mean values. \*  $P < 0.05$ . Scale bar = 5  $\mu\text{m}$ .



**Figure 8.** Graphical summary of the cellular and molecular DNA replication characteristics in mouse embryonic stem cells. At the cellular level, DNA replication is visible as distinct replication foci with a dynamic spatio-temporal organization. In mES cells, three different replication patterns are observed

their replicative potential. Such procedures, however, often introduce undesired genetic aberrations which are known to drastically influence the ploidy of the transformed cell line. Embryonic stem cells, on the other hand, are derived from the inner cell mass (ICM) of the blastocyst and are, by nature, capable of sustaining their proliferative state in culture. Moreover, they have been described to maintain a stable diploid karyotype (133). Nonetheless, we performed karyotype analysis of J1 embryonic stem cells from metaphase chromosome preparations and calculated the approximate genome size taking advantage of published genome data available from the Genome Reference Consortium (92) for haploid mouse genomes. Manual counting of >100 metaphase spreads confirmed a diploid karyotype of the mouse J1 cell line, consisting of 40 acrocentric chromosomes (Figure 7H and Supplementary Table S13). To derive the total genome size, we used the mouse genome assembly GRCm38.p6 mm10 that provides sequencing derived sizes for each chromosome of a haploid mouse genome. Based on this, we calculated a male diploid genome size for mES cells of 5.182 Gb (Figure 7H). Since genome-wide sequencing approaches are affected by unmappable repetitive sequences, this likely is an underestimate of the actual genome size. Indeed, using data from flow cytometry based approaches we obtained diploid mouse genome sizes of 6.03–6.5 Gb (134,135).

#### Model for genome replication in embryonic stem versus somatic cells

Using the above mentioned characteristics of mES cell DNA replication and molecular parameters of the associated replicons (i.e. genome size, S-phase duration, number of nanoRFi, IOD and RFS, see Table 1), we analyzed the relationship between replicons and replication foci in mES. The total number of replicons, reflecting the number of origins activated during S-phase, is given by the genome size divided by the average inter origin distance obtained from DNA fiber experiments. This calculation resulted in a total of 57 700–72 300 origins activated during the S-phase of mES cells, depending on how the genome size was estimated. These values are comparable to the number of replication origins obtained by genome-wide origin mapping techniques (see above and (61,62)). Since IOD measurements may be more affected by the sample quality and the length of DNA fibers obtained during the DNA isolation and combing procedure (128), we additionally focused on RFS to determine the number of active replication foci at

any given time and to compare it with the actual numbers of foci active in parallel obtained from our 3D-SIM data. The genome size divided by the average speed of a replication fork represents the time required to synthesize the entire genome if only one fork would be active for the entire length of S-phase. Dividing this time by the actual measured S-phase duration, thus effectively represents the number of all active replication forks required at any given time during S-phase. These calculations showed that ~4750–5950 forks or half as many bidirectional replicons (~2380–2980) are required to act in parallel in mES cells (Table 1). Comparing the nanoRFi active in parallel counted in 3D-SIM images to the theoretically ‘needed’ RFi active at any given time during S-phase, resulted in a ratio of 0.72–0.9 (2376/3320–2976/3320 RFi), indicating that mES cells activate more replicons than ‘needed’ for genome duplication within the timeframe of S-phase.

The difference in predicted and measured active replication foci can be explained by a higher frequency of nanoRFi that contain single replication forks and could be resolved by 3D-SIM. The latter would involve the presence of 10–28% of unidirectional forks in mES cells to make up for the difference in calculated versus counted RFi. Indeed, we measured a substantial number of unidirectional replication forks in mES cells using DNA combing amounting to 13% of the total forks detected (Figure 7F). On the other hand, the combination of a less compacted chromatin in mES cells (Supplementary Figure S14) and the resolving power of the 3D-SIM system allows to visualize more individual replication forks in mouse ES cells. Another possibility that could affect the spatial positioning of replication forks or replicons is a different chromatin loop organization in mES cells at sites of ongoing DNA replication, e.g. by subdivision into multiple smaller loops. Hence, we may also detect more individual forks, increasing the number of observed nano replication foci. The differences in heterochromatin compaction in embryonic stem cells (Supplementary Figure S14) are representative of a chromatin organization that is generally more open than that of differentiated cells (78,98). This, in turn, could affect the organization of the DNA fiber, i.e., into chromatin loops. In support of this, it was reported using chromosome conformation capture carbon copy (5C) that the organization of TADs (topologically associated domains) into multiple sub-megabase sized domains (i.e. sub-TAD domains located within TADs) is found in several developmentally regulated genetic loci in mouse ES and neuronal progenitor (NP) cells. Interestingly, in mES cells, a series of on average 100 kb sized loops connects the Sox2 gene

during the ~11 h of S-phase (time progression arrow not scaled). (Sub)chromosomal elements were found to replicate at specific time points during genome duplication. While telomeres located at the long arm of the chromosome replicate throughout S-phase, the ones capping the q arms show significant increase in replication during S-phase stage II. In contrast to somatic mouse cells, (peri)centromeric DNA (marked in green) replicates during mid S-phase (stage II). The Y chromosome is replicated synchronously at the end of S-phase (stage Y) in pluripotent as well as in differentiated cells. Replication timing of pericentromeric DNA switches from early/mid to late S-phase upon mES cell differentiation, which correlates with chromocenter compaction and decreased histone acetylation. At the molecular level, mES cell replicons are characterized by short inter-origin distances of about 90 kb. Replication forks progress at 1.7 nucleotides per minute to replicate the 5.2 Gb mouse genome in about 11 h. Replication is initiated from around 3320 replication foci (RFi), although theoretically ~2380 bidirectional mES cells are sufficient for genome duplication within the timeframe of S-phase. Hence, from the 3320 sites observed, 28% correspond to single forks and the remaining to bidirectional forks. This is within the range of (13%) single forks observed in DNA fiber analysis. A characteristic arising from the molecular parameters of DNA replication in mouse pluripotent cells are smaller replicon sizes in mES cells relative to somatic cells. Along a given segment of chromosomal DNA carrying multiple licensed origins of replication, mES cells initiate DNA replication from double the number of origins compared to somatic cells. This results in more but smaller replicons and concomitantly smaller inter-origin distances (~90 kb in mES cell and 160–190 kb in mouse/human somatic cells).

12774 *Nucleic Acids Research*, 2020, Vol. 48, No. 22

with a presumed downstream enhancer element marked by specific histone modifications. Akin to the loss of these epigenetic marks upon differentiation to NP cells, looping interactions were no longer detected in the latter, suggesting a mES cell specific loop organization (136) with loop sizes reflecting replicon sizes that we determined in this study (~90 kb).

In summary, we analyzed the spatio-temporal dynamics of DNA replication progression in (un)differentiated mouse embryonic stem (mES) cells and compared it with somatic cells. We find a developmental switch in the replication order of main chromosomal domains that is dependent on histone acetylation level (Figure 8). Furthermore, we measured and compared the molecular properties of the mES cell replicon, including the number of replication foci active in parallel and their spatial clustering in mES cells versus somatic cells. We conclude that each replication nanofocus in mES cells corresponds to an individual replicon, with approximately one tenth to one quarter representing unidirectional forks. Furthermore, we find that mES cells activate twice as many origins spaced at half the distance than somatic cells. Altogether, our results highlight fundamental developmental differences on progression of genome replication and origin activation in pluripotent cells.

#### DATA AVAILABILITY

All data/code are available at <https://doi.org/10.25534/tudatalib-220>.

#### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

#### ACKNOWLEDGEMENTS

We thank Manuela Mildner and Diana Imblan for excellent technical assistance. We are indebted to Sylvia Ritter (GSI Helmholtzzentrum für Schwerionenforschung, Darmstadt, Germany) for helping with embryonic stem cell metaphase spreads and to Kai Breitwieser for help with E14 pulse chase experiments. We thank Alexander Rapp, Kathrin S. Heinz, Florian D. Hastert and Anne K. Ludwig for experimental advice and discussions.

**Author contribution:** C.R. performed experiments for Figures 1, 2, 4, 5, S8, S10, S14, S15, S18 and S19. P.W. performed experiments for Figures 1, 3–7, S6, S8, S11, S15–S17 and S19. A.M. and H.L. provided the 3D SIM imaging in Figures 1 and 5. A.L. performed growth curve measurement in Figure 4.

C.R. analyzed data from Figures 1–7, S8, S10, S14, S15, S17–S19. Data analysis from Figures 1, 4–7, S8, S11, S15, S16 and S19 were done by P.W.. P.P. performed genome wide origin mapping analysis shown in Figures 7D, 7E, S20–S22. D.H. performed RFI feature analysis from Figure S12 and S13. V.O.C. contributed to image analysis.

C.R., P.W., C.C.-D. and M.C.C. conceived the project, C.R. generated final figures and C.R. and M.C.C. wrote the manuscript. All authors revised, commented and agreed on the manuscript.

#### FUNDING

Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) [Project-ID 393547839 – SFB 1361, CA 198/9-2, CA 198/12-1 to M.C.C.]. Funding for open access charge: Deutsche Forschungsgemeinschaft.

**Conflict of interest statement.** None declared.

#### REFERENCES

- Dimitrova, D.S. and Gilbert, D.M. (1999) The spatial position and replication timing of chromosomal domains are both established in early G1 phase. *Mol. Cell*, **4**, 983–993.
- Hyrien, O., Marheineke, K. and Goldar, A. (2003) Paradoxes of eukaryotic DNA replication: MCM proteins and the random completion problem. *BioEssays*, **25**, 116–125.
- Fragkos, M., Ganier, O., Coulombe, P. and Mechali, M. (2015) DNA replication origin activation in space and time. *Nat. Rev. Mol. Cell Biol.*, **16**, 360–374.
- Parker, M.W., Botchan, M.R. and Berger, J.M. (2017) Mechanisms and regulation of DNA replication initiation in eukaryotes. *Crit. Rev. Biochem. Mol. Biol.*, **52**, 107–144.
- Chagin, V.O., Stear, J.H. and Cardoso, M.C. (2010) Organization of DNA replication. *Cold Spring Harb. Perspect. Biol.*, **2**, a000737.
- Jackson, D.A. and Pombo, A. (1998) Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. *J. Cell Biol.*, **140**, 1285–1295.
- Leonhardt, H., Rahn, H.P., Weinzierl, P., Sporberr, A., Cremer, T., Zink, D. and Cardoso, M.C. (2000) Dynamics of DNA replication factories in living cells. *J. Cell Biol.*, **149**, 271–280.
- Ma, H., Samarabandu, J., Devdhar, R.S., Acharya, R., Cheng, P.C., Meng, C. and Berezney, R. (1998) Spatial and temporal dynamics of DNA replication sites in mammalian cells. *J. Cell Biol.*, **143**, 1415–1425.
- Manders, E.M., Stap, J., Brakenhoff, G.J., van Driel, R. and Aten, J.A. (1992) Dynamics of three-dimensional replication patterns during the S-phase, analysed by double labelling of DNA and confocal microscopy. *J. Cell Sci.*, **103** (Pt 3), 857–862.
- O’Keefe, R.T., Henderson, S.C. and Spector, D.L. (1992) Dynamic organization of DNA replication in mammalian cell nuclei: spatially and temporally defined replication of chromosome-specific alpha-satellite DNA sequences. *J. Cell Biol.*, **116**, 1095–1110.
- Marchal, C., Sima, J. and Gilbert, D.M. (2019) Control of DNA replication timing in the 3D genome. *Nat. Rev. Mol. Cell Biol.*, **20**, 721–737.
- Flockingier, R.A., Freedman, M.L. and Stambrook, P.J. (1967) Generation times and DNA replication patterns of cells of developing frog embryos. *Dev. Biol.*, **16**, 457–473.
- Manders, E.M., Stap, J., Strackee, J., van Driel, R. and Aten, J.A. (1996) Dynamic behavior of DNA replication domains. *Exp. Cell Res.*, **226**, 328–335.
- van Dierendonck, J.H., Keyzer, R., van de Velde, C.J. and Cornelisse, C.J. (1989) Subdivision of S-phase by analysis of nuclear 5-bromodeoxyuridine staining patterns. *Cytometry*, **10**, 143–150.
- Nakamura, H., Morita, T. and Sato, C. (1986) Structural organizations of replicon domains during DNA synthetic phase in the mammalian nucleus. *Exp. Cell Res.*, **165**, 291–297.
- Nakayasu, H. and Berezney, R. (1989) Mapping replicational sites in the eucaryotic cell nucleus. *J. Cell Biol.*, **108**, 1–11.
- Sporberr, A., Gahl, A., Ankerhold, R., Leonhardt, H. and Cardoso, M.C. (2002) DNA polymerase clamp shows little turnover at established replication sites but sequential de novo assembly at adjacent origin clusters. *Mol. Cell*, **10**, 1355–1365.
- Hozak, P., Jackson, D.A. and Cook, P.R. (1994) Replication factories and nuclear bodies: the ultrastructural characterization of replication sites during the cell cycle. *J. Cell Sci.*, **107**, 2191–2202.
- Friedman, K.L., Brewer, B.J. and Fangman, W.L. (1997) Replication profile of *Saccharomyces cerevisiae* chromosome VI. *Genes Cells*, **2**, 667–678.
- Heinz, K.S., Casas-Delucchi, C.S., Torok, T., Cmarko, D., Rapp, A., Raska, I. and Cardoso, M.C. (2018) Peripheral re-localization of constitutive heterochromatin advances its replication timing and

- impairs maintenance of silencing marks. *Nucleic Acids Res.*, **46**, 6112–6128.
21. Lob, D., Lengert, N., Chagin, V.O., Reinhart, M., Casas-Delucchi, C.S., Cardoso, M.C. and Drossel, B. (2016) 3D replicon distributions arise from stochastic initiation and domino-like DNA replication progression. *Nat. Commun.*, **7**, 11207.
  22. Yamashita, M., Hori, Y., Shinomiya, T., Obuse, C., Tsurimoto, T., Yoshikawa, H. and Shirahige, K. (1997) The efficiency and timing of initiation of replication of multiple replicons of *Saccharomyces cerevisiae* chromosome VI. *Genes Cells*, **2**, 655–665.
  23. Berezney, R., Dubey, D.D. and Huberman, J.A. (2000) Heterogeneity of eukaryotic replicons, replicon clusters, and replication foci. *Chromosoma*, **108**, 471–484.
  24. Huberman, J.A. and Riggs, A.D. (1968) On the mechanism of DNA replication in mammalian chromosomes. *J. Mol. Biol.*, **32**, 327–341.
  25. Courbet, S., Gay, S., Arnoult, N., Wronka, G., Anglana, M., Brison, O. and Debatisse, M. (2008) Replication fork movement sets chromatin loop size and origin choice in mammalian cells. *Nature*, **455**, 557–560.
  26. Guillou, E., Ibarra, A., Coulon, V., Casado-Vela, J., Rico, D., Casal, I., Schwob, E., Losada, A. and Mendez, J. (2010) Cohesin organizes chromatin loops at DNA replication factories. *Genes Dev.*, **24**, 2812–2822.
  27. Mamberti, S. and Cardoso, M.C. (2020) Are the processes of DNA replication and DNA repair reading a common structural chromatin unit? *Nucleus*, **11**, 66–82.
  28. Chagin, V.O., Casas-Delucchi, C.S., Reinhart, M., Schermelleh, L., Markaki, Y., Maiser, A., Boliu, J.J., Bensimon, A., Fillies, M., Domaing, P. et al. (2016) 4D Visualization of replication foci in mammalian cells corresponding to individual replicons. *Nat. Commun.*, **7**, 11231.
  29. Reinhart, M. and Cardoso, M.C. (2017) A journey through the microscopic ages of DNA replication. *Protoplasma*, **254**, 1151–1162.
  30. Aladjem, M.I. (2007) Replication in context: dynamic regulation of DNA replication patterns in metazoans. *Nat. Rev. Genet.*, **8**, 588–600.
  31. Casas-Delucchi, C.S., Brero, A., Rahn, H.P., Solovei, I., Wutz, A., Cremer, T., Leonhardt, H. and Cardoso, M.C. (2011) Histone acetylation controls the inactive X chromosome replication dynamics. *Nat. Commun.*, **2**, 222.
  32. Casas-Delucchi, C.S., van Bemmel, J.G., Haase, S., Herce, H.D., Nowak, D., Meilinger, D., Stear, J.H., Leonhardt, H. and Cardoso, M.C. (2012) Histone hypoacetylation is required to maintain late replication timing of constitutive heterochromatin. *Nucleic Acids Res.*, **40**, 159–169.
  33. Jorgensen, H.F., Azuara, V., Amols, S., Spivakov, M., Terry, A., Nesterova, T., Cobb, B.S., Ramsahoye, B., Merkenschlager, M. and Fisher, A.G. (2007) The impact of chromatin modifiers on the timing of locus replication in mouse embryonic stem cells. *Genome Biol.*, **8**, R169.
  34. Vogelauer, M., Rubbi, L., Lucas, I., Brewer, B.J. and Grunstein, M. (2002) Histone acetylation regulates the time of replication origin firing. *Mol. Cell*, **10**, 1223–1233.
  35. Blumenthal, A.B., Kriegstein, H.J. and Hogness, D.S. (1974) The units of DNA replication in *Drosophila melanogaster* chromosomes. *Cold Spring Harb. Symp. Quant. Biol.*, **38**, 205–223.
  36. Hyrien, O., Maric, C. and Mechali, M. (1995) Transition in specification of embryonic metazoan DNA replication origins. *Science*, **270**, 994–997.
  37. Hyrien, O. and Mechali, M. (1993) Chromosomal replication initiates and terminates at random sequences but at regular intervals in the ribosomal DNA of *Xenopus* early embryos. *EMBO J.*, **12**, 4511–4520.
  38. Walter, J. and Newport, J.W. (1997) Regulation of replicon size in *Xenopus* egg extracts. *Science*, **275**, 993–995.
  39. Hamatani, T., Carter, M.G., Sharov, A.A. and Ko, M.S. (2004) Dynamics of global gene expression changes during mouse preimplantation development. *Dev. Cell*, **6**, 117–131.
  40. Sansam, C.G., Goins, D., Siefert, J.C., Clowdus, E.A. and Sansam, C.L. (2015) Cyclin-dependent kinase regulates the length of S phase through TICRR/TRESLIN phosphorylation. *Genes Dev.*, **29**, 555–566.
  41. Torres-Padilla, M.E. and Zernicka-Goetz, M. (2006) Role of TIF1alpha as a modulator of embryonic transcription in the mouse zygote. *J. Cell Biol.*, **174**, 329–338.
  42. Ferreira, J. and Carmo-Fonseca, M. (1997) Genome replication in early mouse embryos follows a defined temporal and spatial order. *J. Cell Sci.*, **110**, 889–897.
  43. Hiratani, I., Ryba, T., Itoh, M., Yokochi, T., Schwaiger, M., Chang, C.W., Lyuu, Y., Townes, T.M., Schubeler, D. and Gilbert, D.M. (2008) Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol.*, **6**, e245.
  44. Rivera-Mulia, J.C., Buckley, Q., Sasaki, T., Zimmerman, J., Didier, R.A., Nazor, K., Loring, J.F., Lian, Z., Weissman, S., Robins, A.J. et al. (2015) Dynamic changes in replication timing and gene expression during lineage specification of human pluripotent stem cells. *Genome Res.*, **25**, 1091–1103.
  45. Li, E., Bestor, T.H. and Jaenisch, R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, **69**, 915–926.
  46. Doetschman, T., Gregg, R.G., Maeda, N., Hooper, M.L., Melton, D.W., Thompson, S. and Smithies, O. (1987) Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature*, **330**, 576–578.
  47. Peters, A.H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schofer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A. et al. (2001) Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell*, **107**, 323–337.
  48. Yaffe, D. and Saxel, O. (1977) Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature*, **270**, 725–727.
  49. Bertulat, B., De Bonis, M.L., Della Ragione, F., Lehmkuhl, A., Mildner, M., Storm, C., Jost, K.L., Scala, S., Hendrich, B., D'Esposito, M. et al. (2012) MeCP2 dependent heterochromatin reorganization during neural differentiation of a novel MeCP2-deficient embryonic stem cell reporter line. *PLoS One*, **7**, e47848.
  50. Sporb, A., Domaing, P., Leonhardt, H. and Cardoso, M.C. (2005) PCNA acts as a stationary loading platform for transiently interacting Okazaki fragment maturation proteins. *Nucleic Acids Res.*, **33**, 3521–3528.
  51. Lindhout, B.I., Franz, P., Tessadori, F., Meckel, T., Hooykaas, P.J. and van der Zaal, B.J. (2007) Live cell imaging of repetitive DNA sequences via GFP-tagged polydactyl zinc finger proteins. *Nucleic Acids Res.*, **35**, e107.
  52. Casas-Delucchi, C.S., Becker, A., Boliu, J.J. and Cardoso, M.C. (2012) Targeted manipulation of heterochromatin rescues MeCP2 Rett mutants and re-establishes higher order chromatin organization. *Nucleic Acids Res.*, **40**, e176.
  53. Chagin, V.O., Reinhart, M. and Cardoso, M.C. (2015) High-resolution analysis of mammalian DNA replication units. *Methods Mol. Biol.*, **1300**, 43–65.
  54. Bialic, M., Coulon, V., Drac, M., Gostan, T. and Schwob, E. (2015) Analyzing the dynamics of DNA replication in Mammalian cells using DNA combing. *Methods Mol. Biol.*, **1300**, 67–78.
  55. Weber, P., Rausch, C., Scholl, A. and Cardoso, M.C. (2019) Repli-FISH (fluorescence in situ hybridization): application of 3D-(immuno)-FISH for the study of DNA replication timing of genetic repeat elements. *OBM Genet.*, **3**, e1325.
  56. Gustafsson, M.G., Shao, L., Carlton, P.M., Wang, C.J., Golubovskaya, I.N., Cande, W.Z., Agard, D.A. and Sedat, J.W. (2008) Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. *Biophys. J.*, **94**, 4957–4970.
  57. Berg, S., Kutra, D., Kroeger, T., Strachle, C.N., Kausler, B.X., Haubold, C., Schiegg, M., Ales, J., Beier, T., Rudy, M. et al. (2019) ilastik: interactive machine learning for (bio)image analysis. *Nat. Methods*, **16**, 1226–1232.
  58. van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Goullart, E., Yu, T. and contributors, t.s.-i. (2014) scikit-image: image processing in Python. *PeerJ*, **2**, e453.
  59. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B. et al. (2012) Fiji: an open-source platform for biological-image analysis. *Nat. Methods*, **9**, 676–682.

60. Natale, F., Rapp, A., Yu, W., Maiser, A., Harz, H., Scholl, A., Grulich, S., Anton, T., Horl, D., Chen, W. *et al.* (2017) Identification of the elementary structural units of the DNA damage response. *Nat. Commun.*, **8**, 15760.
61. Prorok, P., Artufel, M., Aze, A., Coulombe, P., Peiffer, I., Lacroix, L., Guedin, A., Mergny, J.L., Damaschke, J., Schepers, A. *et al.* (2019) Involvement of G-quadruplex regions in mammalian replication origin activity. *Nat. Commun.*, **10**, 3274.
62. Almeida, R., Fernandez-Justel, J.M., Santa-Maria, C., Cadoret, J.C., Cano-Aroca, L., Lombrana, R., Herranz, G., Agresti, A. and Gomez, M. (2018) Chromatin conformation regulates the coordination between DNA replication and transcription. *Nat. Commun.*, **9**, 1590.
63. Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
64. Tubbs, A., Sridharan, S., van Wietmarschen, N., Maman, Y., Callen, E., Stanlie, A., Wu, W., Wu, X., Day, A., Wong, N. *et al.* (2018) Dual roles of Poly(dA:dT) tracts in replication initiation and fork collapse. *Cell*, **174**, 1127–1142.
65. Hansen, R.S., Thomas, S., Sandstrom, R., Canfield, T.K., Thurman, R.E., Weaver, M., Dorschner, M.O., Gartler, S.M. and Stamatoyanopoulos, J.A. (2010) Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *PNAS*, **107**, 139–144.
66. Pope, B.D., Hiratani, I. and Gilbert, D.M. (2010) Domain-wide regulation of DNA replication timing during mammalian development. *Chromosome Res.*, **18**, 127–136.
67. Woodfine, K., Beare, D.M., Ichimura, K., Debernardi, S., Mungall, A.J., Fiegler, H., Collins, V.P., Carter, N.P. and Dunham, I. (2005) Replication timing of human chromosome 6. *Cell Cycle*, **4**, 172–176.
68. Woodfine, K., Fiegler, H., Beare, D.M., Collins, J.E., McCann, O.T., Young, B.D., Debernardi, S., Mott, R., Dunham, I. and Carter, N.P. (2004) Replication timing of the human genome. *Hum. Mol. Genet.*, **13**, 191–202.
69. Natale, F., Scholl, A., Rapp, A., Yu, W., Rausch, C. and Cardoso, M.C. (2018) DNA replication and repair kinetics of Alu, LINE-1 and satellite III genomic repetitive elements. *Epigenet. Chromatin*, **11**, 61.
70. Shermoen, A.W., McClelland, M.L. and O'Farrell, P.H. (2010) Developmental control of late replication and S phase length. *Curr. Biol.*, **20**, 2067–2077.
71. Jeppesen, P., Mitchell, A., Turner, B. and Perry, P. (1992) Antibodies to defined histone epitopes reveal variations in chromatin conformation and underacetylation of centric heterochromatin in human metaphase chromosomes. *Chromosoma*, **101**, 322–332.
72. Guenatri, M., Bailly, D., Maison, C. and Almouzni, G. (2004) Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *J. Cell Biol.*, **166**, 493–505.
73. Brero, A., Easwaran, H.P., Nowak, D., Grunewald, I., Cremer, T., Leonhardt, H. and Cardoso, M.C. (2005) Methyl CpG-binding proteins induce large-scale chromatin reorganization during terminal differentiation. *J. Cell Biol.*, **169**, 733–743.
74. Mayer, R., Brero, A., von Hase, J., Schroeder, T., Cremer, T. and Dietzel, S. (2005) Common themes and cell type specific variations of higher order chromatin arrangements in the mouse. *BMC Cell Biol.*, **6**, 44.
75. Solovei, I. and Joffe, B. (2010) Inverted nuclear architecture and its development during differentiation of mouse rod photoreceptor cells: a new model to study nuclear architecture. *Genetika*, **46**, 1159–1163.
76. Solovei, I., Kreysing, M., Lanctot, C., Kosem, S., Peichl, L., Cremer, T., Guck, J. and Joffe, B. (2009) Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. *Cell*, **137**, 356–368.
77. Nozaki, T., Imai, R., Tanbo, M., Nagashima, R., Tamura, S., Tani, T., Joti, Y., Tomita, M., Hibino, K., Kanemaki, M.T. *et al.* (2017) Dynamic organization of chromatin domains revealed by super-resolution live-cell imaging. *Mol. Cell*, **67**, 282–293.
78. Lopes Novo, C. and Rugg-Gunn, P.J. (2016) Chromatin organization in pluripotent cells: emerging approaches to study and disrupt function. *Brief. Funct. Genomics*, **15**, 305–314.
79. Meshorer, E. and Misteli, T. (2006) Chromatin in pluripotent embryonic stem cells and differentiation. *Nat. Rev. Mol. Cell Biol.*, **7**, 540–546.
80. Gaspar-Maia, A., Alajem, A., Polesso, F., Sridharan, R., Mason, M.J., Heidersbach, A., Ramalho-Santos, J., McManus, M.T., Plath, K., Meshorer, E. *et al.* (2009) Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature*, **460**, 863–868.
81. Kobayakawa, S., Miike, K., Nakao, M. and Abe, K. (2007) Dynamic changes in the epigenomic state and nuclear organization of differentiating mouse embryonic stem cells. *Genes Cells*, **12**, 447–460.
82. Aparicio, J.G., Viggiani, C.J., Gibson, D.G. and Aparicio, O.M. (2004) The Rpd3-Sin3 histone deacetylase regulates replication timing and enables intra-S origin control in *Saccharomyces cerevisiae*. *Mol. Cell Biol.*, **24**, 4769–4780.
83. Bickmore, W.A. and Carothers, A.D. (1995) Factors affecting the timing and imprinting of replication on a mammalian chromosome. *J. Cell Sci.*, **108** (Pt 8), 2801–2809.
84. Goren, A., Tabib, A., Hecht, M. and Cedar, H. (2008) DNA replication timing of the human beta-globin domain is controlled by histone modification at the origin. *Genes Dev.*, **22**, 1319–1324.
85. Kemp, M.G., Ghosh, M., Liu, G. and Leffak, M. (2005) The histone deacetylase inhibitor trichostatin A alters the pattern of DNA replication origin activity in human cells. *Nucleic Acids Res.*, **33**, 325–336.
86. Keohane, A.M., O'Neill, L.P., Belyaev, N.D., Lavender, J.S. and Turner, B.M. (1996) X-Inactivation and histone H4 acetylation in embryonic stem cells. *Dev. Biol.*, **180**, 618–630.
87. Lin, C.M., Fu, H., Martinovsky, M., Bouhassira, E. and Aladjem, M.I. (2003) Dynamic alterations of replication timing in mammalian cells. *Curr. Biol.*, **13**, 1019–1028.
88. Perry, P., Sauer, S., Billon, N., Richardson, W.D., Spivakov, M., Warnes, G., Livesey, F.J., Merckenschlager, M., Fisher, A.G. and Azuara, V. (2004) A dynamic switch in the replication timing of key regulator genes in embryonic stem cells upon neural induction. *Cell Cycle*, **3**, 1645–1650.
89. Schubeler, D., Francastel, C., Cimbara, D.M., Reik, A., Martin, D.I. and Groudine, M. (2000) Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human beta-globin locus. *Genes Dev.*, **14**, 940–950.
90. Heinz, K.S. and Cardoso, M.C. (2019) Targeted manipulation/repositioning of subcellular structures and molecules. *Methods Mol. Biol.*, **2038**, 199–208.
91. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
92. Mouse Genome Sequencing, C., Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M. *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520–562.
93. Bouwman, B.A. and de Laat, W. (2015) Getting the genome in shape: the formation of loops, domains and compartments. *Genome Biol.*, **16**, 154.
94. Krijger, P.H. and de Laat, W. (2013) Identical cells with different 3D genomes; cause and consequences? *Curr. Opin. Genet. Dev.*, **23**, 191–196.
95. Politz, J.C., Scalzo, D. and Groudine, M. (2013) Something silent this way forms: the functional organization of the repressive nuclear compartment. *Annu. Rev. Cell Dev. Biol.*, **29**, 241–270.
96. Solovei, I., Thanisch, K. and Feodorova, Y. (2016) How to rule the nucleus: divide et impera. *Curr. Opin. Cell Biol.*, **40**, 47–59.
97. Ahmad, K. and Henikoff, S. (2001) Centromeres are specialized replication domains in heterochromatin. *J. Cell Biol.*, **153**, 101–110.
98. McCarroll, R.M. and Fangman, W.L. (1988) Time of replication of yeast centromeres and telomeres. *Cell*, **54**, 505–513.
99. Chagin, V.O., Reinhart, B., Becker, A., Mortusewicz, O., Jost, K.L., Rapp, A., Leonhardt, H. and Cardoso, M.C. (2019) Processive DNA synthesis is associated with localized decompaction of constitutive heterochromatin at the sites of DNA replication and repair. *Nucleus*, **10**, 231–253.
100. Weidtkamp-Peters, S., Rahn, H.P., Cardoso, M.C. and Hemmerich, P. (2006) Replication of centromeric heterochromatin in mouse

- fibroblasts takes place in early, middle, and late S phase. *Histochem. Cell Biol.*, **125**, 91–102.
101. Arnoult, N., Schluth-Bolard, C., Letessier, A., Drascovic, I., Bouarich-Bourimi, R., Campisi, J., Kim, S.H., Bousouar, A., Ottaviani, A., Magdinier, F. *et al.* (2010) Replication timing of human telomeres is chromosome arm-specific, influenced by subtelomeric structures and connected to nuclear localization. *PLoS Genet.*, **6**, e1000920.
  102. Jobling, M.A. and Tyler-Smith, C. (2003) The human Y chromosome: an evolutionary marker comes of age. *Nat. Rev. Genet.*, **4**, 598–612.
  103. Singh, N.P., Madabhushi, S.R., Srivastava, S., Senthilkumar, R., Neeraja, C., Khosla, S. and Mishra, R.K. (2011) Epigenetic profile of the euchromatic region of human Y chromosome. *Nucleic Acids Res.*, **39**, 3594–3606.
  104. Aken, B.L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., Fernandez Banet, J., Billis, K., Garcia Giron, C., Hourlier, T. *et al.* (2016) The Ensembl gene annotation system. *Database*, **2016**, baw093.
  105. Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T. *et al.* (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, **423**, 825–837.
  106. Maan, A.A., Eales, J., Akbarov, A., Rowland, J., Xu, X., Jobling, M.A., Charchar, F.J. and Tomaszewski, M. (2017) The Y chromosome: a blueprint for men's health? *Eur. J. Hum. Genet.*, **25**, 1181–1188.
  107. Charlesworth, B. (2003) The organization and evolution of the human Y chromosome. *Genome Biol.*, **4**, 226.
  108. Quintana-Murci, L. and Fellous, M. (2001) The human Y chromosome: the biological role of a “functional wasteland”. *J. Biomed. Biotechnol.*, **1**, 18–24.
  109. Schubeler, D., Scalzo, D., Kooperberg, C., van Steensel, B., Delrow, J. and Groudine, M. (2002) Genome-wide DNA replication profile for *Drosophila melanogaster*: a link between transcription and replication timing. *Nat. Genet.*, **32**, 438–442.
  110. Cadoret, J.C., Meisch, F., Hassan-Zadeh, V., Luyten, I., Guillet, C., Duret, L., Quesneville, H. and Prioleau, M.N. (2008) Genome-wide studies highlight indirect links between human replication origins and gene regulation. *PNAS*, **105**, 15837–15842.
  111. Costantini, M. and Bernardi, G. (2008) Replication timing, chromosomal bands, and isochores. *PNAS*, **105**, 3433–3437.
  112. Picard, F., Cadoret, J.C., Audit, B., Arneodo, A., Alberti, A., Battail, C., Duret, L. and Prioleau, M.N. (2014) The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells. *PLoS Genet.*, **10**, e1004282.
  113. Blow, J.J., Gillespie, P.J., Francis, D. and Jackson, D.A. (2001) Replication origins in *Xenopus* egg extract are 5–15 kilobases apart and are activated in clusters that fire at different times. *J. Cell Biol.*, **152**, 15–25.
  114. Mills, A.D., Blow, J.J., White, J.G., Amos, W.B., Wilcock, D. and Laskey, R.A. (1989) Replication occurs at discrete foci spaced throughout nuclei replicating in vitro. *J. Cell Sci.*, **94**, 471–477.
  115. Newport, J. and Kirschner, M. (1982) A major developmental transition in early *Xenopus* embryos: II. Control of the onset of transcription. *Cell*, **30**, 687–696.
  116. Newport, J. and Kirschner, M. (1982) A major developmental transition in early *Xenopus* embryos: I. Characterization and timing of cellular changes at the midblastula stage. *Cell*, **30**, 675–686.
  117. Li, V.C., Ballabeni, A. and Kirschner, M.W. (2012) Gap 1 phase length and mouse embryonic stem cell self-renewal. *PNAS*, **109**, 12550–12555.
  118. Burdon, T., Smith, A. and Savatier, P. (2002) Signalling, cell cycle and pluripotency in embryonic stem cells. *Trends Cell Biol.*, **12**, 432–438.
  119. Soufi, A. and Dalton, S. (2016) Cycling through developmental decisions: how cell cycle dynamics control pluripotency, differentiation and reprogramming. *Development*, **143**, 4301–4311.
  120. White, J. and Dalton, S. (2005) Cell cycle control of embryonic stem cells. *Stem Cell Rev.*, **1**, 131–138.
  121. Tamm, C., Pijuan Galito, S. and Anneren, C. (2013) A comparative study of protocols for mouse embryonic stem cell culturing. *PLoS One*, **8**, e81156.
  122. Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W. *et al.* (2015) Chromatin architecture reorganization during stem cell differentiation. *Nature*, **518**, 331–336.
  123. Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T. and Misteli, T. (2006) Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell*, **10**, 105–116.
  124. Pekowska, A., Klaus, B., Xiang, W., Severino, J., Daigle, N., Klein, F.A., Oles, M., Casellas, R., Ellenberg, J., Steinmetz, L.M. *et al.* (2018) Gain of CTCF-Anchored chromatin loops marks the exit from naive pluripotency. *Cell Syst.*, **7**, 482–495.
  125. Fraser, P. (2006) Transcriptional control thrown for a loop. *Curr. Opin. Genet. Dev.*, **16**, 490–495.
  126. Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F. and de Laat, W. (2002) Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell*, **10**, 1453–1465.
  127. Buongiorno-Nardelli, M., Micheli, G., Carri, M.T. and Marilley, M. (1982) A relationship between replicon size and supercoiled loop domains in the eukaryotic genome. *Nature*, **298**, 100–102.
  128. Techer, H., Koundrioukoff, S., Azar, D., Wilhelm, T., Carignon, S., Brison, O., Debatisse, M. and Le Tallec, B. (2013) Replication dynamics: biases and robustness of DNA fiber analysis. *J. Mol. Biol.*, **425**, 4845–4855.
  129. Cardoso, M.C. and Leonhardt, H. (1999) DNA methyltransferase is actively retained in the cytoplasm during early development. *J. Cell Biol.*, **147**, 25–32.
  130. Cayrou, C., Gregoire, D., Coulombe, P., Danis, E. and Mechali, M. (2012) Genome-scale identification of active DNA replication origins. *Methods*, **57**, 158–164.
  131. Petryk, N., Kahli, M., d'Aubenton-Carafa, Y., Jaszczyszyn, Y., Shen, Y., Silvain, M., Thernes, C., Chen, C.L. and Hyrien, O. (2016) Replication landscape of the human genome. *Nat. Commun.*, **7**, 10208.
  132. Efroni, S., Dutttagupta, R., Cheng, J., Dehghani, H., Hoepfner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D., Buetow, K.H. *et al.* (2008) Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell*, **2**, 437–447.
  133. Martin, G.R. (1981) Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *PNAS*, **78**, 7634–7638.
  134. Capparelli, R., Cottone, C., D'Apice, L., Viscardi, M., Colantonio, L., Lucretti, S. and Iannelli, D. (1997) DNA content differences in laboratory mouse strains determined by flow cytometry. *Cytometry*, **29**, 261–266.
  135. Vinogradov, A.E. (1998) Genome size and GC-percent in vertebrates as determined by flow cytometry: the triangular relationship. *Cytometry*, **31**, 100–109.
  136. Phillips-Cremmins, J.E., Sauria, M.E., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y. *et al.* (2013) Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*, **153**, 1281–1295.

## SUPPLEMENTARY TABLES, FIGURES AND MOVIES

### Developmental differences in genome replication program and origin activation

Cathia Rausch<sup>1,†</sup>, Patrick Weber<sup>1,†</sup>, Paulina Prorok<sup>1</sup>, David Hörl<sup>2</sup>, Andreas Maiser<sup>2</sup>, Anne Lehmkuhl<sup>1</sup>, Vadim O. Chagin<sup>1,3</sup>, Corella S. Casas-Delucchi<sup>1</sup>, Heinrich Leonhardt<sup>2</sup> and M. Cristina Cardoso<sup>1,\*</sup>

1 Department of Biology, Technical University of Darmstadt, 64287 Darmstadt, Germany

2 Department of Biology II, LMU Munich, 81377 Munich, Germany

3 Institute of Cytology, Russian Academy of Sciences, St. Petersburg, Russia

† These authors contributed equally to this work.

\* To whom correspondence should be addressed. Tel: +49 6151 16 21882; Fax: +49 6151 16 21880; Email: cardoso@bio.tu-darmstadt.de.

Present address: Corella S. Casas-Delucchi, The Institute of Cancer Research | Chester Beatty Laboratories, London, SW3 6JB, United Kingdom.

## SUPPLEMENTARY TABLES

Supplementary Table 1: Cell line characteristics.

Name	Species	Type	Genotype	Reference
J1 wt	Mus musculus	embryonic stem cell	wildtype	(1)
E14 wt	Mus musculus	embryonic stem cell	wildtype	(2)
C2C12	Mus musculus	myoblast	wildtype	(3)
MEF W8	Mus musculus	embryonic fibroblast	wildtype	(4)
Primary fibroblast	Mus musculus	primary ear fibroblast*	wildtype	-

\* from adult C57BL/6 mice.

Supplementary Table 2: Plasmid characteristics.

Name	pc number*	Fluorophore	Gene species	Promoter	Reference
pmRFP-PCNA	1054	mRFP	Homo sapiens	CMV	(5)
pMaSat-GFP	1803	GFP	-	CMV	(6)
pGBP-MaSat	2469	-	-	CMV	(7)
peGFP-HDAC1	2447	eGFP	Homo sapiens	CMV	This study
peGFP-C1	592	eGFP	Aequorea victoria	CMV	Clontech Laboratories, Mountain View, CA, USA

\* pc: plasmid collection.

**Supplementary Table 3: Nucleotide characteristics.**

Name	Application	Detection	Cat #	Company
Biotin-16-dUTP	Labeling of FISH* probes	Streptavidin	-	Self made (8)
Dig-11-dUTP	Labeling of FISH probes	Antibody detection	11573152910	Sigma-Aldrich, St Louis, MO, USA
Cy3-dUTP	Labeling of FISH probes	-	PA53022	GE Healthcare, Chicago, IL, USA
Cy5-dUTP	Labeling of FISH probes	-	PA55022	GE Healthcare, Chicago, IL, USA
dATP, dTTP, dCTP & dGTP	Generation of FISH probes	-	10297018	Thermo Fisher Scientific, Waltham, MA, USA
5-ethynyl-2'-deoxyuridine (EdU)	Labeling of nascent DNA in pulse (chase) experiments	ClickIT chemistry	E10415	Thermo Fisher Scientific, Waltham, MA, USA
5-bromo-2'-deoxyuridine (BrdU)	Labeling of nascent DNA in pulse (chase) experiments	Antibody detection	B5002	Sigma-Aldrich, St Louis, MO, USA
Thymidine	Labeling of nascent DNA in pulse (chase) experiments**	-	T9250	Sigma-Aldrich, St Louis, MO, USA
5-iodo-2'-deoxyuridine (IdU)	Labeling of nascent DNA in molecular DNA combing experiments	Antibody detection	I7125	Sigma-Aldrich, St Louis, MO, USA
5-chloro-2'-deoxyuridine (CldU)	Labeling of nascent DNA in molecular DNA combing experiments	Antibody detection	C6891	Sigma-Aldrich, St Louis, MO, USA

\* fluorescence *in situ* hybridization, \*\* added only during chase period in pulse chase experiments.

**Supplementary Table 4: Primary and secondary antibody characteristics.**

Reactivity	Host	Dilution	Application	Cat #	Company
anti-PCNA (PC-10)	Mouse	1:100	IF*	M0879	Dako, Hamburg, Germany
anti-Oct3/4 (40/Oct-3)	Mouse	1:1000	IF	611203	BD Bioscience, San Jose, CA, USA
anti-H3K9ac (1B10)	Mouse	1:500	IF	61251	Active Motif, La Hulpe, Belgium
anti-BrdU/IdU (B44)	Mouse	1:200	Molecular combing IF	347580	Becton Dickinson, Franklin Lakes, NJ, USA
anti-ssDNA (16-19, IgG2a)	Mouse	1:200	Molecular combing IF	MAB3034	Millipore, Burlington, MA, USA
anti-Sox2 (EPR3131)	Rabbit	1:200	IF	ab92494	Abcam, Cambridge, UK
anti-H4K8ac	Rabbit	1:100	IF	06-760	Upstate, Lake Placid, NY, USA
anti-H3K9m3	Rabbit	1:200	IF	39161	Active Motif, La Hulpe, Belgium
anti-H4K5ac (EP1000Y)	Rabbit	1:500	IF	1808	Epitomics, Burlingame, CA, USA
anti-digoxigenin	Rabbit	1:500	FISH** probe detection	700772	Thermo Fisher Scientific, Waltham, MA, USA
anti-BrdU/CldU (BU1/75 (ICR1))	Rat	1:100/1:200	IF/Molecular combing IF	OBT0030CX	BioRad, Puchheim, Germany
anti-GFP (7.1 & 13.1)	Mouse	1:500	FISH IF	11814460001	Roche, Penzberg, Germany
anti-mouse Alexa488	Donkey	1:300	IF (fluorescent secondary)	715-545-150	The Jackson Laboratory, Bar Harbor, ME, USA
anti-mouse IgG Cy3	Donkey	1:300	IF (fluorescent secondary)	715-165-151	The Jackson Laboratory, Bar Harbor, ME, USA
anti-mouse IgG Chromeo 546	Goat	1:200	Molecular combing IF (fluorescent secondary)	15033	Active Motif, La Hulpe, Belgium
anti-mouse IgG AlexaFluor 594	Donkey	1:500	IF (fluorescent secondary)	715-585-151	The Jackson Laboratory, Bar Harbor, ME, USA

Reactivity	Host	Dilution	Application	Cat #	Company
anti-mouse IgG Alexa647	Goat	1:300	IF (fluorescent secondary)	A-32728	Molecular Probes Inc., Eugene, OR, USA
anti-mouse IgG2a AlexaFluor 647	Goat	1:200	Molecular combing IF (fluorescent secondary)	A-21241	Fisher Scientific GmbH, Hampton, NH, USA
anti-rabbit IgG Cy3	Donkey	1:300	IF (fluorescent secondary)/FISH probe detection	711-165-152	The Jackson Laboratory, Bar Harbor, ME, USA
anti-rabbit IgG Alexa647	Goat	1:300	IF (fluorescent secondary)	A-21245	Molecular Probes Inc., Eugene, OR, USA
anti-rat IgG AlexaFluor 488	Donkey	1:500/1:200	IF/Molecular combing IF (fluorescent secondary)	712-545-153	The Jackson Laboratory, Bar Harbor, ME, USA

\* IF: immunofluorescence, \*\* fluorescence *in situ* hybridization.

Supplementary Table 5: Imaging systems characteristics.

Microscope/ Company	Lasers/lamps	Filters (ex. & em. [nm])*	Objectives/ lenses	Detection system	Incubation system	Application
Ultra-View VoX spinning disk on an inverted Nikon Ti-E microscope/ PerkinElmer Life Sciences, UK	solid state diode lasers (405 nm, 488 nm, 561 nm, 640 nm)	405/488/568/640** 405: 415–475 488: 505–549 561: 580–650 640: 664–754	oil immersion 60x Plan- Achromat (NA 1.45)	cooled 14-bit Hamamatsu® C9100-50 EMCCD	closed live-cell microscopy chamber (ACU control, Olympus) for time-lapse microscopy	time-lapse microscopy & confocal z- stack imaging
Widefield Axiovert 200 /Zeiss, Germany	HBO100 mercury lamp	488: 473-491 & 506-534 561: 550-580 & 590-650 640: 590-650 & 663-738	oil immersion 63x Plan- Achromat (1.4 NA)	12-bit AxioCam mRM	-	molecular combing imaging
Leica SP5 II/Wetzlar, Germany	405 nm diode 488 nm Argon, 561 nm DPSS, 633 nm HeNe gas	AOBS beam splitter	HCX PL APO 63x / 1.4-0.6 oil lambda blue & HCX PL APO 100x / 1.44 oil Corr CS	2 HyD Hybrid Detectors	-	confocal z- stack imaging
DeltaVision OMX V3/ GE, Chicago, IL, USA	405, 488 and 593 nm diode lasers	405: 401-447, 488: 500-550, 594: 603-627	100x 1.4 oil immersion objective UPlanSApo (Olympus)	Cascade II:512 EMCCD cameras (Photometrics)	-	3D SIM imaging

\* ex.: excitation & em.: emission, \*\* dichroic specification, \*\*\* WD: working distance.

**Supplementary Table 6: Genome-wide replication origin profiling information.**

Dataset	Sample	Cells	Sequencing method	Webpage	Reference
GSE126477	-	mES	Small nascent DNA strand sequencing (SNS-seq)	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE126477">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE126477</a>	(9)
	GSM3602315			<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM3602315">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM3602315</a>	
	GSM3602316			<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM3602316">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM3602316</a>	
	GSM3602317			<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM3602317">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM3602317</a>	
GSE99740	-	mES	Small nascent DNA strand sequencing (SNS-seq)	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99740">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99740</a>	(10)
	GSM2651111			<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2651111">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2651111</a>	
	GSM2651112			<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2651112">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2651112</a>	
GSE99740	-	MEF	Small nascent DNA strand sequencing (SNS-seq)	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99740">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99740</a>	(10)
	GSM2651107			<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2651107">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2651107</a>	
	GSM2651108			<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2651108">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2651108</a>	
GSE116321	-	activated mouse B cells	Okazaki fragment sequencing (OK-seq)	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE116321">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE116321</a>	(11)
	GSM3227970			<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM3227970">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM3227970</a>	
	GSM3227971			<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM3227971">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM3227971</a>	
	GSM3227972			<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM3227972">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM3227972</a>	

Supplementary Table 7: Statistic parameters for Figure 1A, S8D and S12.

Figure	Stage	N	Mean	Median	SD	SEM	CI	P-value
Live cell (1A & S8D)	I	15	3.6	-	0.6	-	-	-
	II	23	4.6	-	0.9	-	-	-
	III	13	1.5	-	0.4	-	-	-
	IV	16	1.1	-	0.2	-	-	-
	IV <sub>2</sub>	18	1.0	-	0.1	-	-	-
RFi distance from nuclear border (S12)	I	32	0.55	0.55	0.13	0.02	0.001	-
	II	32	0.35	0.35	0.08	0.01	0.001	8.2e-10
	III	29	0.3	0.3	0.1	0.02	0.001	4e-12
	IV	24	0.28	0.28	0.07	0.02	0.001	3.4e-13
	IV <sub>2</sub>	27	0.33	0.28	0.1	0.02	0.001	1.8e-09
RFi solidity (S12)	I	32	0.65	0.66	0.01	0.002	0.0001	-/5.2e-13
	II	32	0.66	0.66	0.01	0.003	0.0002	0.49/7.3e-13
	III	29	0.67	0.67	0.02	0.004	0.0003	0.002/4.6e-10
	IV	24	0.69	0.69	0.02	0.005	0.0003	8.5e-08/1.9e-05
	IV <sub>2</sub>	27	0.73	0.73	0.03	0.006	0.0004	5.2e-13/-
RFi numbers (S12)	I	32	232	213	92.12	16.28	1.03	0.091/8.0e-10
	II	32	278	273	123.2	21.78	1.38	-/3.7e-10
	III	29	212	202	100.85	18.73	1.18	0.024/5.0e-07
	IV	24	188	194	87.4	17.84	1.13	0.002/1.4e-05
	IV <sub>2</sub>	27	83	77	59.57	11.46	0.73	3.7e-10/-

Figure	Stage	N	Mean	Median	SD	SEM	CI	P-value
norm. DAPI intensity in RFi (S12)	I	32	0.34	0.33	0.05	0.01	0.0005	0.013
	II	32	0.37	0.37	0.05	0.01	0.0005	-
	III	29	0.34	0.34	0.06	0.01	0.0007	0.012
	IV	24	0.32	0.32	0.06	0.01	0.0007	0.002
	IV <sub>2</sub>	27	0.36	0.35	0.07	0.01	0.0008	0.47

Supplementary Table 8: Statistic parameters for Figure 2 and 3.

Figure	Days/stage/ celltype	N	Mean	Median	SD	SEM	CI	P-value
Oct3/4 (2C)	0	107	1.41	1	1.19	0.115	0.0072	-
	3	117	0.2	0.2	0.07	0.007	0.0004	< 2.16e-16
	7	102	0.14	0.14	0.05	0.005	0.0003	< 2.16e-16
Sox2 (2C)	0	37	1.01	1	0.11	0.02	0.0011	-
	3	33	0.54	0.53	0.02	0.004	0.0002	< 2.16e-16
	7	38	0.53	0.53	0.03	0.005	0.00003	< 2.16e-16
EdU (2E)	early	12	480301	342456	45857	3309	2078	-
	mid	16	685792	422070	68445	3392	2129	0.002417
	late	13	847384	551027	106878	11329	7124	1.74e-05
H3K9ac (2H)	undiff. mESC	24	0.74	0.74	0.18	0.028	0.055	-
	diff. mESC	28	0.61	0.64	0.09	0.017	0.034	0.00054
	primary	25	0.51	0.48	0.11	0.021	0.042	2.23e-07
H3K9m3 (2H)	undiff. mESC	20	1.93	1.99	0.36	0.081	0.159	-
	diff. mESC	25	2.33	2.12	0.74	0.149	0.292	0.1221
	primary	19	2.77	2.52	0.64	0.147	0.288	1.627e-05
H4K8ac (2H)	undiff. mESC	26	0.86	0.78	0.16	0.032	0.062	-
	diff. mESC	26	0.63	0.64	0.11	0.021	0.043	0.000995
	primary	17	0.5	0.48	0.08	0.018	0.036	5.864e-06
H4K5ac (3B)	control	28	1	-	0.25	-	-	-
	HDAC1	34	0.4	-	0.16	-	-	< 2.16e-16
H4K8ac (3B)	control	21	1	-	0.25	-	-	-
	HDAC1	28	0.2	-	0.04	-	-	< 2.16e-16

Figure	Days/stage/ celltype	N	Mean	Median	SD	SEM	CI	P-value
II to III (3D)	control	28	100	-	1.14	-	-	-
	HDAC1	16	51.6	-	9.3	-	-	-
'III' to 'II' (3D)	control	0	0	-	0	-	-	-
	HDAC1	15	48.4	-	7.4	-	-	-
Chromocenter volume (3E)	control	13	14.13	6.24	17.82	2.09	0.13	-
	HDAC1 II to III	10	10.31	6.86	15.07	1.53	0.1	0.14
	HDAC1 'III' to 'II'	9	5.88	4.36	6.43	0.6	0.04	0.0003
Chromocenter shape (3F)	control	13	0.54	0.56	0.12	0.01	0.001	-
	HDAC1 II to III	10	0.6	0.61	0.07	0.01	0.0004	9.4e-10
	HDAC1 'III' to 'II'	9	0.66	0.67	0.07	0.01	0.0004	6.3e-12

**Supplementary Table 9: Statistic parameters for Figure S14.**

Figure	Celltype	N	Mean	Median	SD	SEM	CI	P-value
<b>Number of chromocenters (S14B)</b>	undiff. mESC	28	11.15	10.00	5.54	1.24	0.080	-
	diff. mESC	28	22.00	20.50	7.70	1.72	0.110	1.17e-05
	primary	31	23.35	22.00	9.76	2.18	0.140	3.42e-05
<b>Chromocenter volume (S14C)</b>	undiff. mESC	28	6.88	4.80	9.38	0.63	0.04	-
	diff. mESC	28	3.50	2.05	4.13	0.20	0.01	5.25e-07
	primary	31	3.60	2.07	4.68	0.22	0.01	1.31e-06
<b>Chromocenter compaction (S14D)</b>	undiff. mESC	28	0.04	0.037	0.02	0.0012	0.00007	-
	diff. mESC	28	0.05	0.05	0.04	0.0015	0.00009	< 2.16e-16
	primary	31	0.06	0.051	0.03	0.0013	0.00008	< 2.16e-16
<b>Chromocenter shape factor (S14E)</b>	undiff. mESC	28	0.48	0.47	0.10	0.006	0.004	-
	diff. mESC	28	0.56	0.56	0.08	0.004	0.002	< 2.16e-16
	primary	31	0.58	0.58	0.07	0.003	0.002	< 2.16e-16

**Supplementary Table 10: Statistic parameters for Figure 4, S16, S18 and S19.**

Figure	Stage	N	Mean	Median	SD	SEM	CI	P-value
<b>MaSat/MiSat/ telomeres (4B)</b>	I	12	27.26/ 29.22/ 34.39	25.97/ 27.5/ 33.94	4.4/ 5.15/ 4.47	1.27/ 1.49/ 1.29	0.08/ 0.1/ .08	-
	II	11	43.13/ 38.03/ 38.44	41.37/ 39.05/ 40.39	5.16/ 5.14/ 5.37	1.56/ 1.55/ 1.62	0.1/ 0.1/ 0.1	1.5e-07/ 0.0005/ 0.06
	III	12	21.59/ 26.36/ 27.21	20.79/ 26.46/ 28.3	6.14/ 5.47/ 5.81	1.77/ 1.58/ 1.68	0.11/ 0.1/ 0.11	0.02/ 0.2/ 0.003
	IV	10	14.97/ 17.25/ 21.05	15.01/ 17.6/ 20.69	4.11/ 3.63/ 4.99	1.3/ 1.15/ 1.58	0.08/ 0.07/ 0.1	1.5e-06/ 3.6e-06/ 3.4e-06
	Y	10	9.3/ 10.87/ 11.26	8.41/ 9.49/ 9.75	3.04/ 4.35/ 3.8	0.96/ 1.38/ 1.2	0.06/ 0.09/ 0.08	5.7e-10/ 1.6e-08/ 2.8e-11
<b>Y chromosome (4C &amp; S16D)</b>	I	13	273.67	208.8	265.82	23.97	1.51	4.8e-12
	II	13	278.99	185.16	265.11	36.08	2.27	3.6e-12
	III	8	302.93	158.63	357.69	103.26	6.62	2.2e-11
	IV	8	323.07	53.94	537.48	155.16	9.95	6.6e-9
	Y	10	2049.87	2077.22	849.74	157.79	9.98	-
<b>Telomeres in vs out (4D)</b>	in	14	4014	2951	3725	235	14.73	-
	out	14	2172	1429	2231	111	6.97	6.77e-12
<b>MaSat (S16A)</b>	I	19	234.39	175.48	225.9	14.9	0.94	-
	II	19	896.88	857.6	492.27	36.79	2.3	< 2.16e-16
	III	5	229.56	173.97	216.2	27.24	1.72	0.8766
	IV	6	82.89	43.93	106.67	11.37	0.72	1.51e-14
	Y	6	79.33	63.19	70.1	12.59	0.8	6.07e-13

Figure	Stage	N	Mean	Median	SD	SEM	CI	P-value
<b>MiSat (S16B)</b>	I	13	15.87	13.67	12.48	0.83	0.052	-
	II	16	28.13	24.22	20.79	1.51	0.072	< 2.16e-16
	III	13	7.1	4.31	7.69	0.45	0.028	< 2.16e-16
	IV	10	3.41	2.22	4.04	0.27	0.017	< 2.16e-16
	Y	5	3.21	1.26	5.28	0.54	0.034	< 2.16e-16
<b>Telomeres (S16C)</b>	I	14	28.4	25.63	17.33	0.72	0.045	-
	II	11	33.8	25.64	28.02	1.27	0.080	0.00023
	III	10	29.53	25.74	20.65	1.11	0.070	0.392
	IV	11	18.07	12.94	16.99	1.03	0.065	1.32e-15
	Y	10	11.07	6.7	14.15	1.03	0.064	< 2.16e-16
<b>Y/MaSat/ telomeres (S18A)</b>	I	8	21.4/ 24.19/ 27.4	22.05/ 23.85/ 27.95	2.43/ 2.33/ 2.96	0.86/ 0.82/ 1.05	0.06/ 0.05/ 0.07	-
	II	8	25.31/ 44.54/ 32.56	24.59/ 44.92/ 31.97	5.17/ 2.43/ 3.44	1.83/ 0.86/ 1.22	0.12/ 0.06/ 0.08	0.08/ 9.1e-11/ 0.006
	III	8	23.01/ 21.22/ 21.79	22.58/ 22.41/ 23.41	5.23/ 5.43/ 5.58	1.85/ 1.92/ 1.97	0.12/ 0.12/ 0.11	0.45/ 0.19/ 0.03
	IV	8	22.9/ 13.23/ 16.93	23.36/ 13.29/ 16.25	3.57/ 2.82/ 4.85	1.26/ 1.0/ 1.71	0.08/ 0.06/ 0.11	0.34/ 9e-07/ 0.0002
	Y	9	50.67/ 7.22/ 9.13	50.05/ 6.88/ 9.53	5.92/ 1.32/ 2.42	1.97/ 0.44/ 0.81	0.13/ 0.03/ 0.05	3.6e-08/ 1.9e-09/ 2.2e-09

Figure	Stage	N	Mean	Median	SD	SEM	CI	P-value
<b>Y/MaSat/ MiSat (S18B)</b>	I	9	18.93/ 26.99/ 25.97	16.46/ 24.95/ 25.84	10.43/ 3.88/ 4.73	3.48/ 1.29/ 1.58	0.23/ 0.08/ 0.1	-
	II	11	20.3/ 48.37/ 41.08	18.61/ 48.39/ 42.05	8.87/ 7.18/ 6.74	2.67/ 2.16/ 2.03	0.17/ 0.14/ 0.13	0.76/ 2.7e-07/ 1.6e-05
	III	9	20.03/ 20.49/ 21.45	17.17/ 19.2/ 22.14	8.51/ 5.32/ 3.75	2.84/ 1.77/ 1.25	0.18/ 0.09/ 0.12	0.8/ 0.01/ 0.04
	IV	10	31.32/ 19.17 21.46	33.16/ 19.29/ 20.96	5.595/ 4.52/ 6.06	1.88/ 1.43/ 1.92	0.12/ 0.09/ 0.12	0.008/ 0.0008/ 0.09
	Y	9	51.04/ 10.27/ 10.68	53.52/ 11.07/ 10.87	9.45/ 4.43/ 6.23	3.15/ 1.48/ 2.08	0.2/ 0.1/ 0.13	4.2e-06/ 2.7e-07/ 3.2e-05
<b>Y/MiSat/ telomeres (S18C)</b>	I	9	21.26/ 21.13/ 26.72	20.22/ 20.73/ 25.67	2.97/ 3.54/ 3.57	0.99/ 1.18/ 1.19	0.06/ 0.08/ 0.08	-
	II	8	23.14/ 38.49/ 34.38	23.45/ 38.96/ 34.41	4.49/ 3.28/ 1.76	1.59/ 1.16/ 0.62	0.1/ 0.08/ 0.04	0.33/ 2.7e-08/ 0.0001
	III	7	22.87/ 26.79/ 26.45	22.41/ 26.39/ 27.51	2.96/ 2.72/ 2.8	1.12/ 1.02/ 1.06	0.07/ 0.07/ 0.07	0.3/ 0.003/ 0.9
	IV	9	24.01/ 20.32/ 19.27	23.92/ 20.78/ 19.34	6.17/ 2.12/ 1.8	2.06/ 0.71/ 0.6	0.13/ 0.05/ 0.04	0.3/ 0.7/ 0.0001
	Y	10	51.91/ 14.27/ 13.29	50.75/ 12.64/ 12.47	5.73/ 3.61/ 2.63	1.81/ 1.14/ 0.83	0.12/ 0.07/ 0.05	7e-10/ 0.0006/ 1.7e-07
<b>Y chromosome MEF W8 (S19A)</b>	early	50	285.24	125.95	413.16	58.43	3.68	< 2.16e-16
	mid	10	401.8	284.74	399.81	126.43	8.15	1.7e-07
	late	55	413.38	345.17	236.73	31.92	2.01	< 2.16e-16
	Y	83	1581.58	1323.29	1013.69	111.22	7.00	-

**Supplementary Table 11: Statistic parameters for Figure 1C and 5.**

Stage	N	Mean	SD
I	368	24.7	0.6
II	458	30.7	0.85
III	239	15.5	0.5
Y	91	6.1	0.1
non S	356	22.9	0.8

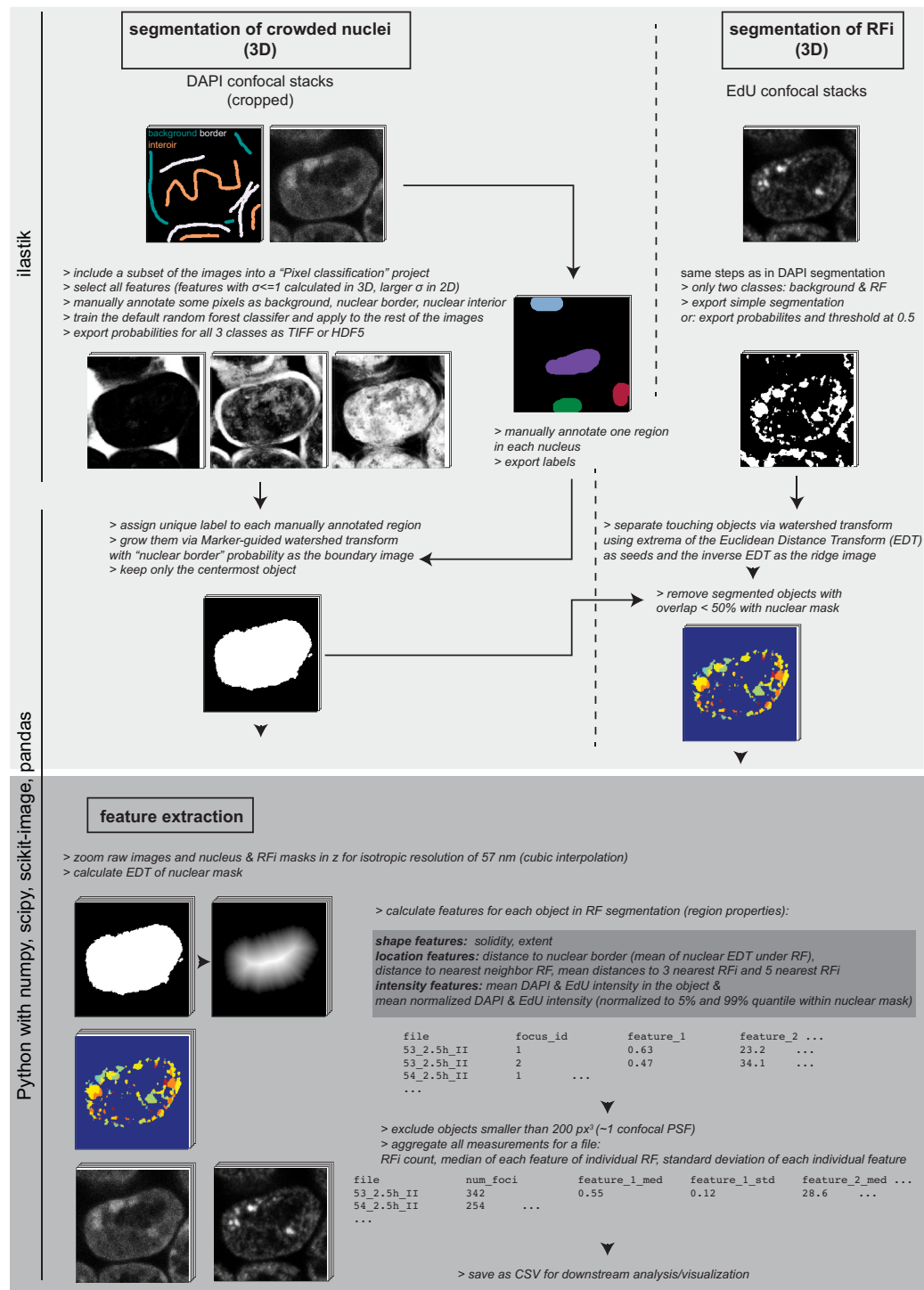
**Supplementary Table 12: Statistic parameters for Figure 6.**

Figure	Substage	N	Mean	Median	SD	SEM	CI	P-value
<b># RFi (6B)</b>	I	24	3182	3160	557	114	7.2	-
	II	28	3567	3398	542	102	6.5	-
	III	36	3217	3248	526	88	5.5	-
	Y	7	1881	1607	831	314	314	-
<b>nanoRFi/pWF RFi (6D)</b>	C2C12	16	5.99	-	0.42	0.11	0.23	-
	mES cell	30	6.5	-	0.39	0.12	0.26	0.0034
<b>Clustered nanoRFi per pWF (6D)</b>	C2C12	16	3.82	3	2.3	0.021	0.0014	-
	mES cell	30	4.01	3	2.6	0.027	0.0017	2.52e-08
<b>Volume pWF RFi (6E)</b>	C2C12	16	103.9	82.4	88.4	0.87	0.0054	-
	mES cell	30	121.6	71.6	115.6	1.07	0.0067	< 2.16e-16
<b>Volume nanoRFi (6F)</b>	C2C12	16	12.4	9	115.1	0.0046	0.0009	-
	mES cell	30	13.3	9.6	114.2	0.0049	0.0003	0.75

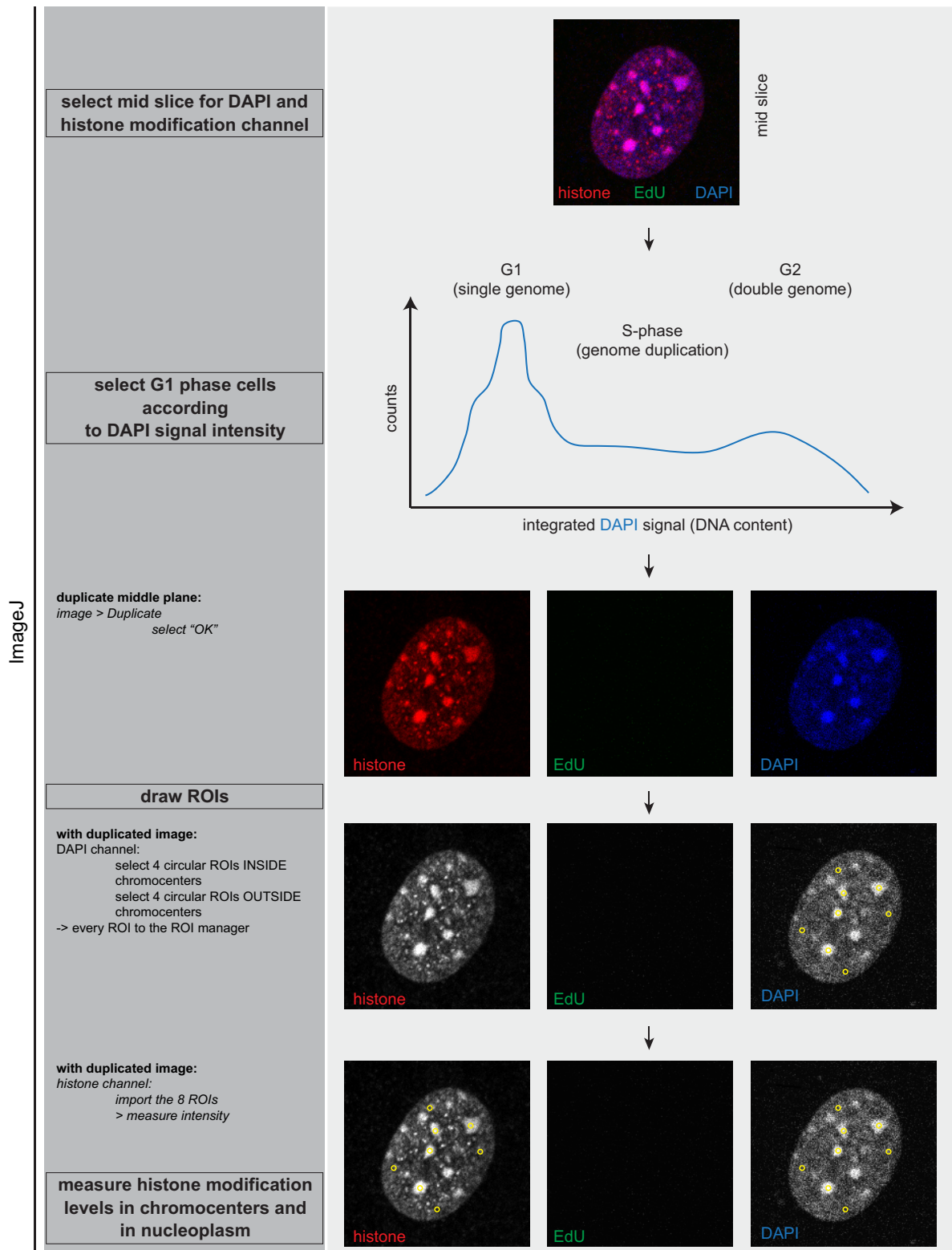
**Supplementary Table 13: Statistic parameters for Figure 7, Supplementary Figure S21 and S22.**

Figure	N	Mean	Median	SD	SEM	CI	P-value
mESC (7E)	71111	33.62	21.02	44.88	0.17	0.011	-
MEF (7E)	34159	69.93	44.96	85.21	0.46	0.029	< 2.16e-16
J1 karyotype (7H)	143	37.18	39	3.98	0.33	0.02	-
mESC (X chr) (S21A)	2645	62.23	37.17	188.49	3.67	0.23	-
MEF (X chr) (S21A)	1998	82.36	49.2	235.93	5.28	0.33	0.0017
mESC (autos.) (S21A)	71111	33.62	21.02	44.88	0.17	0.011	-
MEF (autos.) (S21A)	34159	69.93	44.96	85.21	0.46	0.029	< 2.16e-16
mESC (S22A)	71257	709.18	596	327.15	1.23	0.077	-
MEF (S22A)	34198	621.86	569	209.75	1.13	0.071	-
act. B cells (S22A)	8897	21415.53	17000	13618.36	144.38	9.054	-
mESC (S22B)	33746	70.85	58.89	54.52	0.3	0.019	-
MEF (S22B)	23843	100.19	75.07	90.57	0.59	0.037	< 2.16e-16
act. B cells (S22B)	8878	266.21	129	583.6	6.19	0.0388	< 2.16e-16

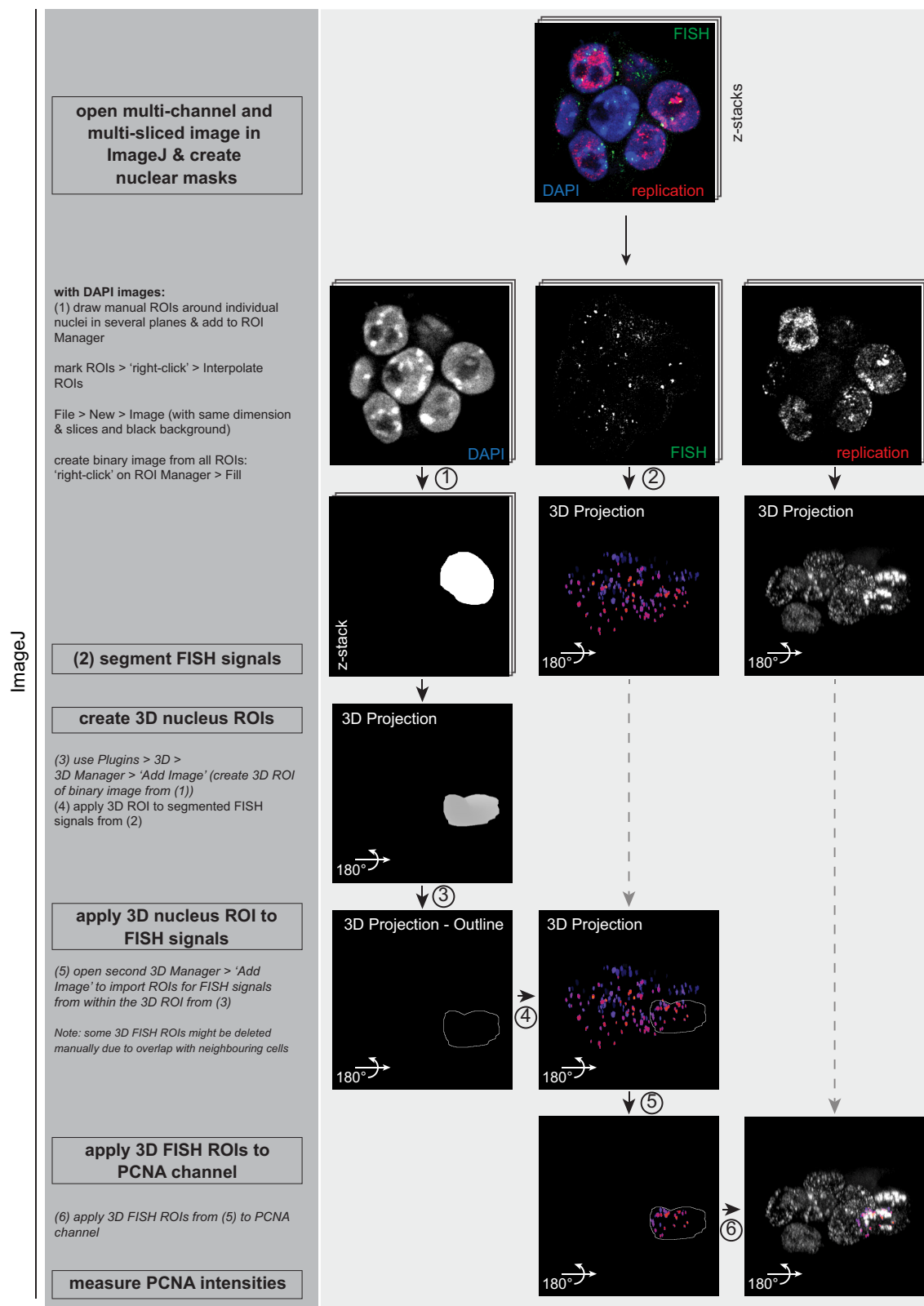
## SUPPLEMENTARY FIGURES



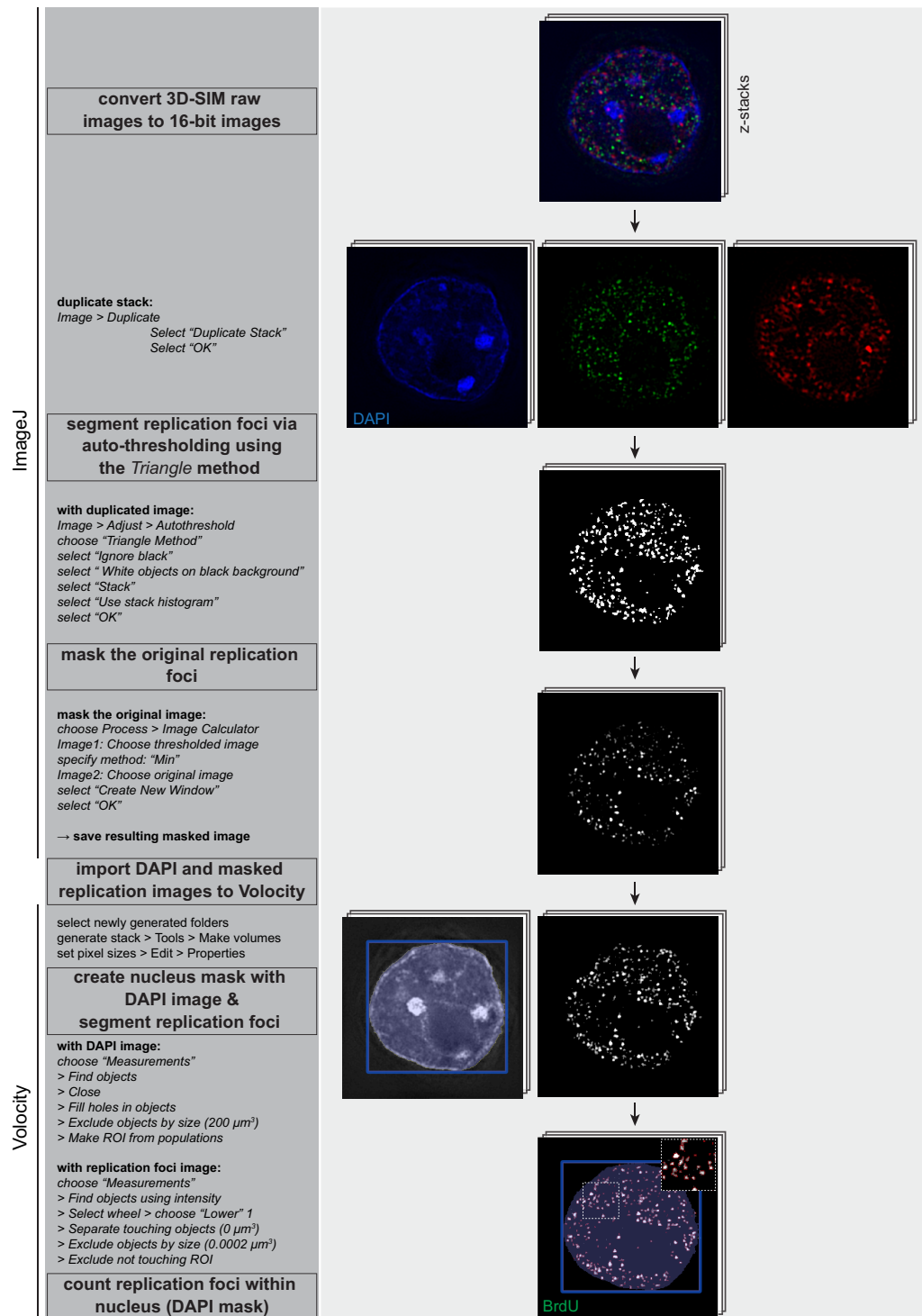
**Supplementary Figure S1 – Analysis pipeline of replication foci (RFi) features.** To extract quantitative features of RFi, DAPI stained nuclei and EdU labeled RFi were segmented in confocal stacks of mES cells. The central nucleus as well as RFi were segmented using supervised pixel classification (ilastik) and watershed transform (Python with scipy/scikit-image). Several features were calculated for each RF and used to create an aggregate feature vector for each image. A detailed analysis pipeline is described in the Materials and Methods section.



**Supplementary Figure S2 – Analysis pipeline of histone modification levels.** To determine the histone modification levels at chromocenters, DAPI intensities of all cells were measured at middle planes and cells were grouped into G1, S and G2 phases, accordingly. Four circular regions of interest (ROI) were drawn in middle plane G1 phase cells inside DAPI intense stained chromocenters. Similarly, four circular ROIs were drawn outside of chromocenters and intensity levels within these ROIs were measured in the histone modification channel. A detailed analysis pipeline is described in the Materials and Methods section.

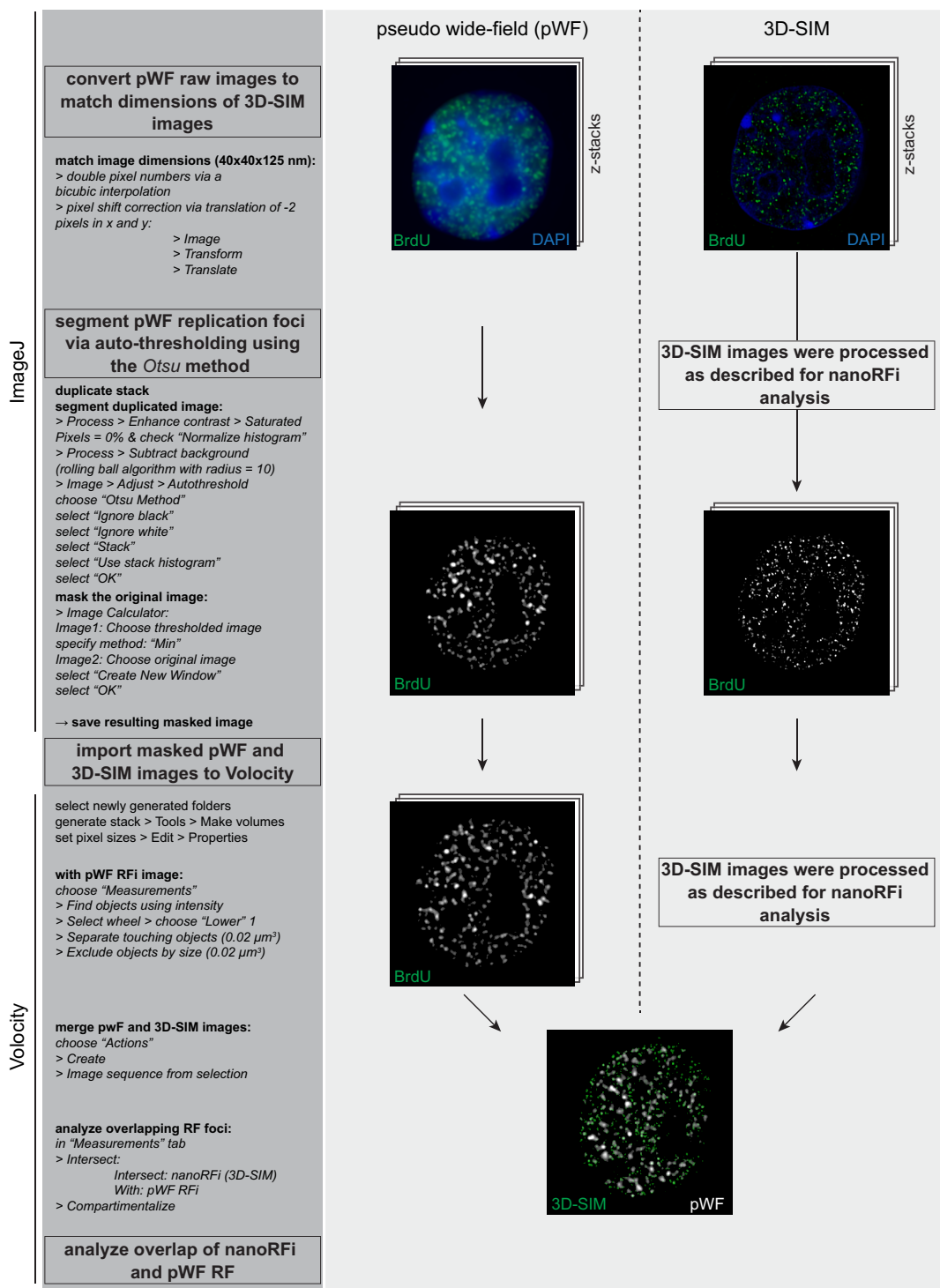


**Supplementary Figure S3 – Analysis pipeline of Repli-FISH analysis.** To determine the replication timing of specific (sub)chromosomal elements, 3D masks of individual cell nuclei were manually generated based on the DAPI channel. FISH signals were segmented. The 3D nucleus mask was applied to the segmented FISH signals and the generated nuclear FISH ROIs were used to mask the PCNA signals and PCNA intensities within the mask were measured. A detailed analysis pipeline is described in the Materials and Methods section.

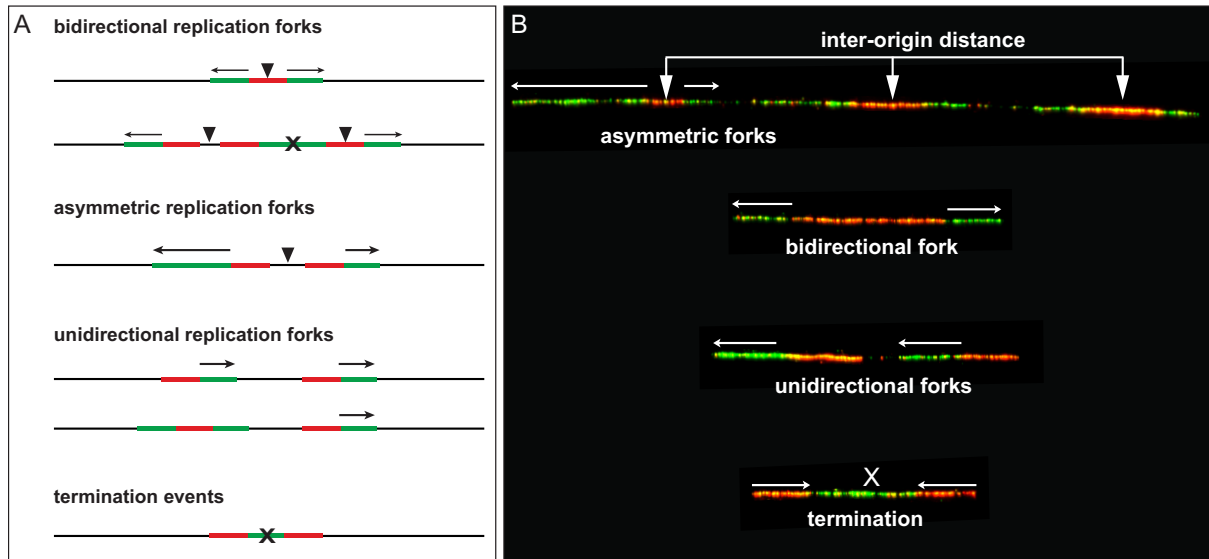


### Supplementary Figure S4 – Analysis pipeline of 3D structured illumination microscopy (3D-SIM) resolved nano replication foci counting.

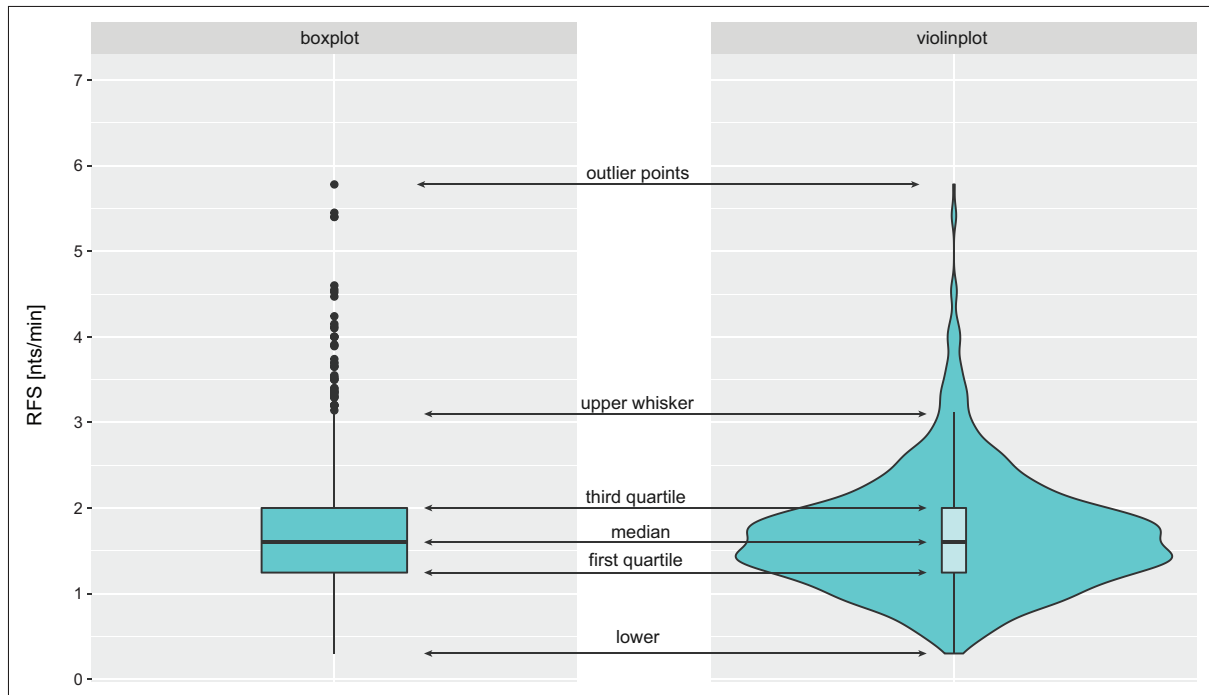
For counting of nano replication foci (nanoRFi) in 3D-SIM images, analysis was done using ImageJ and the PerkinElmer Volocity software. Raw image files were converted to 16-bit images, individual cell nuclei were segmented based on the DAPI channel, replication signals were thresholded and the binary images were used to mask the original replication foci signals. DAPI and masked replication foci images were imported to Volocity to quantify replication foci for individual nuclei by creating nuclear masks and corresponding ROIs (regions of interest) were based on the DAPI images. Next, 3D-SIM replication foci were detected by intensity and only foci within the nuclear ROI were counted. A detailed analysis pipeline is described in the Materials and Methods section.



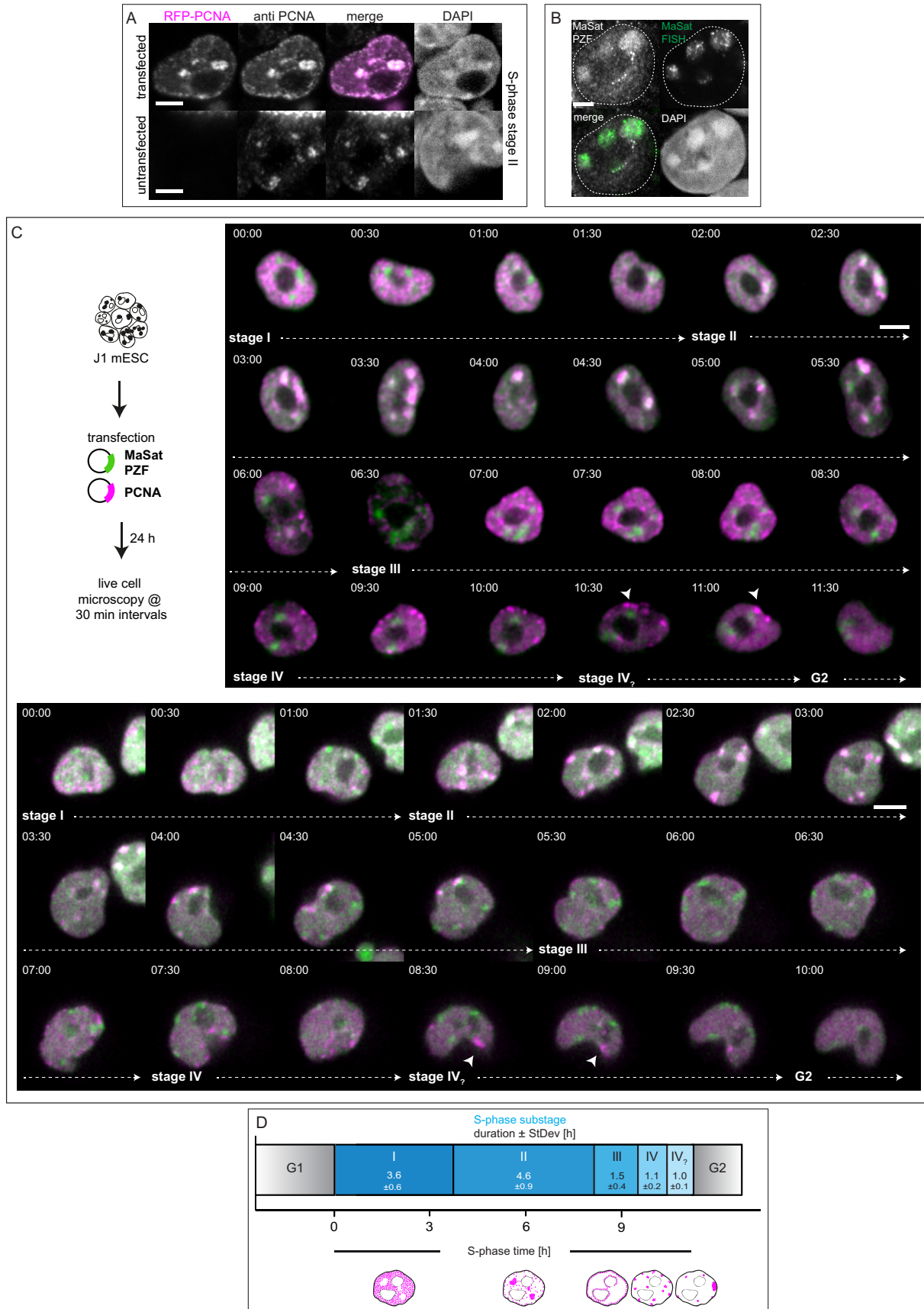
**Supplementary Figure S5 – Analysis pipeline of nano replication foci per pseudo-widefield replication focus quantification.** For counting nano replication foci (nanoRFi) per pseudo-widefield (pWF) replication focus, analysis was done using ImageJ and the PerkinElmer Volocity software. Dimensions of pWF replication raw images were matched to the image dimension of the 3D-SIM data (40 x 40 x 125 nm). Individual cell nuclei were segmented based on the DAPI channel, histograms of replication signals were normalized, background was subtracted, signals were thresholded and the binary images were used to mask the original replication foci signals. 3D-SIM replication signals were processed as described in Supplementary Figure S4. Segmented and masked pseudoWF and 3D-SIM image stacks were imported to Volocity and merged. Detection of pWF RFI was based on intensity as for 3D-SIM images and the number of nanoRFi contained in one pWF RF were counted.



**Supplementary Figure S6 – Analysis of the mouse embryonic stem cell replicon by DNA fiber assays. (A)** Classification of typical DNA fiber signals obtained from molecular combing experiments. Replication fork speed is calculated from measurements of green replication signals (arrows) flanking red tracks that mark initiation events during the first pulse labeling period (origins of replication marked by black arrowheads). Merging forks within clusters of origins and termination events are discarded (marked with X). Fork asymmetry can be assessed by calculating the ratio of left/right replication forks, emanating from an origin of replication. Replication signals spreading in only one direction are considered as unidirectional replication forks. **(B)** Representative examples of replication signals from DNA combing experiments. Labeling is according to the classification described in (A). The inter-origin distance (IOD) is the distance between neighboring initiation sites within a cluster, identified based on the first nucleotide label (IdU, red).



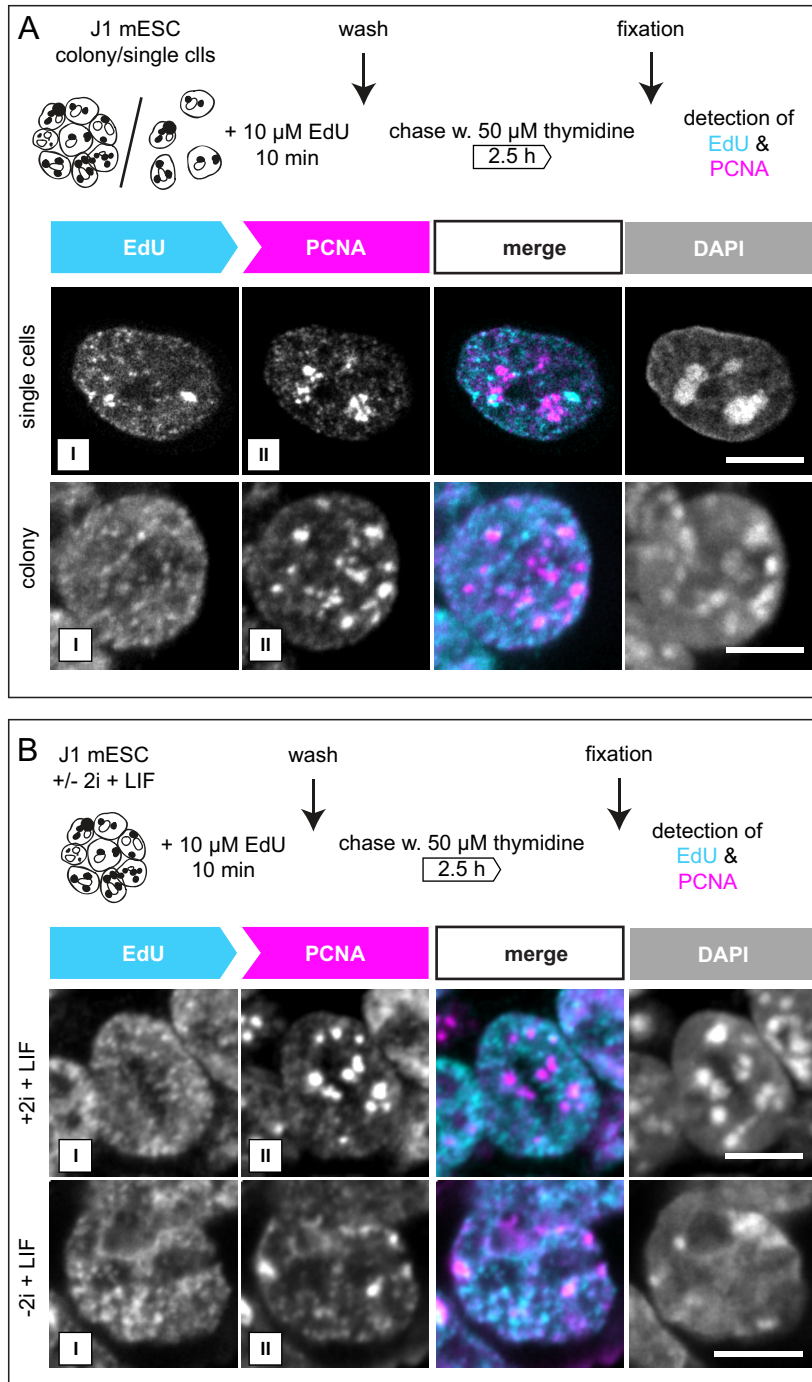
**Supplementary Figure S7 – Statistics representation using violin plots.** Statistical analyses were represented by violin plots, a variation of boxplots with a kernel density plot on each side. Similar to a boxplot, boxes and whiskers represent 25–75 percentiles and 1.5 times the IQD (inter-quartile distance), respectively and the center line depicts the median. Additionally, violin plots represent the probability density of the data at different values.



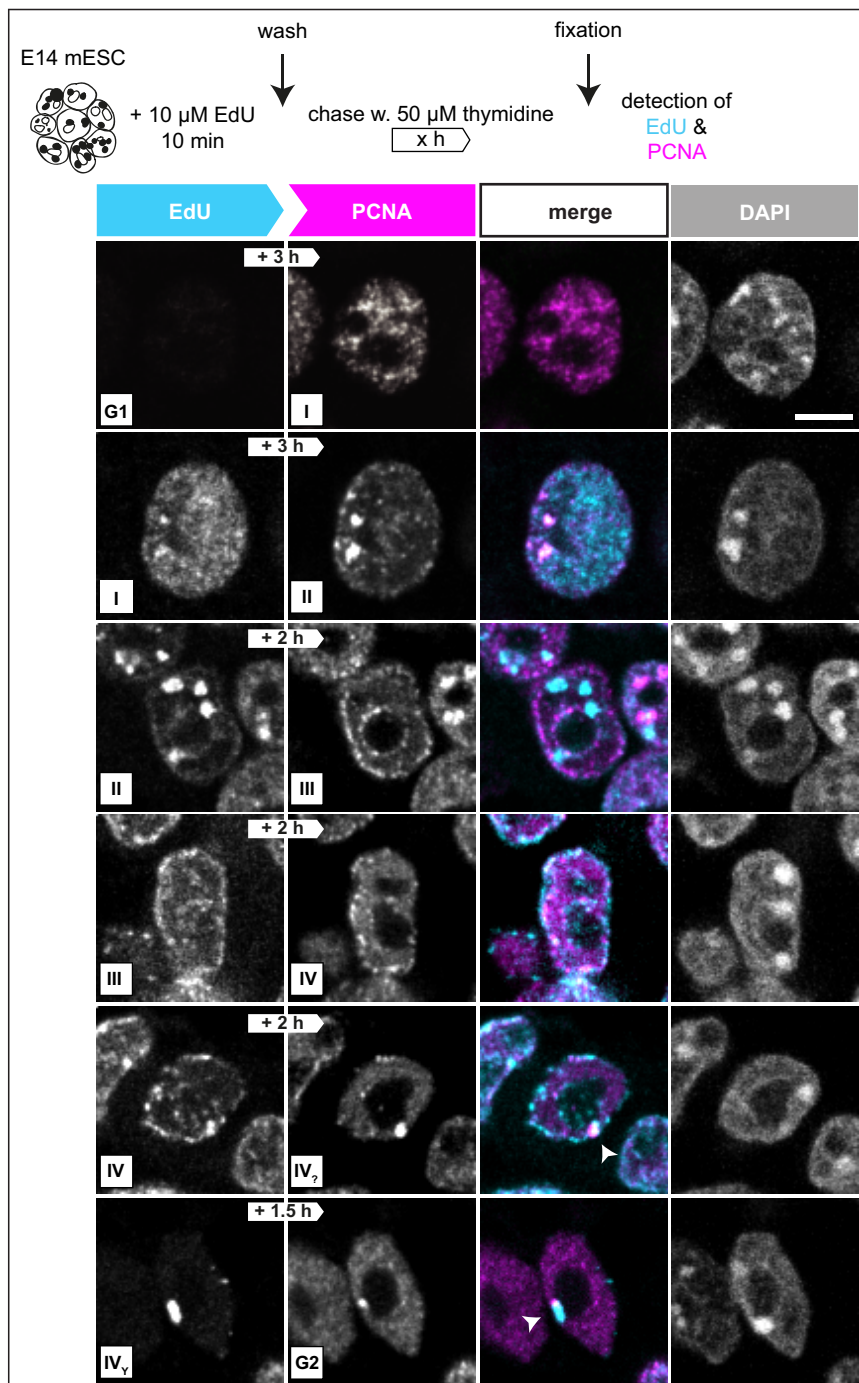
**Supplementary Figure S8 – Live-cell analysis of DNA replication dynamics in mouse embryonic stem cells.** **(A)** J1 mES cells were transfected with mRFP-PCNA and stained with a PCNA specific antibody. Representative confocal images of the anti-PCNA staining (grey) in an mRFP-PCNA (magenta) transfected and an untransfected cell are shown. **(B)** MaSat-GFP polydactyl zinc finger (PZF) transfected cells were fixed using a formaldehyde gradient with simultaneous permeabilization and major satellite repeat sequences were visualized by fluorescence *in situ* hybridization (FISH) with a MaSat specific probe (green). Representative confocal images are shown. **(C)** mES cells were co-transfected with plasmids encoding fusion constructs of mRFP-PCNA (magenta) and GFP-tagged polydactyl zinc finger protein binding to major satellite sequences (MaSat-GFP, green) to follow S-phase progression and to mark pericentromeric heterochromatin (chromocenters), respectively, and imaging was performed for 24 hours with 30 minutes intervals. Spinning disk confocal images of a representative cell progressing through S-phase (stage I to IV<sub>?</sub>) and G2 phase. **(D)** Approximate durations of the individual substages (mean  $\pm$  StDev) were calculated from at least 15 cells (Supplementary Table 7) from different live cell experiments. The arrowheads in the IV<sub>?</sub> stage mark a prominent accumulation of replication signals observed at the end of S-phase. Scale bar = 5  $\mu$ m.



**Supplementary Figure S9 – Description of the pulse chase experiments. (A)** Experimental setup of a pulse-chase experiment to determine the spatio-temporal progression of DNA replication in mES cells. Asynchronously growing mES cell cultures were pulse labeled with the nucleotide analog EdU, followed by various thymidine chase periods (white arrows) and fixation. Subsequently, replication foci marked by the incorporated EdU, i.e. nascent DNA during the first pulse labeling (cyan), and the replisome component PCNA, i.e. ongoing replication at the time point of fixation (magenta), were (immuno)fluorescently detected (Figure 1C). The different chase times resulted in different degrees of replication foci separation and, with increasing chase times, eventually also to transition from one S-phase pattern to the next. Representative confocal images of EdU pulsed cells followed by different chase times before fixation, are shown. **(B)** Schematic overview of the S-phase progression with different chase durations. Durations of the different S-phase stages are indicated on the left. Chase durations in the fixed cell approach were changed according to S-phase substage duration. While shorter chase times (< 2 hours/ 2.5 hours for stage I - II and II - III, respectively and < 0.75 hours for stage III - IV and IV - IV<sub>2</sub> transitions) do not visualize substage transitions in a substantial amount of cells, increasing chase times reveals changes in S-phase patterns. This approach allowed us to get a spatio-temporal resolution of DNA replication in fixed cells. **(C)** Schematic representation and representative confocal images of a pulse-chase-pulse-chase experiment, underlining the domino-like DNA replication model. Scale bars = 5  $\mu\text{m}$ .

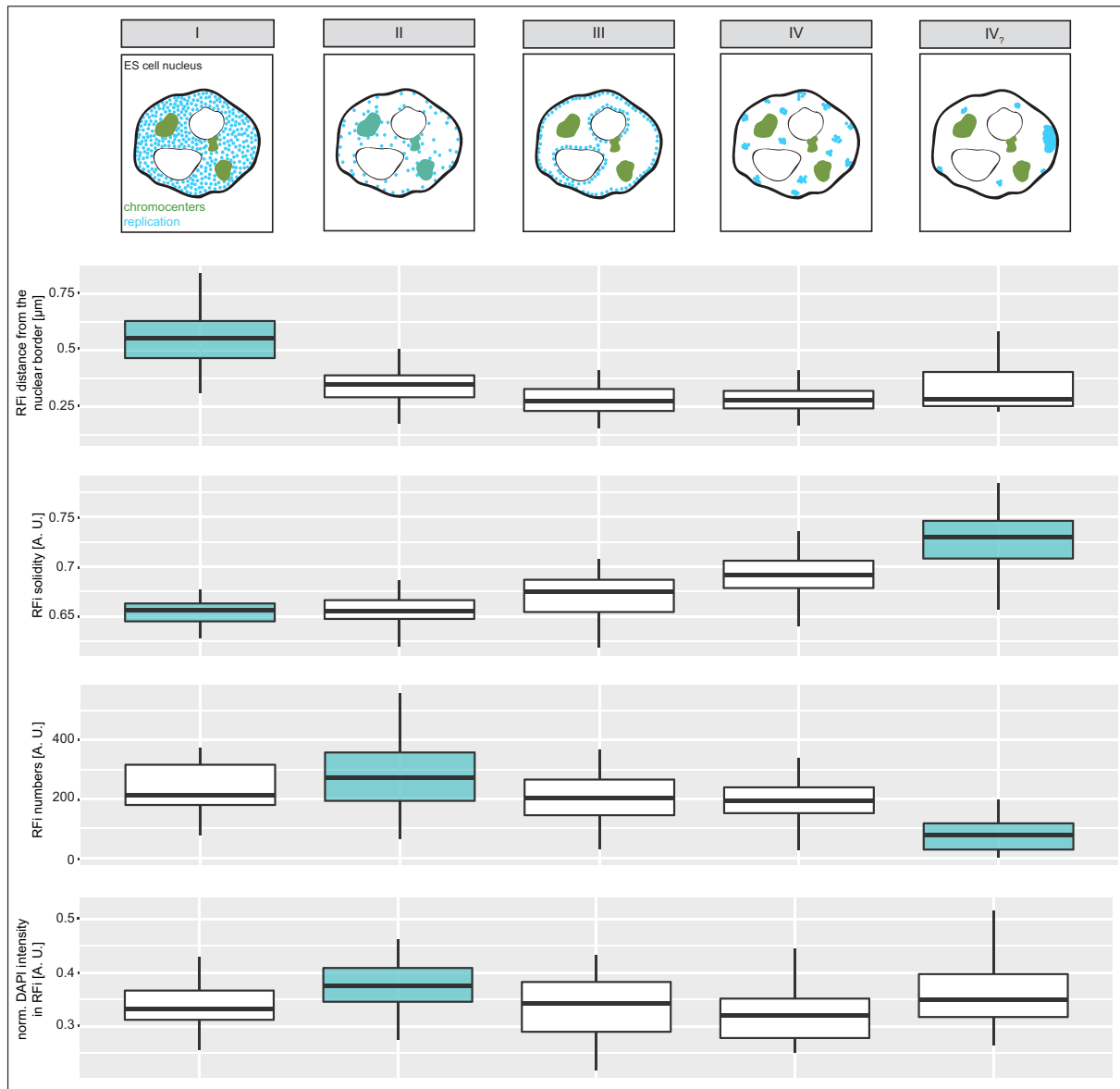


**Supplementary Figure S10 – DNA replication dynamics of mouse embryonic cells in different microenvironments. (A)** J1 mES cells were pulse chased as described in Figure 1, 6 hours (single cells) or 24 hours (colonies) after seeding, EdU and PCNA were (immuno)fluorescently detected and S-phase progression was analyzed. Representative confocal images of cells passing from stage I to stage II are depicted for single cells and cells grown in colonies. **(B)** J1 mouse ES cells were grown for 15 days in medium containing only LIF or 2i and LIF and pulse chased 24 hours after seeding. Confocal images of S-phase progression from stage I to stage II are depicted as above. Scale bars = 5  $\mu$ m.

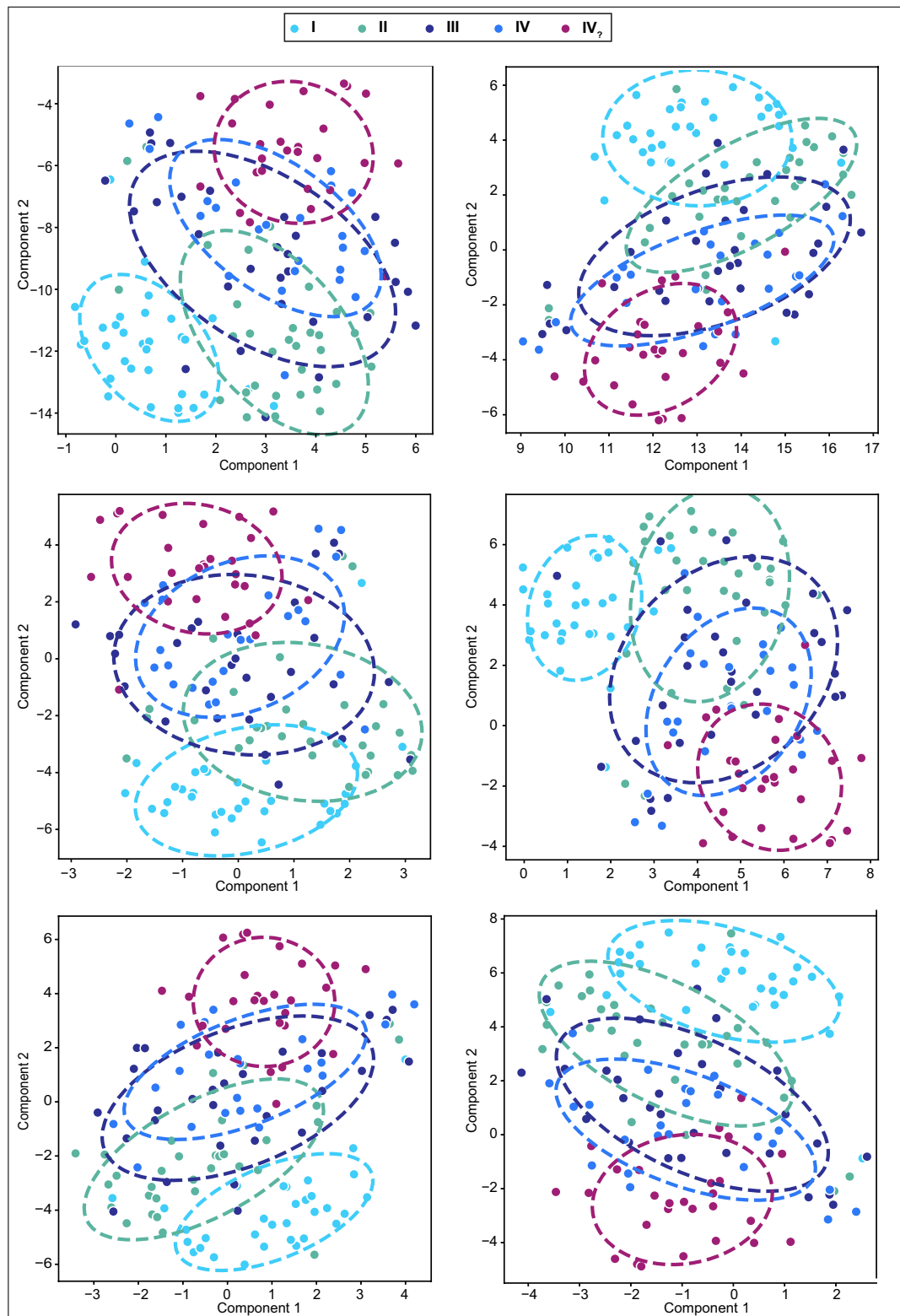


**Supplementary Figure S11 – DNA replication dynamics in mouse E14 embryonic stem cells.**

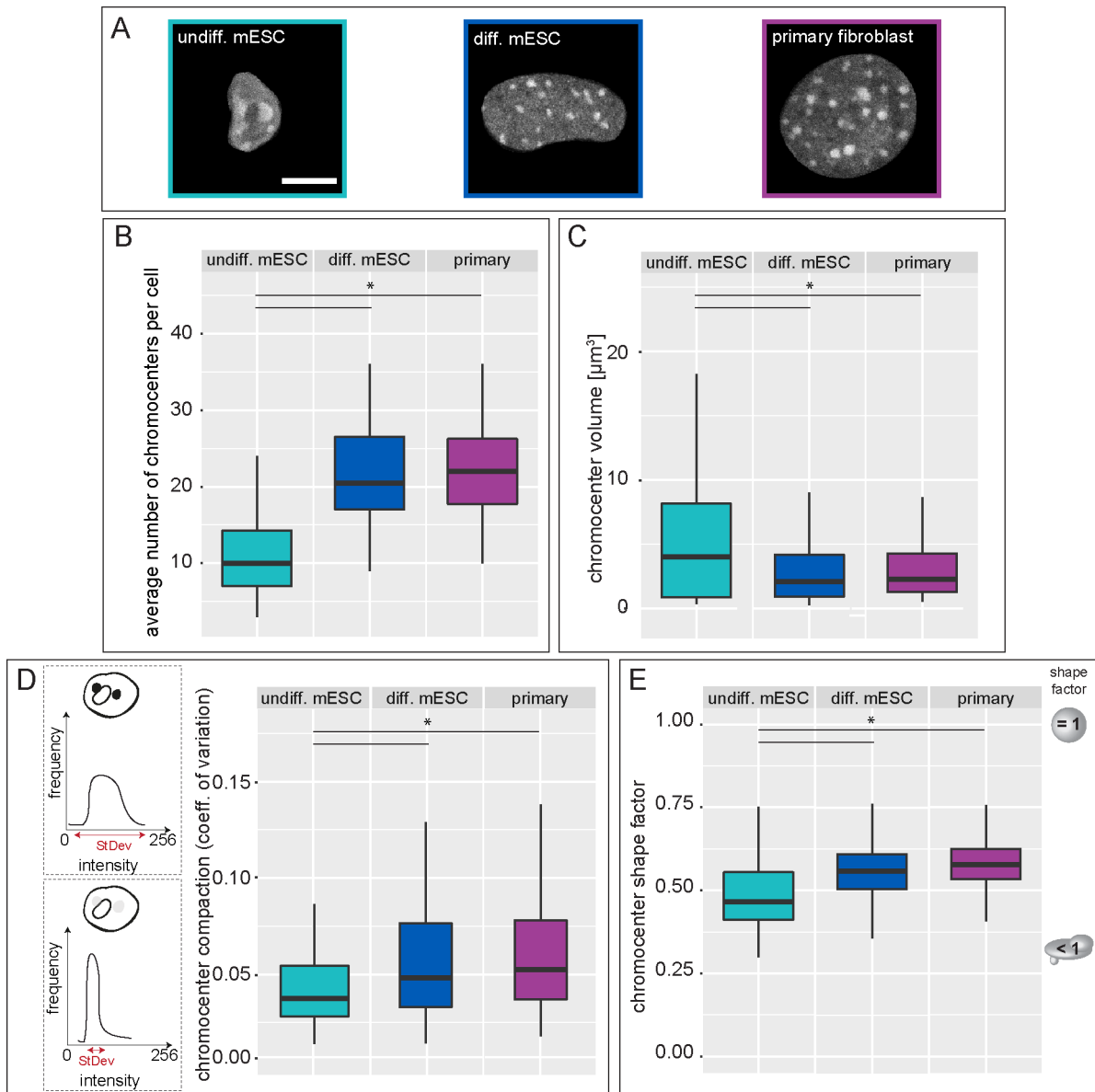
The experimental pulse chase setup (described in Figure 1), allowed the identification and temporal order of five main replication patterns in mouse E14 ES cells. Representative spinning disk confocal images of transitions from G1 to S-phase, S-phase substage transitions and S-phase to G2 progression are shown. The arrowheads in the IV<sub>2</sub> stage mark a prominent accumulation of replication signals observed at the end of S-phase. White arrows represent chase periods. Scale bar = 5  $\mu\text{m}$ .



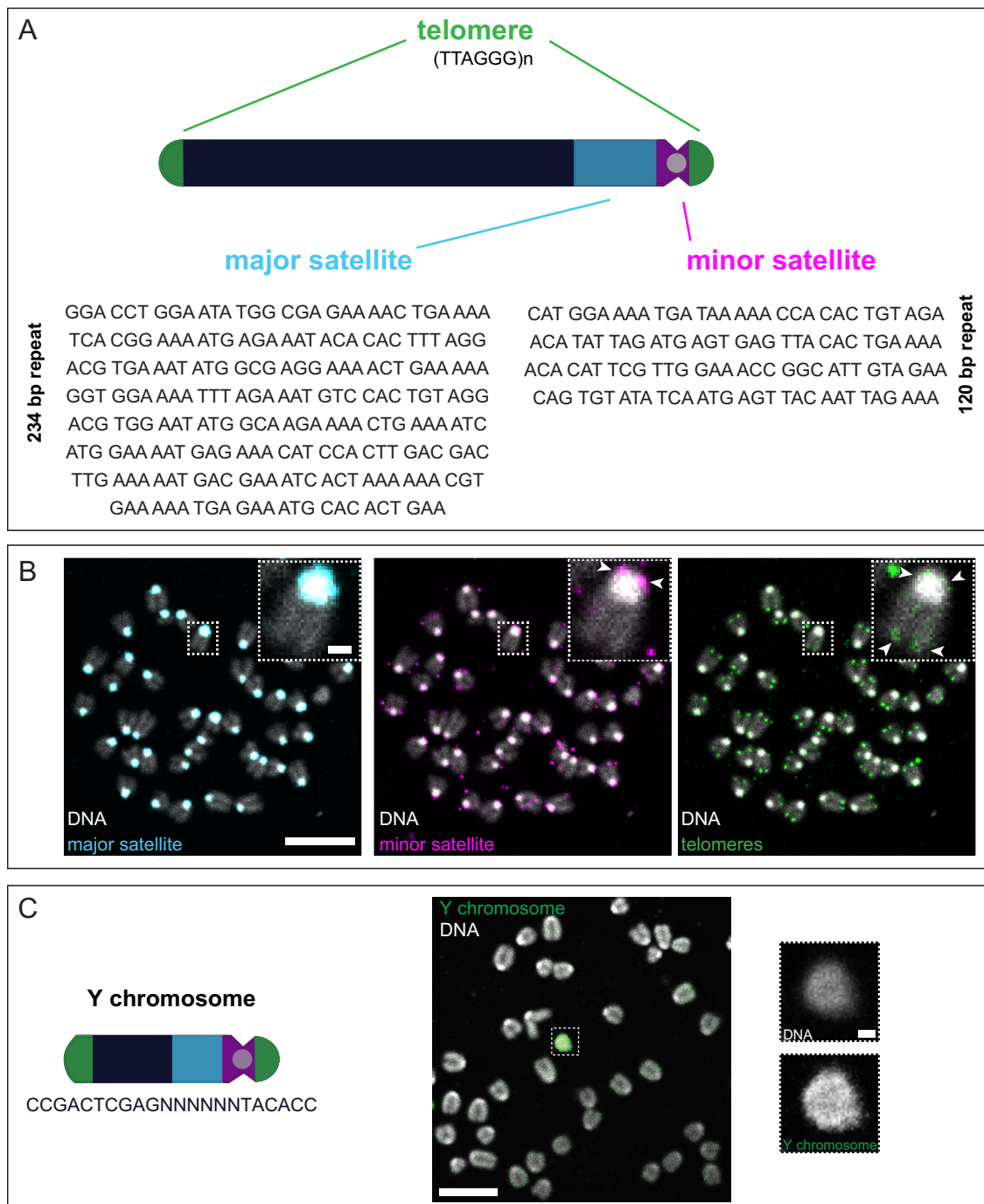
**Supplementary Figure S12 – Characteristics of replication foci in the five S-phase substages of mouse embryonic stem cells.** Replication foci (RFi) distance from the nuclear border, RFi solidity, RFi numbers and normalized DAPI intensities within RFi are plotted for the five different mES cell S-phase stages. Cyan marked boxplots represent features that differ significantly from one stage to the other. All boxplots are as in Figure 2 and Supplementary Figure S7. Detailed statistics are summarized in Supplementary Table 7.



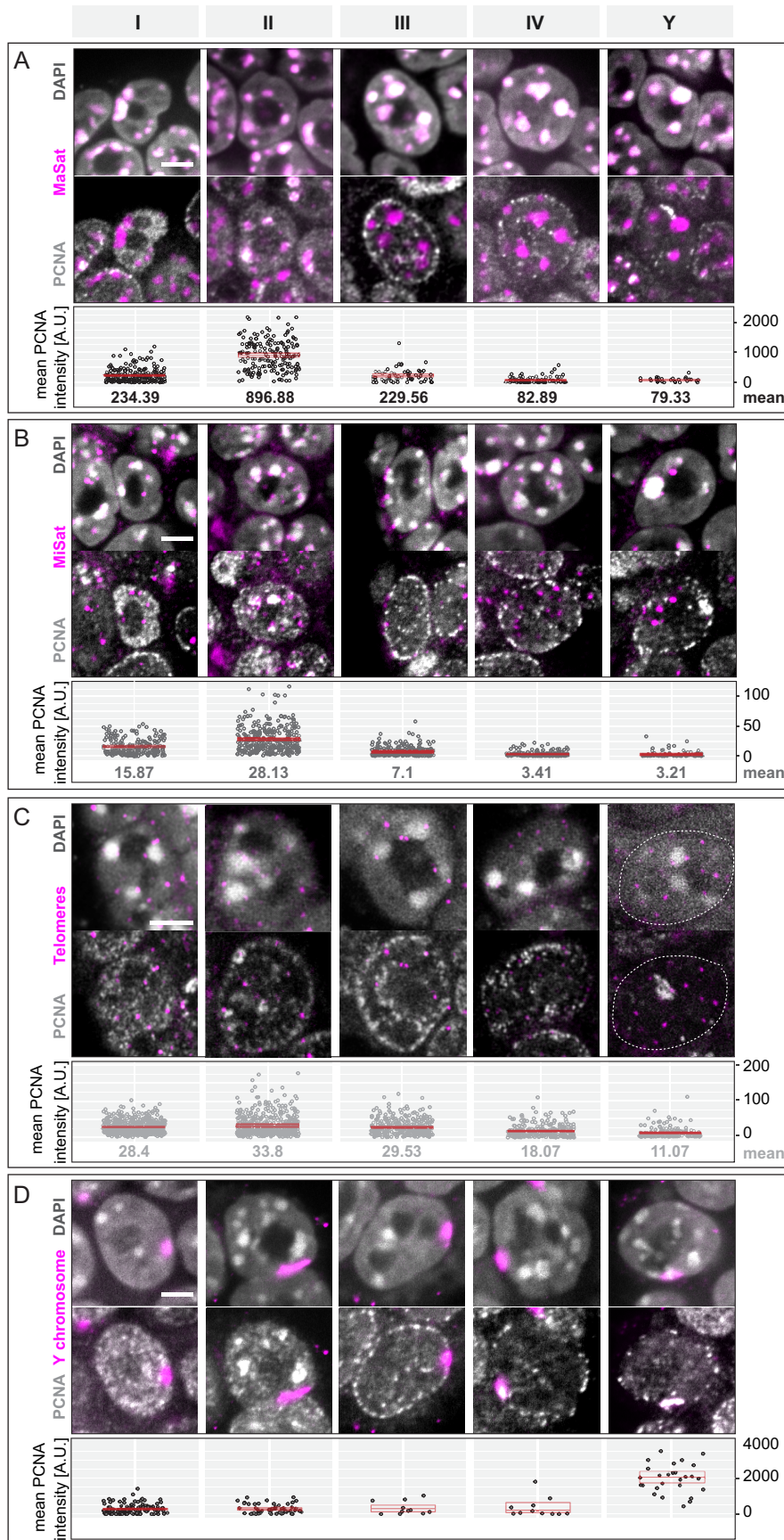
**Supplementary Figure S13 – t-distributed stochastic neighbor embedding (t-SNE) analysis and replication foci (RFi) characteristics distributions. (A)** t-SNE embedding was used to visualize the high-dimensional RFi characteristics. Each high-dimensional object was modeled by a 2 dimensional point. Similar objects are modeled by nearby points and dissimilar objects are modeled by distant points. Six variants of t-SNE representations are shown. Multinomial logistic regression of the S-phase stage from the image features reached >70% accuracy. When also allowing for one-off errors (e.g. stage I classified as II), the accuracy was over >90% (accuracy determined in 5-fold cross-validation).



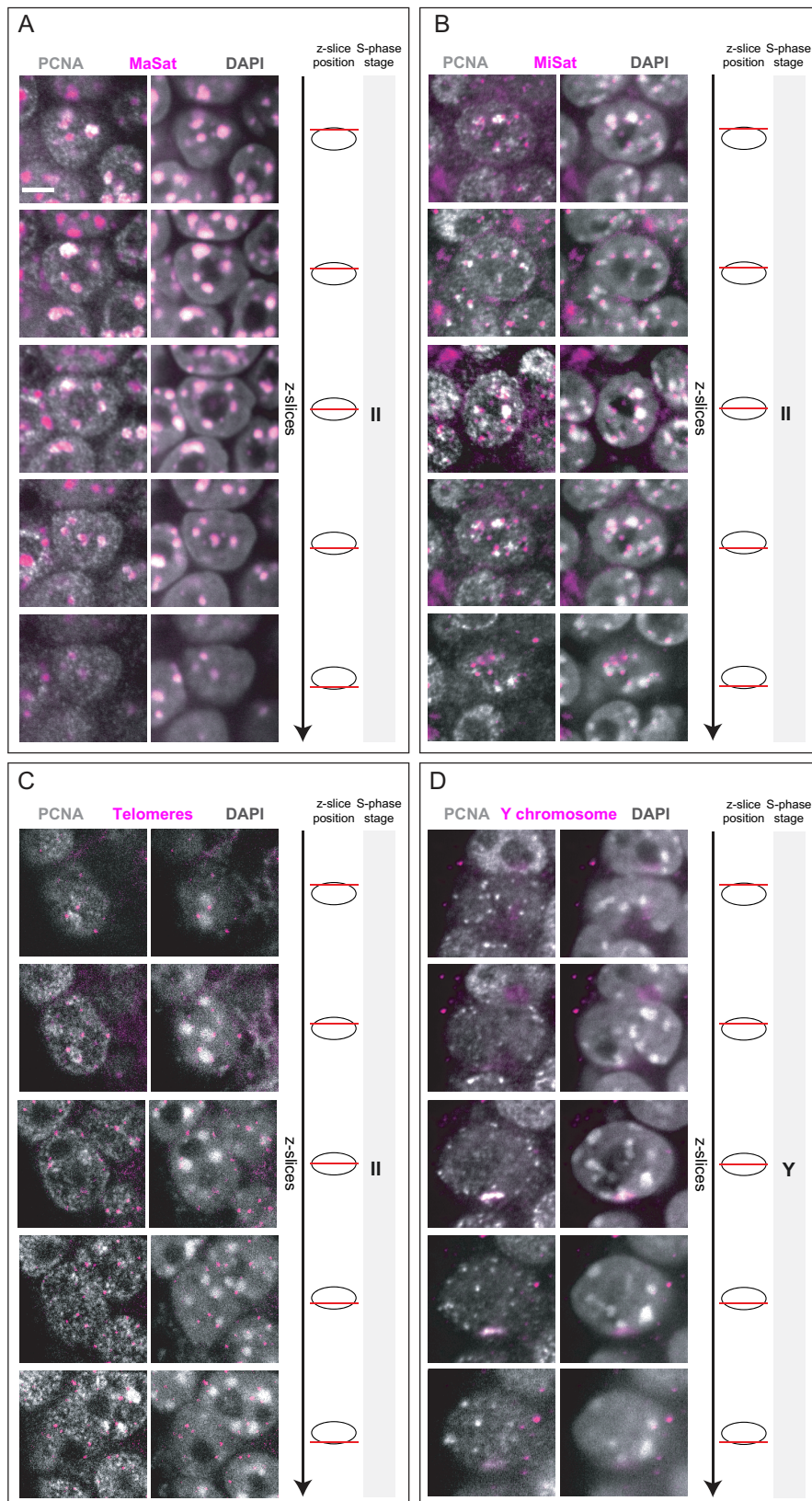
**Supplementary Figure S14 – Characteristics of pericentromeric DNA clusters (chromocenters) in (un)differentiated mouse embryonic stem cells and mouse myoblast cells.** (A) Representative cell nuclei from undifferentiated J1 mES cells (cyan), differentiated J1 mES cells (blue) and primary mouse ear cells (magenta). Average numbers of chromocenters per cell (B), volumes of chromocenters (C), chromocenter compaction (coefficient of variation of DAPI (StDev/mean, (D)) and chromocenter shape factor (E) show the different organization of pericentromeric heterochromatin in mouse somatic and pluripotent mES cells. In (D), condensed chromatin is characterized by high DAPI intensities. Thus, cell nuclei containing condensed chromatin show a higher DAPI standard deviation than cells containing decondensed chromatin, which is characterized by low DAPI standard deviations. Normalized DAPI standard deviation values, as a proxy for DNA decondensation levels, are plotted. All boxplots are as in Figure 2 and Supplementary Figure S7. Detailed statistics are summarized in Supplementary Table 9. Scale bar = 5  $\mu\text{m}$ . \*  $P < 0.05$ .



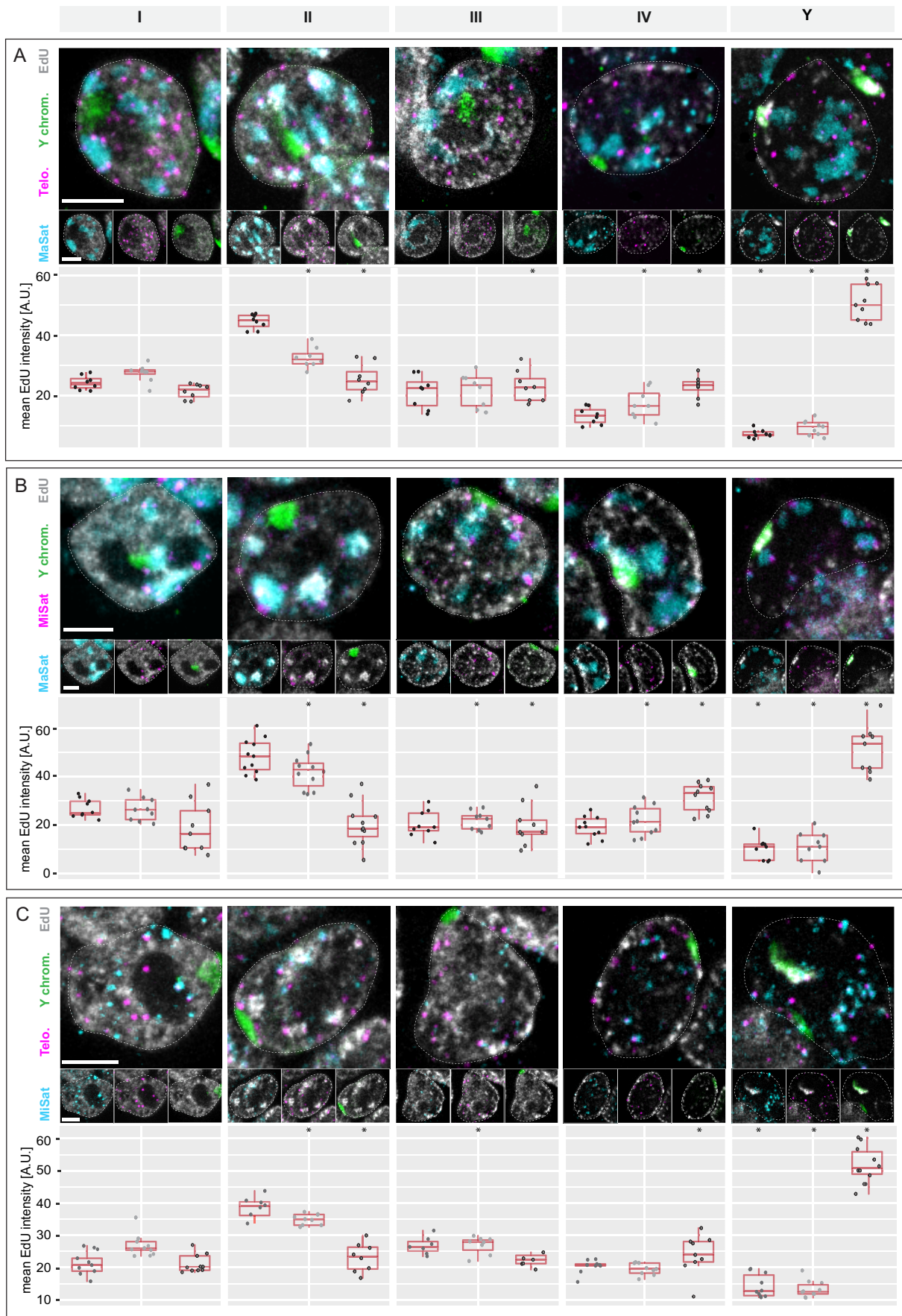
**Supplementary Figure S15 – Validation of FISH probes for (sub)-chromosomal elements. (A)** Schematic overview of the chromosomal location and sequence motifs of the analyzed major and minor satellites (MaSat (cyan) and MiSat (magenta)) and telomere (green) repeat elements in mouse cells. **(B)** Triple-color FISH combining probes labeled with different nucleotide analogs, specific for major satellites (cyan), minor satellites (magenta) and telomeres (green) hybridized to metaphase chromosome preparations from diploid mouse cells. Insets show a magnification of one chromosome and arrows show specific probe binding. **(C)** Schematic overview of the mouse Y chromosome and the primer sequence used for probe generation (left). Hybridization of the Y chromosome specific probe to mouse metaphase spreads. Inset shows specific probe annealing to one chromosome (right). Scale bar = 10  $\mu$ m and 1  $\mu$ m for main graphs and magnified regions, respectively.



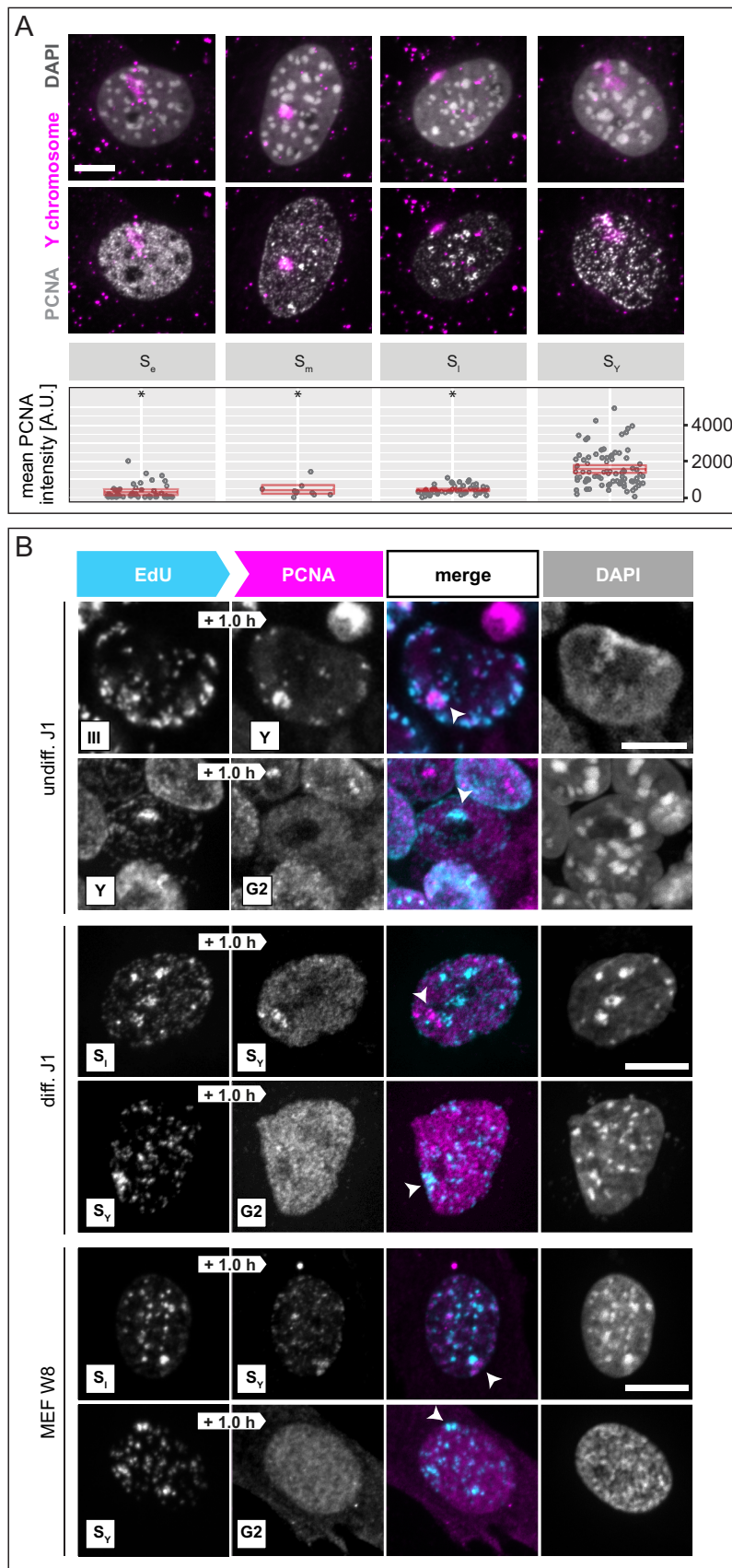
**Supplementary Figure S16 – Replication timing of (sub-)chromosomal structures in mouse embryonic stem cells.** Tandem repeat elements (magenta, major satellites (MaSat, **(A)**), minor satellites (MiSat, **(B)**) and telomeres (**(C)**) and the Y chromosome (**(D)**) were co-visualized with PCNA (grey) via Repli-FISH in mES interphase cells. Cells were classified into S-phase stages I to Y according to their PCNA pattern, mean PCNA intensities within the marked elements were measured as described in Supplementary Figure S3 and plotted. Red boxes represent mean  $\pm$  95% confidence intervals and mean values are indicated below each plot. DAPI (grey) and hybridization signals (magenta) are shown as a merge. Dotted lines represent cell contours. Detailed statistics are summarized in Supplementary Table 10. Scale bar = 5  $\mu$ m.



**Supplementary Figure S17 – 3-dimensional nuclear distribution of (sub-)chromosomal structures in mouse embryonic stem cells.** Tandem repeat elements (magenta, major satellites (MaSat, **(A)**), minor satellites (MiSat, **(B)**) and telomeres (**(C)**) and the Y chromosome (**(D)**) were co-visualized with PCNA (grey, left column) or DAPI (grey, right column). Mid z slice from stage II cells (A-C) or stage Y (D) depicted in Supplementary Figure S16 are shown. Additionally 2 slices above and two slices below the mid slice are represented to show the distribution of the marked elements within the mES cell nucleus. Scale bar = 5  $\mu\text{m}$ .

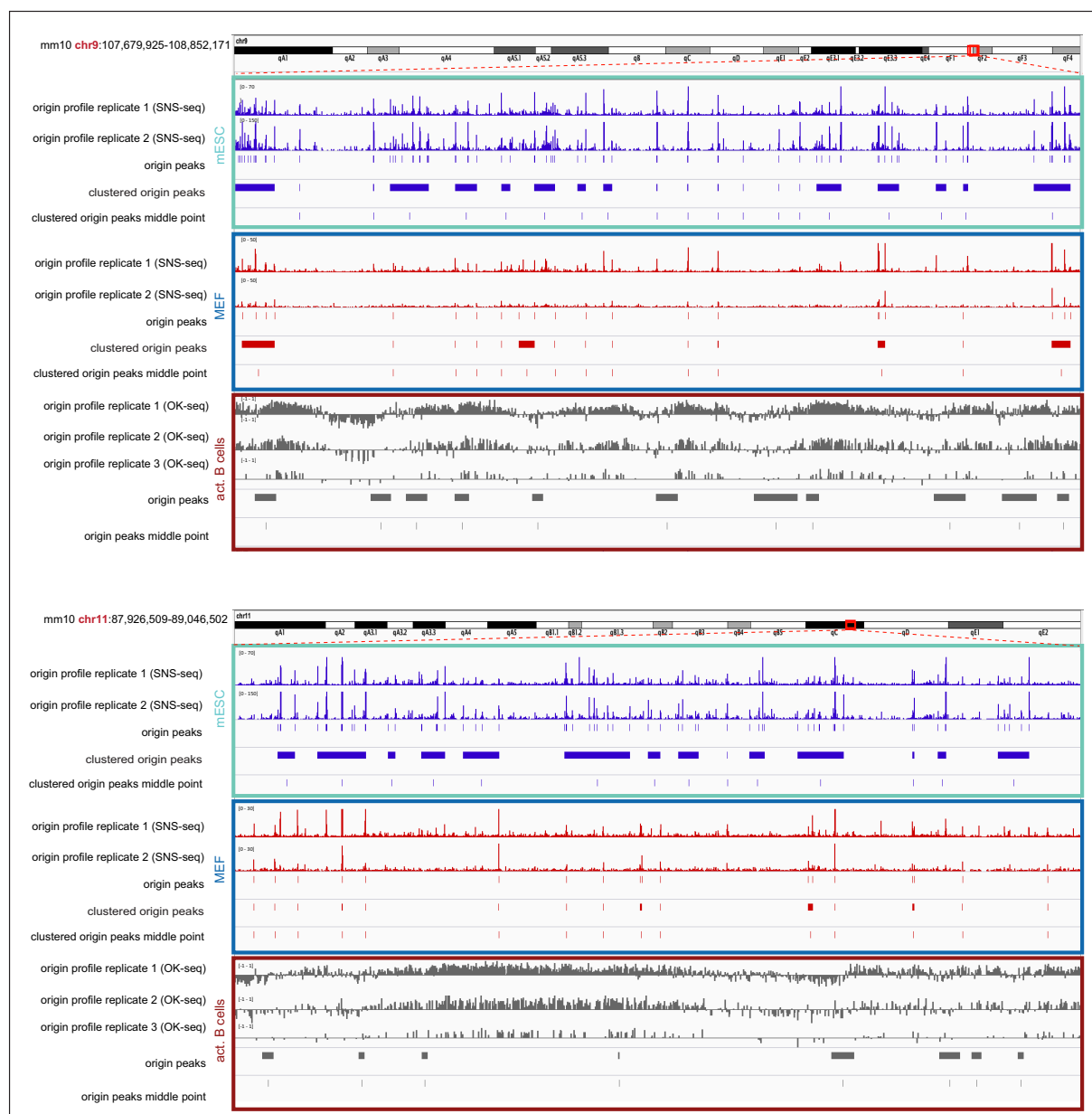


**Supplementary Figure S18 – Replication timing of subchromosomal structures and the Y chromosome in mouse embryonic stem cells.** EdU (DNA replication) and the Y chromosome were co-visualized with major satellite repeats and telomeres (**A**), with major and minor satellite repeats (**B**) and with minor satellite repeats and telomeres (**C**) in mES cell interphase nuclei. Cells were classified into S-phase stages I to Y according to their EdU pattern, mean EdU intensities within the marked elements were measured as described in Supplementary Figure S3 and plotted. Detailed statistics are summarized in Supplementary Table 10. Scale bar = 5  $\mu\text{m}$ . \*  $P < 0.05$  (calculated among the same element against the respective stage I value).

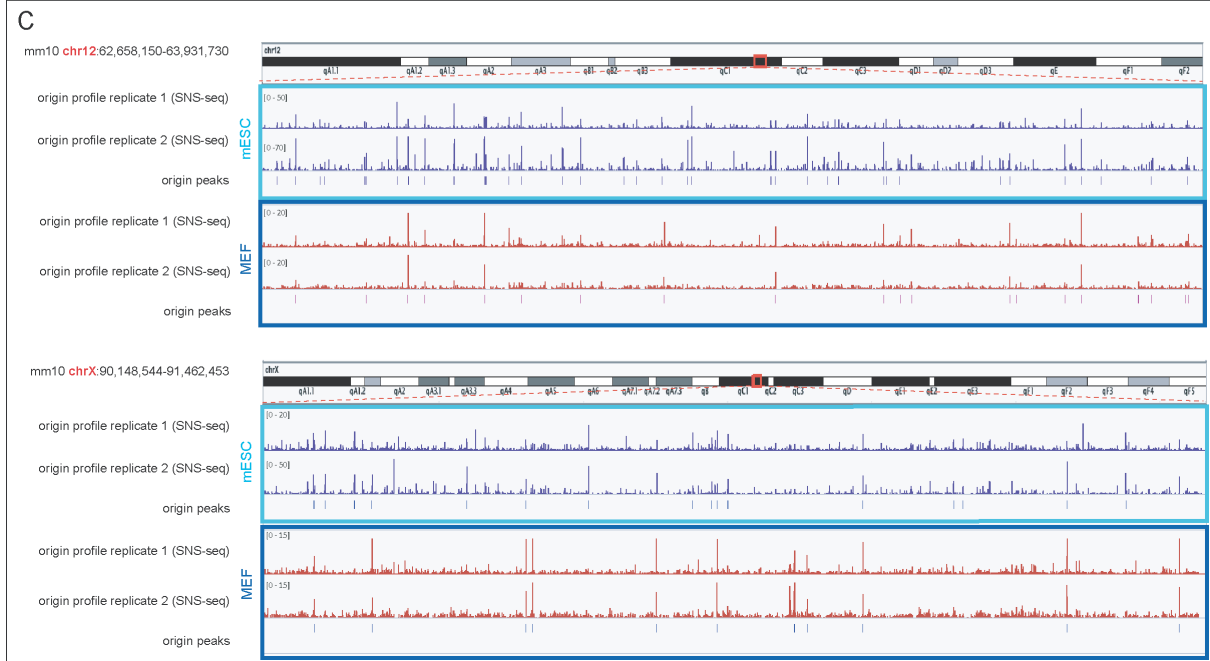
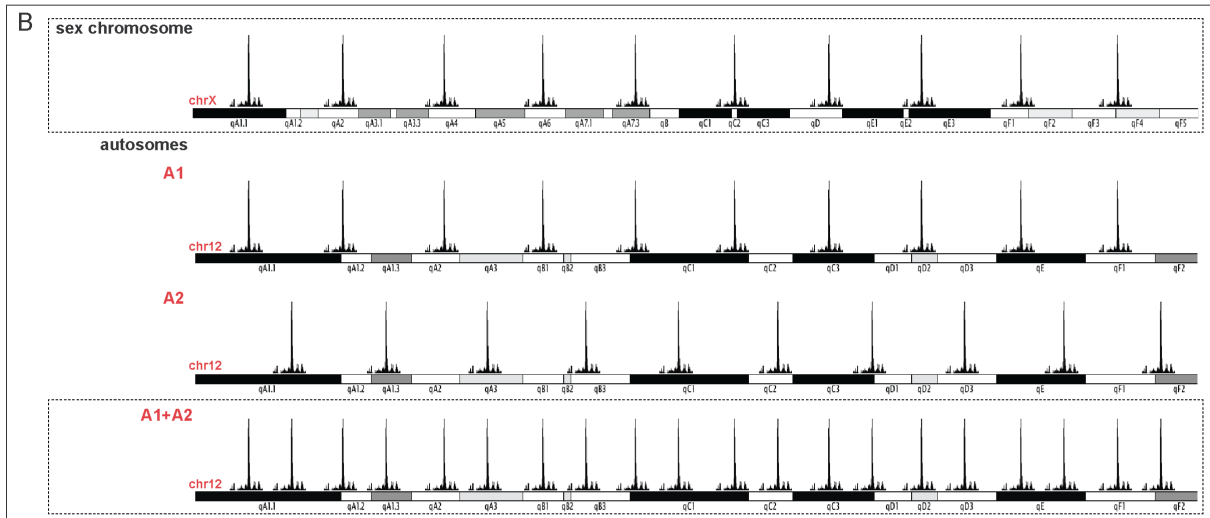
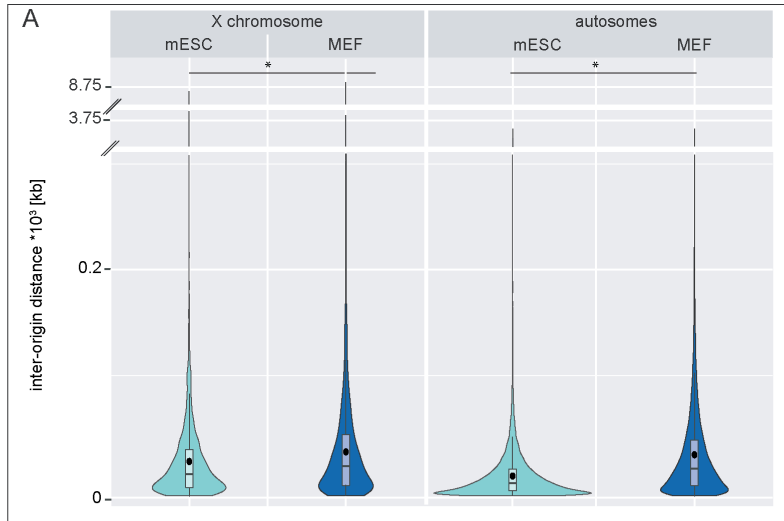


**Figure S19 – Y chromosome replication marks the end of S-phase independently of the cell potency state.** (A)

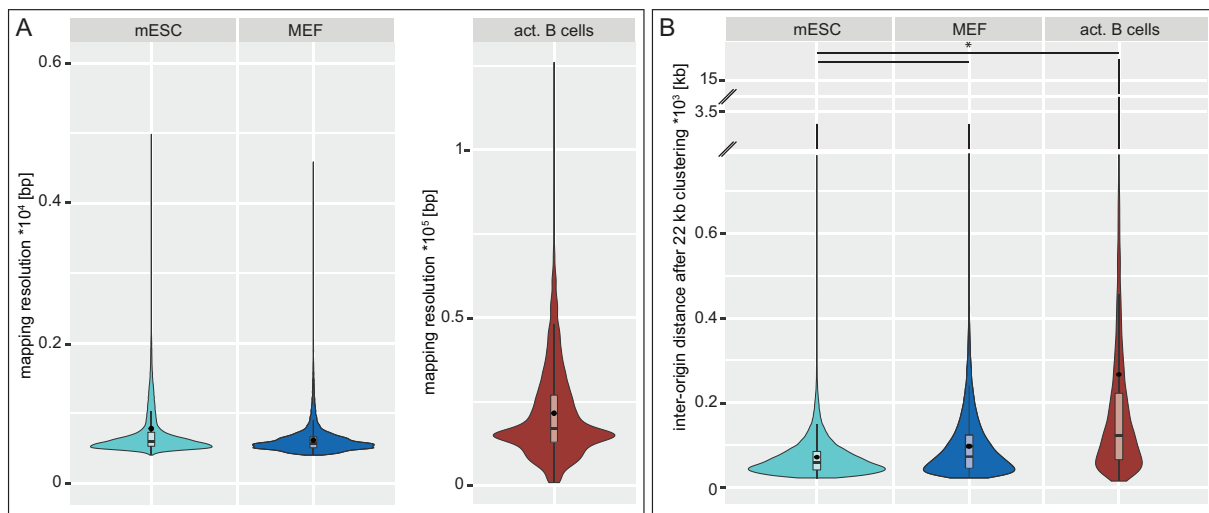
Replication timing analysis of the Y chromosome in mouse embryonic fibroblasts (MEF) via Repli-FISH. PCNA quantification within hybridization signals and representation are as in Figure 4. (B) Pulse chase experiment in male (un)differentiated mouse J1 ES cells and mouse embryonic fibroblasts (MEF W8). Representative spinning disk confocal images of transitions from stage III/late to stage Y and from stage Y to G2 are shown for all three cell lines. Arrowheads mark the Y chromosome. Boxplots are as in Figure 2 and Supplementary Figure S7. Scale bar = 5  $\mu\text{m}$ . Detailed statistics are summarized in Supplementary Table 10. \*  $P < 0.05$  (calculated against  $S_y$ ).



**Supplementary Figure S20 – Origin profiling in mouse embryonic stem cells, mouse embryonic fibroblasts and mouse activated B cells.** Graphical representation of origin profiling in mES, MEF and mouse activated B cells (act. B cells) in all replicates analyzed are shown. mES and MEF origins clustered at the distance of 22 kb (performed for equalization of origin mapping resolution in the SNS-seq and the OK-seq methods) as well as the middle point of clustered origins are indicated. The OK-seq profiles in the B cell line in 3 replicates are represented in the following rows as well as origins localization and their middle point. The replication profile scale is indicated in the upper left corner.



**Supplementary Figure S21 – Origin distribution on the chromosome X and autosomal chromosomes.** (A) The comparison of IOD calculated for the mES and MEF cell lines on autosomes and the chromosome X using the SNS-seq origin mapping data. (B) Schematic representation of origin distribution on the chromosome X and 2 parental copies of autosomal chromosomes. (C) Graphical representation of origin profiling in mES and MEF cells on a somatic chromosome (chromosome 12) and an arbitrary chosen region on the chromosome X. Two independent replicates of origin profiling and the localization of origins reproducibly identified in the two replicates are shown. The replication profile scale is indicated in the upper left corner. Detailed statistics are summarized in Supplementary Table 13. \*  $P < 0.05$ .



**Supplementary Figure S22 – Origin mapping resolution in genome-wide techniques and origin clustering.** (A) Origin mapping resolutions measured as an average origin peak width in SNS-seq and OK-seq methods are shown. (B) For comparison purposes the differences in methods resolution were equalized by merging the neighboring origins detected in SNS-seq in the distance of the average peak size in OK-seq analysis (22 kb). The IODs calculated for clustered origins in mES and MEF cell lines as well as origins identified in activated B cells (act. B cells) are shown. Detailed statistics are summarized in Supplementary Table 13. \*  $P < 0.05$ .

**Movies 1 & 2 - Time lapse movie of cell cycle progression in a mouse embryonic stem cell.** mES cells were transfected with mRFP-PCNA and MaSat-GFP (polydactyl zinc finger protein with specific binding to major satellite repeats) and time lapse confocal microscopy was performed 24 hours post-transfection. z-stacks were acquired at 30 minutes intervals. Chromocenters (green) and DNA replication (magenta) are shown as merged images and time intervals are shown on the upper left. Depicted mES cell progresses from S-phase stage I to mitosis (movie 1) and from S-phase stage I to G2 (movie 2). Scale bar = 5  $\mu$ m.

## REFERENCES

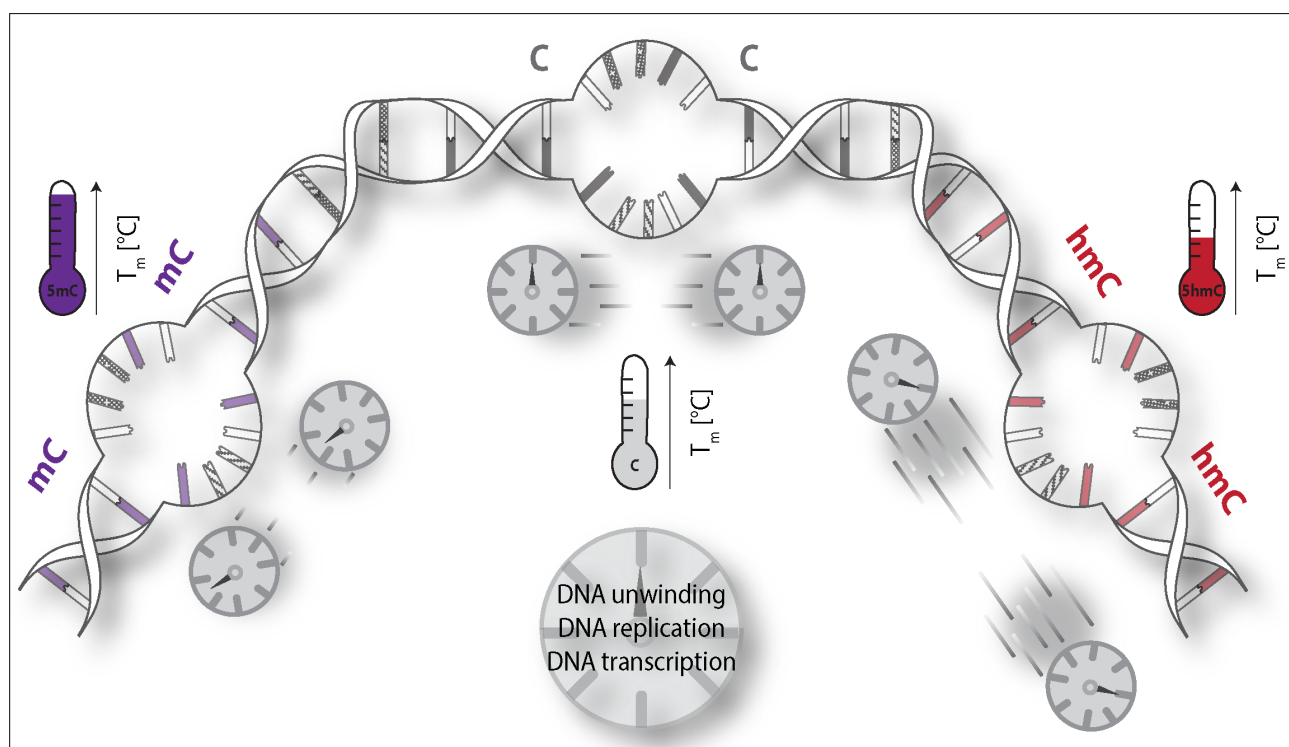
1. Li, E., Bestor, T.H. and Jaenisch, R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, **69**, 915-926.
2. Doetschman, T., Gregg, R.G., Maeda, N., Hooper, M.L., Melton, D.W., Thompson, S. and Smithies, O. (1987) Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature*, **330**, 576-578.
3. Yaffe, D. and Saxel, O. (1977) Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature*, **270**, 725-727.
4. Peters, A.H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schofer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A. *et al.* (2001) Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell*, **107**, 323-337.
5. Sporbert, A., Domaing, P., Leonhardt, H. and Cardoso, M.C. (2005) PCNA acts as a stationary loading platform for transiently interacting Okazaki fragment maturation proteins. *Nucleic acids research*, **33**, 3521-3528.
6. Lindhout, B.I., Fransz, P., Tessadori, F., Meckel, T., Hooykaas, P.J. and van der Zaal, B.J. (2007) Live cell imaging of repetitive DNA sequences via GFP-tagged polydactyl zinc finger proteins. *Nucleic acids research*, **35**, e107.
7. Casas-Delucchi, C.S., Becker, A., Bolius, J.J. and Cardoso, M.C. (2012) Targeted manipulation of heterochromatin rescues MeCP2 Rett mutants and re-establishes higher order chromatin organization. *Nucleic acids research*, **40**, e176.
8. Weber, P., Rausch, C., Scholl, A. and Cardoso, M.C. (2019) Repli-FISH (Fluorescence in Situ Hybridization): Application of 3D-(Immuno)-FISH for the study of DNA replication timing of genetic repeat elements. *OBM Genetics*, **3**.
9. Prorok, P., Artufel, M., Aze, A., Coulombe, P., Peiffer, I., Lacroix, L., Guedin, A., Mergny, J.L., Damaschke, J., Schepers, A. *et al.* (2019) Involvement of G-quadruplex regions in mammalian replication origin activity. *Nature communications*, **10**, 3274.
10. Almeida, R., Fernandez-Justel, J.M., Santa-Maria, C., Cadoret, J.C., Cano-Aroca, L., Lombrana, R., Herranz, G., Agresti, A. and Gomez, M. (2018) Chromatin conformation regulates the coordination between DNA replication and transcription. *Nature communications*, **9**, 1590.
11. Tubbs, A., Sridharan, S., van Wietmarschen, N., Maman, Y., Callen, E., Stanlie, A., Wu, W., Wu, X., Day, A., Wong, N. *et al.* (2018) Dual Roles of Poly(dA:dT) Tracts in Replication Initiation and Fork Collapse. *Cell*, **174**, 1127-1142 e1119.



## 5.4 CHAPTER IV: Cytosine Base Modifications Regulate DNA Duplex Stability and Metabolism

C. Rausch prepared all figures and tables and wrote the manuscript together with M. C. Cardoso. C. Rausch performed all experiments except for figures 1A, 4A, 7D, S3, S13A and S15A. Experiments from figures 1B-C, S8A, 8C and S8E-G were performed in collaboration with P. Zhang. C. Rausch analyzed all data except from figures 4A, S13A and S15A and from figures 1B-C, S8A, 8C and S8E-G analyzed together with P. Zhang.

### Graphical Abstract





# Cytosine base modifications regulate DNA duplex stability and metabolism

Cathia Rausch<sup>1</sup>, Peng Zhang<sup>1,2,\*</sup>, Corella S. Casas-Delucchi<sup>3,†</sup>, Julia L. Daiß<sup>4,†</sup>, Christoph Engel<sup>4</sup>, Gideon Coster<sup>3</sup>, Florian D. Hastert<sup>1</sup>, Patrick Weber<sup>1</sup> and M. Cristina Cardoso<sup>1,\*</sup>

<sup>1</sup>Cell Biology and Epigenetics, Department of Biology, Technical University of Darmstadt, 64287 Darmstadt, Germany, <sup>2</sup>Center for Tissue Engineering and Stem Cell Research, Guizhou Medical University, Guiyang, Guizhou 550004, China, <sup>3</sup>Chester Beatty Laboratories, The Institute of Cancer Research, London SW3 6JB, UK and <sup>4</sup>Regensburg Center for Biochemistry, University of Regensburg, 93053 Regensburg, Germany

Received July 13, 2020; Revised May 26, 2021; Editorial Decision May 27, 2021; Accepted June 03, 2021

## ABSTRACT

DNA base modifications diversify the genome and are essential players in development. Yet, their influence on DNA physical properties and the ensuing effects on genome metabolism are poorly understood. Here, we focus on the interplay of cytosine modifications and DNA processes. We show by a combination of *in vitro* reactions with well-defined protein compositions and conditions, and *in vivo* experiments within the complex networks of the cell that cytosine methylation stabilizes the DNA helix, increasing its melting temperature and reducing DNA helicase and RNA/DNA polymerase speed. Oxidation of methylated cytosine, however, reverts the duplex stabilizing and genome metabolic effects to the level of unmodified cytosine. We detect this effect with DNA replication and transcription proteins originating from different species, ranging from prokaryotic and viral to the eukaryotic yeast and mammalian proteins. Accordingly, lack of cytosine methylation increases replication fork speed by enhancing DNA helicase unwinding speed in cells. We further validate that this cannot simply be explained by altered global DNA decondensation, changes in histone marks or chromatin structure and accessibility. We propose that the variegated deposition of cytosine modifications along the genome regulates DNA helix stability, thereby providing an elementary mechanism for local fine-tuning of DNA metabolism.

## INTRODUCTION

It has long been known that besides the four canonical DNA bases, adenine (A), thymine (T), cytosine (C) and guanine (G), chemically modified variants are found within the genome of almost all phyla (1). The predominant DNA modification within the mammalian genome is cytosine methylation, consisting of the addition of a methyl group to the C5 of the cytosine ring [5-methylcytosine (5mC) (2)], and catalyzed by DNA methyltransferases (Dnmts) (3,4). In mammals, cytosine methylation is involved in numerous cellular processes, such as X-chromosome inactivation and imprinting, silencing of gene expression, control of cell development and differentiation (5–7). The impact of DNA methylation on gene expression and ultimately cellular differentiation stresses the need for a tight regulation of d5mC levels. Oxidation of d5mC to 5-hydroxymethylcytosine (d5hmC) (8) and the identification of the responsible enzymes, the ten-eleven Translocation protein family (Tet1, Tet2 and Tet3) (9), established a long sought active DNA demethylation pathway dependent on thymine-DNA glycosylase (TDG) and base excision repair (BER) (10–12). Tet proteins catalyze the oxidation of 5mC to 5hmC in a 2-oxoglutarate-, iron(II)- and oxygen-dependent manner and were shown to further oxidize 5hmC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (13) in DNA. Subsequent studies could identify factors that specifically recognize and bind modified cytosine bases (epigenetic readers) (14), thereby regulating chromatin structure, DNA accessibility and gene expression (15,16).

Generally, cytosine base modifications are thought to not influence the global canonical B-DNA structure, although local changes around the modification site can arise [reviewed in (17)]. Recent reports, however, show that chem-

\*To whom correspondence should be addressed. Tel: +49 6151 16 21882; Fax: +49 6151 16 21880; Email: cardoso@bio.tu-darmstadt.de  
 Correspondence may also be addressed to Peng Zhang. Tel: +86 85180988028; Fax: +86 85180988028; Email: peng12zhang@outlook.com

†The authors wish it to be known that, in their opinion, they should be regarded as contributing equally.

Present address: Florian D. Hastert, Paul-Ehrlich-Institut, Bundesinstitut für Impfstoffe und biomedizinische Arzneimittel, 63225 Langen, Germany.

© The Author(s) 2021. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

ically modified bases can directly affect primary DNA double helix structure and conformation *in vitro* (18), as well as flexibility (19) and thermostability (20). Yet, the influence of base modifications on DNA structure and stability is still highly debated (21,22). These chemical modifications of DNA bases do not alter the Watson–Crick base pairing per se; they, however, add a chemical functionality within the DNA double helix, changing the hydrophobicity of the major groove, introducing steric hindrance and changing base pair stacking, structure and mechanical properties of the DNA double helix (18,22–27). Interestingly, the size and especially the hydrophobicity of the modifications influence fluctuations and rotary motions within the base pairs and are essential determinants of the solvation state of the DNA stretch and consequently of its rigidity and stability (28).

Since an alteration in DNA stability and shape would be predicted to have major consequences for global genome processes, we investigated the effects of cytosine modifications on DNA double helix stability and DNA-related enzymatic processes *in vitro* and *in vivo*. We therefore analyzed, on the one hand, DNA melting temperature *in vitro* and, on the other hand, manipulated genomic DNA base modification levels *in vivo*. To this end, we altered expression levels of modification writer proteins (Dnmts and Tets) and transfected cells with modified nucleotides. Overall, methylated cytosines resulted in higher melting temperatures, slower helicase-mediated DNA unwinding and decreased rates of DNA replication and transcription. In contrast, its oxidation product d5hmC reverted this negative effect and led to lower melting temperatures, and overall higher DNA unwinding, and replication and transcription rates. We additionally show that the loss of DNA modifications in mouse embryonic stem cells altered S-phase progression, due to faster replication fork speed in the absence of d5mC. These results imply that, in addition to previously described effects of DNA modifications on chromatin structure, their effect on the stability of the DNA duplex per se creates an additional way by which DNA base modifications can influence genomic processes, thereby (locally) fine-tuning DNA accessibility and enzymatic activity.

## MATERIALS AND METHODS

### Plasmids

Mammalian expression constructs encoding a mcherry-tagged fusion of the catalytic domain of mouse Tet1 (Tet1CD) and the mcherry tag alone were previously described (15). GFP-tagged human RPA34 (29) and RFP-tagged human proliferating cell nuclear antigen (PCNA) (30) or miRFP670-tagged human PCNA were used to study replication protein A (RPA) accumulation and to identify S-phase stages, respectively. The expression vector containing a monomeric near-infrared (miRFP670) PCNA fusion was generated by amplifying miRFP670 from the pmiRFP670-N1 [Addgene plasmid # 79887 and (31)] vector with primers containing AgeI and BsrGI sites (Supplementary Table S1). The fragment was inserted into AgeI/BsrGI-digested mRFP-hPCNA (30). The resulting plasmid was verified by restriction enzyme and DNA sequencing analyses. The correct subcellular localization was validated by microscopi-

cal analysis and the expected protein was detected by western blotting analysis upon expression into mammalian cells (Supplementary Figure S1). Supplementary Figure S2 and Supplementary Table S2 summarize the characteristics of fusion proteins and expression constructs.

To generate the plasmid containing the yeast ARS306 origin of replication (pGC504), two oligonucleotides (GC497 and GC498, Supplementary Table S1) were annealed and ligated into pZN3 (32) digested with BamHI and PstI.

### Cells, culture, transfection, inhibitor treatment and nucleoside/nucleotide incorporation

All cells used were tested for mycoplasma and deemed free of contamination. C2C12 mouse myoblast cells (33), mouse embryonic fibroblasts (MEFs) W8 (34) and PM [p53<sup>-/-</sup> and Dnmt1<sup>-/-</sup>, (35)] and HEK 293 human embryonic kidney (36) cells were grown in Dulbecco's modified Eagle Medium (high glucose, DMEM, Sigma-Aldrich, St Louis, MO, USA) supplemented with 20% (C2C12), 15% (MEF PM) or 10% (MEF W8 and HEK 293) fetal bovine serum, 1 × glutamine (cat# G7513, Sigma-Aldrich, St Louis, MO, USA) and 1 μM gentamicin (cat# G1397, Sigma-Aldrich, St Louis, MO, USA) in a humidified atmosphere with 5% CO<sub>2</sub> at 37°C. J1 wt and TKO (Dnmt1<sup>-/-</sup>, Dnmt3a<sup>-/-</sup> and Dnmt3b<sup>-/-</sup>) mouse embryonic stem cells were cultured under feeder-free conditions on gelatin-coated culture dishes (0.2% gelatin in ddH<sub>2</sub>O) in DMEM (high glucose) supplemented with 16% fetal bovine serum, 1 × nonessential amino acids (cat# M7145, Sigma-Aldrich, St Louis, MO, USA), 1 × penicillin/streptomycin (cat# P4333, Sigma-Aldrich, St Louis, MO, USA), 2 mM L-glutamine, 0.1 mM β-mercaptoethanol (cat# 4227, Carl Roth, Karlsruhe, Germany), 1000 U/ml LIF (cat# ESG1107, Merck, Kenilworth, NJ, USA), 1 μM PD0325901 and 3 μM CHIR99021 (Axon Medchem BV, Reston, VA, USA). Supplementary Table S3 summarizes the cell line characteristics.

Transfections of C2C12 and MEF mouse cells with expression plasmids were performed using AMAXA nucleofection (AMAXA Nucleofector II, Lonza, Basel, Switzerland) as described previously (37). Modified nucleotides (dCTP, d5mCTP and d5hmCTP) were added at a final concentration of 0.2 mM and Cy3-dUTP at a final concentration of 0.02 mM to the transfection mix (cells resuspended in 5 mM KCl, 15 mM MgCl<sub>2</sub>, 120 mM Na<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub>, pH 7.2, and 50 mM mannitol). Cells were analyzed 24 h after transfection. Nucleotide characteristics are summarized in Supplementary Table S4.

Cells used for immunofluorescence were grown on glass coverslips. For time-lapse microscopy, cells were plated in μ-Slide eight-well slides (Ibidi GmbH, Planegg/Martinsried, Germany).

Trichostatin A (TSA, cat# T8552, Sigma-Aldrich, St Louis, MO, USA) treatment was performed with a final concentration of 20 nM for 72 h changing with fresh TSA containing medium every day, as described in (38).

EdU (5-ethynyl-2'-deoxyuridine) and EU (5-ethynyl-uridine) labeling and the subsequent detection of nascent DNA/RNA are described below.

### Growth curve analysis, doubling time, cell cycle distribution and progression

For growth curve analysis,  $2 \times 10^5$  J1 wt and TKO cells were seeded as technical quadruplicates at day 0, and cell numbers were counted with a Neubauer hemocytometer for four consecutive days. Population doubling times were derived with  $\log_2(n_x/n_0)/t$  [ $n_x$ : cell number at day  $x$ ,  $n_0$ : cell number at day 0,  $t$ : hours after seeding]. To determine the percentage of cells in every cell cycle and S-phase substage, asynchronously growing J1 wt and TKO cell cultures were pulse labeled with 10  $\mu$ M EdU for 12 min, fixed and EdU was detected as described below. Cells were manually grouped into the S-phase substages (stages I to Y, (39)), non S-phase or mitosis, and percentages were calculated. S-phase (substage) duration was derived by multiplying the doubling time with the percentage of cells in the respective phase. To compare the spatio-temporal progression through S-phase, J1 wt and TKO cells were pulse labeled for 12 min with 10  $\mu$ M EdU, washed twice and chased for 2 h with 50  $\mu$ M thymidine (cat# T-1895, Sigma-Aldrich, St Louis, MO, USA). Fixation, permeabilization and MeOH treatment were done as described for immunofluorescence. EdU was detected prior to PCNA detection with mouse anti-PCNA and donkey anti-mouse IgG Cy3. All antibody characteristics are summarized in Supplementary Table S5.

### Immunofluorescence

C2C12, MEF W8 and PM and J1 wt and TKO cells were fixed in 3.7% formaldehyde for 10 min and permeabilized for 20 min in 0.5% Triton X-100. For detection of 5mC, 5hmC, BrdU, H3 acetylation and histone modifications, cells were incubated for 5 min in ice-cold methanol. RNaseA treatment (10  $\mu$ g/ml in 1 $\times$  PBS [phosphate-buffered saline]) was performed for 30 min at 37°C and was followed by blocking in 0.2% fish skin gelatin (Sigma-Aldrich, St Louis, MO, USA) in 1 $\times$  PBS. 5mC, 5hmC and BrdU were detected using mouse anti-5mC antibody, rabbit anti-5hmC antibody and rabbit anti-BrdU in conjunction with 25 U/ml DNaseI (Sigma-Aldrich, St Louis, MO, USA) for 60 min at 37°C in digestion/staining buffer (2% BSA, 30 mM Tris/HCl pH 8, 0.33 mM MgCl<sub>2</sub>, 0.5 mM  $\beta$ -mercaptoethanol). The following primary anti-histone antibodies were diluted in 1% BSA/PBS (bovine serum albumin in 1 $\times$  phosphate-buffered saline) and incubated for 1 h at room temperature: rabbit anti-H3 acetyl, mouse anti-H3K9ac, rabbit anti-H3K18ac, rabbit anti-H3K56ac, rabbit anti-H4K8ac, rabbit anti-H4K16ac, rabbit anti-H3K9m3 and rabbit anti-H4K20m3. To stop DNaseI digestion, cells were washed with PBS containing 1 mM EDTA and 0.02% Tween-20. Following incubation with the secondary antibodies goat anti-rabbit IgG Chromeo 488, donkey anti-rabbit IgG Cy3, donkey anti-rabbit IgG Cy5, goat anti-mouse IgG Chromeo 488 or donkey anti-mouse IgG Cy3 for 60 min at room temperature, cells were washed with PBS containing 0.02% Tween-20. DNA was counterstained with 4,6-diamidino-2-phenylindole (DAPI, 1  $\mu$ g/ml, cat# D9542, Sigma-Aldrich, St Louis, MO, USA) for 10 min, and cells were mounted in Mowiol 4-88 (cat# 81381, Sigma-Aldrich, St Louis, MO, USA) with 2.5% DABCO antifade (1,4-diazabicyclo[2.2.2]octan, cat#

D27802, Sigma-Aldrich, St Louis, MO, USA). All antibody characteristics are summarized in Supplementary Table S5.

### Fluorescence *in situ* hybridization (FISH)

For the detection of major satellite DNA repeats, probes were amplified and labeled from mouse myoblast C2C12 genomic DNA by PCR (primer sequences in Supplementary Table S1) using self-made biotin dUTPs and Taq polymerase as described in (40). Probes were ethanol precipitated in the presence of 0.13 M sodium acetate and 50  $\mu$ g/ml fish sperm DNA, resuspended in hybridization solution (70% formamide, 2 $\times$  SSC [saline sodium citrate] and 10% dextran sulfate, pH 7), denatured at 80°C for 5 min and immediately put on ice. Cells were fixed for 10 min in 3.7% formaldehyde in 1 $\times$  PBS, washed and subsequently permeabilized two times with 0.5% Triton X-100 for 15 min, equilibrated for 3 min in 2 $\times$  SSC followed by 30 min in 50% formamide/2 $\times$  SSC. After combining probes and cells in a metal chamber, denaturation was performed at the indicated temperatures with continuous temperature monitoring for 5 min in a water bath followed by overnight annealing at 37°C. Post-hybridization washing was performed with 2 $\times$  SSC; blocking and antibody dilution were done with 1% BSA/4 $\times$  SSC. To detect mcherry positive cells after FISH denaturation, rat anti-RFP antibody and donkey anti-rat IgG Cy3 (Supplementary Table S5) were used. Biotin was detected with Alexa488 conjugated streptavidin (1:800, cat# S11223, Invitrogen, PA49RF Paisley, UK) for 1 h at room temperature. DNA was counterstained and cells were mounted as described for immunofluorescence.

### EdU/EU labeling and visualization of nascent DNA/RNA

To analyze DNA replication, cells were labeled with 10  $\mu$ M EdU (Supplementary Table S4) for 10, 20 or 30 min. To analyze DNA transcription, cells were labeled with 1 mM EU (Supplementary Table S4) for 25, 45, 60 or 120 min. Cells were fixed 10 min in 3.7% formaldehyde. EdU and EU were detected using the Click-IT assay (Thermo Fisher Scientific, Waltham, MA, USA) with 6-FAM azide (1:1000) according to the manufacturer's instructions. DNA was counterstained and cells were mounted as described for immunofluorescence.

### DNA preparation

All oligonucleotide, primer, plasmid and nucleotide characteristics are summarized in Supplementary Tables S1, S2 and S4, respectively. PCR products used in HRM, dot blot, DNA/RNA polymerization and exonuclease assays were purified by gel electrophoresis followed by silica column purification using the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany). Template purification for soluble *in vitro* replication assays with yeast proteins was performed using phenol/chloroform (described below).

Genomic DNA (gDNA) for FISH probe generation was prepared as previously described (40).

*DNA oligonucleotide annealing.* DNA oligonucleotides (20-mer) for high-resolution melting temperature

4 *Nucleic Acids Research*, 2021

(HRM) analysis were purchased from IBA Lifescience (Goettingen, Germany) and IDT (Integrated DNA Technologies, Skokie, IL, USA). Oligonucleotide sequences are summarized in Supplementary Table S1. To prepare double-stranded DNA oligonucleotides, upper and lower oligos were mixed at equal molar amounts in NEBuffer 2 (50 mM NaCl, 10 mM Tris/HCl, 10 mM MgCl<sub>2</sub>, 1 mM dithiothreitol (DTT); NEB, Ipswich, MA, USA), denatured at 95°C for 2 min and annealed by slowly cooling down to room temperature.

*PCR fragments for HRM, DNA polymerization and exonuclease assays.* PCR fragments for HRM analysis were amplified using Q5 polymerase (cat# M0491, NEB, Ipswich, MA, USA) and dCTP, d5mCTP, d5hmCTP, d5fCTP or d5caCTP from a plasmid containing the MINX sequence ((41), Supplementary Table S2) with MINX specific primers or from human HEK gDNA with LINE1 promoter primers. Cycles were set to 98°C for 2 min, 40× (98°C for 15 s, 62.5°C for 30 s, 72°C for 40 s), 72°C for 2 min. To amplify d5fC- and d5caC-containing products, the following cycle conditions were used: 98°C for 1 min, 45× (98°C for 10 s, 65.2°C for 30 s, 72°C for 40 s), 72°C for 2 min.

PCR fragments for DNA polymerization and exonuclease assays were amplified as described above and using dCTP, d5mCTP or d5hmCTP and Q5 polymerase.

*DNA templates for in vitro DNA replication assays.* DNA templates (3 kb) for soluble DNA replication assays with eukaryotic replisome components were generated using Q5 polymerase and dCTP, d5mCTP or d5hmCTP from a plasmid containing the yeast ARS306 origin (pGC504 ARS306) with specific pGC504 primers. Primers were designed in a way to generate amplicons with centrally arranged ARS306. Additionally, to generate amplicons with different levels of cytosine modifications, reactions containing 100% (100% C<sub>modif</sub> and 0% C), 50% (50% C<sub>modif</sub> and 50% C) or 12.5% (12.5% C<sub>modif</sub> and 87.5% C) d5mCTP or d5hmCTP were set up. Cycles were set to 98°C for 2 min, 40× (98°C for 20 s, 68.4°C for 30 s, 72°C for 100 s), 72°C for 10 min. All reactions were checked on agarose gels (0.3 µg/ml ethidium bromide), DNA was extracted from the reactions using phenol/chloroform/isoamyl alcohol (cat# 3215, Carl Roth, Karlsruhe, Germany and cat# A1935, AppliChem, Darmstadt, Germany) and ethanol precipitated in the presence of 0.3 M sodium acetate.

*DNA templates for in vitro DNA transcription assays.* PCR fragments for T7 RNA polymerization reactions were amplified using dCTP, d5mCTP or d5hmCTP, MINX plasmid as template and Q5 polymerase. For the generation of the T7 transcription assays, the forward primer binds the T7 promoter region, therefore, the resulting T7 promoter region amplicon contained unmodified cytosines on the forward strand, allowing T7 RNA polymerase-mediated transcription.

PCR fragments for tailed template elongation with eukaryotic RNA polymerase were amplified using dCTP or 100% d5mCTP or d5hmCTP, MINX plasmid as template and Q5 polymerase. Nb.BsmI nicking site containing for-

ward primer and the corresponding reverse primer are listed in Supplementary Table S1.

**PCR efficiency and yield assays**

PCR efficiency and yield assays contained MINX or LINE1 primers and MINX plasmid or HEK gDNA as template, respectively, dATP, dTTP, dGTP and dCTP, d5mCTP or d5hmCTP and the indicated polymerases. DNA denaturing was performed at the indicated temperatures. Agarose gels were stained with 0.3 µg/ml ethidium bromide (EtBr, cat# 2218, Carl Roth, Karlsruhe, Germany) in 1× TAE (tris acetic acid buffer, 40 mM Tris, 1 mM EDTA, 20 mM glacial acetic acid) for 10 min, destained for 5 min in ddH<sub>2</sub>O and imaged using a AI600 imager (Supplementary Table S6).

**Dot and western blot analysis**

Dot blot analysis of modified PCR fragments was performed as described (15). For PCR fragment analysis, 20, 50, 75 or 150 ng of DNA was spotted on a nitrocellulose membrane. To confirm (equal) DNA loading, the membrane was stained for 2 min in 0.2% methylene blue (cat# M9140, Sigma-Aldrich, St Louis, MO, USA) in 0.3 M sodium acetate, pH 5.2, followed by destaining in ddH<sub>2</sub>O. On a separate membrane, the modifications were detected with primary mouse anti-5mC, rabbit anti-5hmC, rabbit anti-5fC and rabbit anti-5caC antibodies, respectively, and secondary sheep anti-mouse IgG HRP (horseradish peroxidase) conjugated or goat anti-rabbit IgG HRP conjugated antibodies.

To analyze the miRFP-hPCNA (pc3385) fusion protein, (transfected) HEK293 cells were lysed in 4× SDS loading buffer (200 mM Tris/HCl pH 6.8, 400 mM DTT, 8% SDS, 0.4% bromophenol blue and 40% glycerol), and whole protein lysates were analyzed by SDS-PAGE and blotted on a nitrocellulose membrane (GE Healthcare, Chicago, IL, USA). PCNA was detected using a mouse anti-PCNA primary antibody and a sheep anti-mouse IgG HRP conjugated secondary antibody. Visualization of immunoreactive bands was achieved by ECL plus Western Blot detection reagent (GE Healthcare, Munich, Germany).

Membranes were imaged with the CCD camera of an AI600 imager (Supplementary Table S6) and antibody characteristics are summarized in Supplementary Table S5.

**In vitro DNA and RNA polymerization and DNA exonuclease assay with prokaryotic and phage enzymes**

*DNA polymerization assays using prokaryotic enzymes.* DNA polymerization reactions contained LINE1 PCR amplicons containing dC, d5mC or d5hmC, respectively, as DNA template, LINE1 forward and reverse primers (Supplementary Table S1) and dATP, dTTP, dGTP and dCTP (Supplementary Table S4), respectively, and were incubated for 3 min at 99°C to denature the dsDNA template. Primer annealing was done for 1 min at 85°C. Reactions were cooled down to 37°C, and 0 U, 0.0016 U, 0.008 U, 0.04 U, 0.2 U or 1 U of Klenow polymerase (cat# M0210, NEB, Ipswich, MA, USA) was added and incubated for 20 min to allow DNA polymerization. Products were analyzed on

a 2% agarose gel; band intensities were measured (readout for DNA polymerization rate) and plotted.

**DNA exonuclease assays.** DNA exonuclease reactions contained MINX PCR amplicons with dC, d5mC or d5hmC, respectively, as DNA template and 0 U, 0.03125 U, 0.0625 U, 0.125 U, 0.25 U or 0.5 U of T4 DNA polymerase (cat# M0203, NEB, Ipswich, MA, USA) was added and incubated for 7 min at room temperature. Exonuclease activity was analyzed on a 2% agarose gel, and band intensities were measured (readout for DNA exonuclease rate) and plotted.

All agarose gels were imaged using an AI600 imager (Supplementary Table S6), integrated band intensities were measured using ImageJ (<https://imagej.nih.gov/ij/>), and background was subtracted and plotted as a ratio to the control (DNA without enzyme).

**RNA polymerization assays using phage enzymes.** To generate single-stranded DNA (ssDNA) templates for *in vitro* transcription reactions (IVT), 5' phosphorylated (forward strand, Supplementary Table S1) PCR amplicons were digested with  $\lambda$  exonuclease (cat# EN0561, Thermo Fisher, Waltham, MA, USA). The ssDNA generated was purified by ethanol precipitation as described in (40) and in the presence of 80  $\mu$ g glycogen (cat# 10901393001, Sigma Aldrich, St Louis, MO, USA). To obtain a double-stranded (dsDNA) T7 promoter region, a forward oligonucleotide containing the T7 promoter (Supplementary Table S1) was annealed to the ssDNA fragments. Therefore, equimolar amounts of ssDNA and oligonucleotide were mixed, heated for 3 min at 95°C and slowly cooled down to room temperature to allow annealing.

For RNA polymerization assays via T7 run-off IVT, PCR amplicons containing dC, d5mC or d5hmC respectively (dsDNA or ssDNA) DNA template, ATP, UTP, GTP and CTP (Supplementary Table S4), T7 RNA polymerase (cat# M0251, NEB, Ipswich, MA, USA) and 0.5 $\times$  SYBR Green II (cat# S-7564, Thermo Fisher, Waltham, MA, USA) were used. Cycling reactions were performed in a StepOnePlus™ Real-Time PCR machine (Applied Biosystems), SYBR Green II intensities were measured every 30 s for 450 min and plotted using RStudio (version 1.1.447).

#### Soluble *in vitro* DNA replication assays with purified eukaryotic replisome components

DNA replication assays with yeast replication proteins were performed at 30°C and as described in (42) with the following changes: MCM loading and phosphorylation were performed (5  $\mu$ l per lane) in a buffer containing 25 mM HEPES-KOH (pH 7.6), 100 mM potassium glutamate, 10 mM magnesium acetate, 100  $\mu$ g/ml BSA, 2 mM DTT, 0.01% NP-40-S, 5 mM ATP, 45 nM Cdc6, 20 nM origin recognition complex (ORC), 75 nM Cdt1-Mcm2-7, 50 nM DBF4 dependent kinase (DDK) and 5 nM linear DNA template (described above). After 20 min incubation, the reaction volume was increased 2-fold by the addition of pre-equilibrated buffer to give a final replication reaction buffer of 25 mM HEPES-KOH (pH 7.6), 100 mM potassium glutamate, 10 mM magnesium acetate, 100  $\mu$ g/ml BSA, 1 mM

DTT, 0.01% NP40-S, 3 mM ATP, 22.5 nM Cdc6, 10 nM ORC, 37.5 nM Cdt1-Mcm2-7, 25 nM DDK, 2.5 nM linear DNA template, 200  $\mu$ M CTP, GTP and UTP, 30  $\mu$ M dCTP, dGTP, dATP and dTTP and 33 nM  $\alpha$ -<sup>33</sup>P-dATP. Replication was initiated by adding a master mix of proteins to achieve final concentrations of 20 nM Sld3/7, 20 nM Sld2, 30 nM Dpb11, 100 nM GINS, 40 nM Cdc45, 20 nM Pol  $\epsilon$ , 10 nM Mcm10, 40 nM RPA, 10 nM S-CDK, 40 nM Pol  $\alpha$ , 25 nM Csm3/Tof1, 25 nM Mrc1, 30 nM RFC, 40 nM PCNA and 5 nM Pol  $\delta$ . Nucleotide characteristics are summarized in Supplementary Table S4. The replication reaction was stopped after 15 min by the addition of 100 mM EDTA, separated on 0.8% alkaline agarose gels as described in (43) and signals detected using a Typhoon FLA-9500 (Supplementary Table S6).

#### *In vitro* tailed template transcription (RNA polymerization) assays with purified eukaryotic DNA-dependent RNA polymerase I

Tailed template PCR products (described above) were purified using the QIAquick PCR purification kit (cat# 28104, Qiagen, Hilden, Germany). Purified templates were cut with the nicking endonuclease Nb.BsmI (cat# R0706S, NEB, Ipswich, MA, USA) at 65°C for 2 h. The reaction was inactivated at 80°C for 20 min. After 10 min, a 10-fold molar excess of competitor oligo (Supplementary Table S1) was added and then cooled down slowly to anneal with the cleaved-off single-stranded fragment to obtain a 3'-overhang on the template. The DNA was ethanol-precipitated, dried and resuspended in water to reach a concentration of 125 nM.

*Saccharomyces cerevisiae* DNA-dependent RNA polymerase (Pol) I was purified as described (44,45). The polymerization assay is based on previously published protocols (46). Per transcription reaction 0.25 pmol Pol I, 0.125 pmol template (described above), 50 pmol GpC dinucleotide and HEPES (pH 7.8; end concentration 50 mM) were mixed. Five transcription reactions (final volume 25  $\mu$ l) were treated as one and started by adding the equal amount of 2 $\times$  transcription buffer (40 mM HEPES, pH 7.8, 10 mM MgCl<sub>2</sub>, 5 mM EGTA, 0.05 mM EDTA, 2.5 mM DTT, 0.2 mM ATP, 0.2 mM UTP, 0.2 mM GTP, 0.01 mM CTP supplemented with 5  $\mu$ Ci of  $\alpha$ -<sup>32</sup>P-CTP per transcription reaction (0.0025 mM)). After 30, 60, 90, 120 and 600 s, 10  $\mu$ l were removed and the reaction was stopped by adding the equal amount of 2 $\times$  RNA loading dye (8 M urea, 2 $\times$  TBE, 0.03% bromophenol blue, 0.03% xylene cyanol, 10 mM EDTA, 10 mM EGTA). The samples were heated to 95°C for 5 min, briefly centrifuged and applied to urea-PAGE (6% polyacrylamide gel containing 7 M urea in 1 $\times$  TBE). The gel was dried for 30 min at 80°C in a vacuum dryer. Radiolabeled transcripts were detected by phosphor imaging (Supplementary Table S6). Nucleotide characteristics are summarized in Supplementary Table S4.

#### High-resolution melting analysis (HRM)

To determine the melting temperature of 20-mer oligonucleotide containing single cytosine modifications and fully modified (except primer sites) PCR fragments, Platinum®

SYBR® Green qPCR SuperMix-UDG w/ROX (cat# 11744100, Thermo Fisher, Waltham, MA, USA; UDG: uracil DNA glycosylase, ROX: 6-carboxyl-X-rhodamine) was diluted 1:1 with ddH<sub>2</sub>O, UDG was activated at 50°C for 2 min and added to 5 pmol of DNA samples. HRM temperature analysis was performed using a StepOnePlus™ Real-Time PCR machine (Applied Biosystems). Cycling conditions were set to 90°C for 30 s, then temperature was decreased to 50°C with a decreasing rate of 2%, followed by a temperature increase to 90°C with 0.1°C steps. Data normalization and plotting was performed using RStudio (version 1.1.447) and as described previously (47).

#### Micrococcal nuclease (MNase) chromatin accessibility analysis

A total of  $1 \times 10^7$  cells were harvested and cell pellets were resuspended in hypotonic lysis buffer (10 mM Tris, pH 7.5, 10 mM NaCl, 3 mM MgCl<sub>2</sub> and 0.5% NP-40) for 5 min on ice. Cell nuclei were pelleted (5 min, 120 × g), washed once in digestion buffer (50 mM Tris pH 8, 5 mM CaCl<sub>2</sub>) and resuspended in digestion buffer. MNase (cat# EN0181, Thermo Fisher Scientific, Waltham, MA, USA) was diluted in MNase storage buffer (20 mM Tris pH 7.6, 50 mM NaCl, 50% glycerol), and 0 U, 0.1 U, 0.5 U, 0.75 U, 1 U and 3 U were added to the nuclei in a final volume of 100 μl. Reactions were incubated for 5 min at 37°C, stopped by the addition of stopping buffer (20 mM EDTA, 0.4% SDS, 0.5 mg/ml proteinase K) and incubated at 50°C overnight. DNA was recovered by adding 0.1 volumes of 3 M NaCl and 2 volumes of 100% ethanol and storage at -20° for 1 h. After centrifugation (15 min, 15 000 × g) and washing with 70% ethanol, the DNA pellet was air-dried and resuspended in 50 μl ddH<sub>2</sub>O. Equal amounts of DNA were loaded on a 1.5% agarose gel and run for 45 min at 100 V. Gels were stained with EtBr in 1× TAE as described above and imaged using a AI600 imager (Supplementary Table S6). Integrated intensities of monomeric DNA and total DNA in each lane were measured with ImageJ and plotted as a DNA<sub>monomer</sub>/DNA<sub>total</sub> ratio.

#### Replication labeling and molecular combing

Molecular combing experiments were performed using the FiberPrep®-DNA Extraction Kit (cat# EXTR-001, Genomic Vision, Bagnex, France) and as described by the manufacturer. In brief, wild-type J1 and knockout TKO mouse embryonic stem cells were first labeled for 15 min with 10 μM 5-iodo-2'-deoxyuridine (IdU, Supplementary Table S4), washed twice with pre-warmed PBS, labeled for 15 min with 100 μM 5-chloro-2'-deoxyuridine (CldU, Supplementary Table S4) and chased for 1 h with 50 μM thymidine (Supplementary Table S4). Cells were harvested by trypsinization and embedded in low melting point agarose (NuSieveGTG agarose, Biozym Scientific GmbH, Hessisch Oldendorf, Germany), genomic DNA was isolated by proteinase K digestion at a final concentration of 0.2 mg/ml, and single DNA molecules were stretched on vinylsilane-coated glass coverslips (cat# COV-002-RUO, Genomic Vision, Bagnex, France), using the

FiberComb®-Molecular Combing System (cat# MCS-001, Genomic Vision, Bagnex, France) as described by the manufacturer.

Incorporated IdU and CldU nucleotides were immunofluorescently detected using mouse anti-BrdU/IdU and goat anti-mouse IgG Chromeo 546 before detecting CldU with rat anti-BrdU and donkey anti-rat IgG AlexaFluor 488. ssDNA was detected using mouse anti-ssDNA and goat anti-mouse IgG2a AlexaFluor 647. Samples were mounted with Vectashield antifade mounting medium (Vectorlabs, Burlingame, CA, USA). Antibody characteristics are summarized in Supplementary Table S5.

#### Time-lapse microscopy

All characteristics of the microscopy systems, including lasers, filters and objectives used, are summarized in Supplementary Table S6.

To analyze RPA accumulation as a proxy for DNA helicase activity in cells, confocal z-stacks of live cells were acquired using the Ultra-View VoX spinning disk microscopy system. Time-lapse microscopy was carried out in a closed live-cell microscopy chamber at 37°C with 5% CO<sub>2</sub> and 60% humidity. Somatic cells were grouped into early, mid and late S-phase cells according to their RPA and PCNA replication pattern (Supplementary Figure S1 and (48)). Somatic early S-phase cells are characterized by a homogeneous distribution of replication foci throughout the nucleus while in mid S-phase cells, replication signals are predominantly found around the nuclear and nucleolar border. In late S-phase cells, replication signals decrease in number while increasing in size and are co-localizing with highly condensed constitutive heterochromatin clusters (chromocenters). To analyze RPA accumulation in embryonic stem cells, only S-phase cells replicating chromocenters (stage II (39), Supplementary Figure S3) were taken into consideration. Cells were imaged once prior to adding aphidicolin (cat# A0781, Sigma Aldrich, St Louis, MO, USA) to a final concentration of 50 μg/ml (150 μM) or DMSO (cat# 41639, Sigma Aldrich, St Louis, MO, USA) as a control and subsequently imaged every 5 min for 30 min (long intervals) or every 10 s for 12 min (short intervals). To ensure the complete inhibition of the DNA polymerase by the aphidicolin, cells were incubated with 10 μM BrdU (Supplementary Table S4) for 10 min directly after imaging and fixed in 3.7% formaldehyde, and BrdU was detected as described above.

#### High content screening, wide-field and confocal microscopy

All characteristics of the microscopy systems, including lasers, filters and objectives used, are summarized in Supplementary Table S6.

DAPI, 6-FAM, Alexa Fluor 488, mcherry and Cy3 signals were imaged using high content screening microscopy with 20× or 40× long working distance objectives.

Confocal z-stacks of fixed cells were acquired using the Ultra-View VoX spinning disk microscopy system.

Molecular combing samples were imaged using a Zeiss Axiovert 200 widefield microscope.

### Image analysis, statistical analysis and graphical representation

Z-stacks and montages of confocal microscopy images were created using ImageJ (<https://imagej.nih.gov/ij/>).

Full gel and blot images are depicted in Supplementary Figures S6 and S7.

**High content screening microscopy analysis.** For high content screening microscopy, image analysis was performed with the Harmony Software (PerkinElmer Life Sciences). Briefly, cell nuclei were segmented in the DAPI channel, creating DAPI masks that were used to segment the GFP and mcherry channels. For chromocenter segmentation, DAPI intense pericentric heterochromatin spots were detected (spot intensity > threshold) in the DAPI channel and the resulting mask was used to segment the GFP and mcherry channels. Cells touching the image border were excluded from the analysis. The analysis was further restricted to cells with a roundness coefficient > 0.8 and a nuclear area between 60 and 300  $\mu\text{m}^2$ . Lastly, intensity values for all three channels in the whole nucleus and in the chromocenters (Supplementary Figure S4) were measured and plotted with RStudio (version 1.1.447, <https://rstudio.com/>). Based on previous findings showing that the Tet1 levels of low Tet1CD-expressing cells and mouse embryonic stem cells (mESC) are equivalent (49), the group with low Tet1-expressing cells (fluorescence intensity 0–1000, group 1) was divided in three subgroups reflecting endogenous and physiological Tet1 expression levels (fluorescence intensity, group a: 0–100, group b: 100–500, group c: 500–1000), and the average intensity of single cells was plotted (Figure 2A). High Tet1CD-expressing cells were also divided in three groups (2: 1000–2000, 3: 2000–3000 and 4: >3000).

**RPA accumulation analysis.** To analyze the focal RPA accumulation upon aphidicolin/DMSO treatment, cell nuclei were segmented using the Volocity software (Version 6.3, Perkin Elmer), GFP-RPA intensities were measured and the coefficient of variation ( $c_v = \sigma/\mu$ , with  $\sigma$  = standard deviation and  $\mu$  = mean), as a proxy for helicase speed/activity was calculated for all time-points (Supplementary Figure S5). All values were normalized to the pretreatment  $c_v = c_v(tp_x)/c_v(tp_0)$  with  $tp_x$ : any given time point imaged,  $tp_0$ : pretreatment time point) and plotted using RStudio (version 1.1.447).

Slopes of the RPA accumulation curves were calculated as a ratio of the rise (difference of the  $y$ -coordinates) over the run (difference of the  $x$ -coordinates) of the respective linear regression line of the average RPA accumulation curves. To avoid taking plateau phases of RPA accumulation into consideration, slope calculations were restricted to the first 20 min of drug treatment (0–20 min).

**DNA fiber analysis.** Replication signals were measured using ImageJ, and replication fork speed (RFS) was calculated as a ratio of the track length (1  $\mu\text{m}$  corresponds to 2 kb) and the time of nucleoside application. Fluorescent DNA fiber tracks were selected according to their pattern and only lengths of the second pulse (CldU) of progressing forks (CldU track preceded by a clear IdU signal) were considered.

**DNA and RNA polymerization quantification.** To quantify DNA polymerization in radioactive yeast replication reactions, the complete signal below the 3 kb end-labeling band was measured, background was corrected and for each template the respective –DDK signal was subtracted from the +DDK reaction. All values were normalized to the dC control template.

To quantify RNA polymerization in radioactive yeast transcription reactions, the signal intensity of the full-length transcript was determined for each reaction, using QuantityOne (4.6.6 Basic, Bio-Rad, Hercules, CA, USA). The background intensity was subtracted from each signal and then divided by the signal intensity of the full-length transcript generated from the unmodified dC template after 10 min (normalized signal intensity). The normalized signal intensities of 12 repeats performed with different Pol I purifications, different template preparations and different  $\alpha^{32}\text{P}$ -CTP batches were plotted using RStudio (version 1.1.447).

**Statistical analysis.** Independent two-group Student's  $t$ -tests and Mann–Whitney–Wilcoxon tests were performed using RStudio (version 1.1.447). Statistical significance is indicated in the plots and detailed values for statistics [number of cells ( $n$ ), median, mean, standard deviation (StDev), 95% confidence interval (95% CI) and  $P$ -values] are summarized in Supplementary Table S7 (main figures) and Supplementary Table S8 (supplementary figures).

**Graphical representation.** Boxplots and violin plots represent the median (center line) with the box depicting the 25–75 percentiles, and the dotted lines represent the upper and lower whiskers with 1.5 times the IQD (interquartile distance). Bar and line plots show normalized averaged values, and error bars show the respective standard deviation. HRM curves show the normalized average derivative of the technical replicates of one biological replicate over the temperature increase. Line intensity profile plots represent the fluorescence intensities along the distance of the selected arrow segment.

## RESULTS

### DNA base modifications affect DNA helix stability *in vitro*

There is evidence that DNA base modifications by themselves might affect the DNA primary structure (18,24,50) and consequently may influence global genome processes, such as DNA replication, transcription and repair. In a first step, we investigated the stability of the double-stranded DNA helix *in vitro*. To that end, we performed polymerase chain reactions (PCR) using different DNA polymerases and modified base containing nucleotides. The replication of double-stranded DNA *in vitro* consists of three major steps: double-stranded DNA denaturation, primer annealing and primer elongation, where, in mammals, only the four canonical DNA bases are incorporated into newly synthesized DNA. All naturally occurring base modifications in mammalian DNA are generated post-replicatively by specific enzymes (3,4,9,13,51). To ensure that the DNA polymerase used in our *in vitro* experiments did not recognize incorporated modified nucleotides as mismatch and, consequently, removed the modified nucleotide

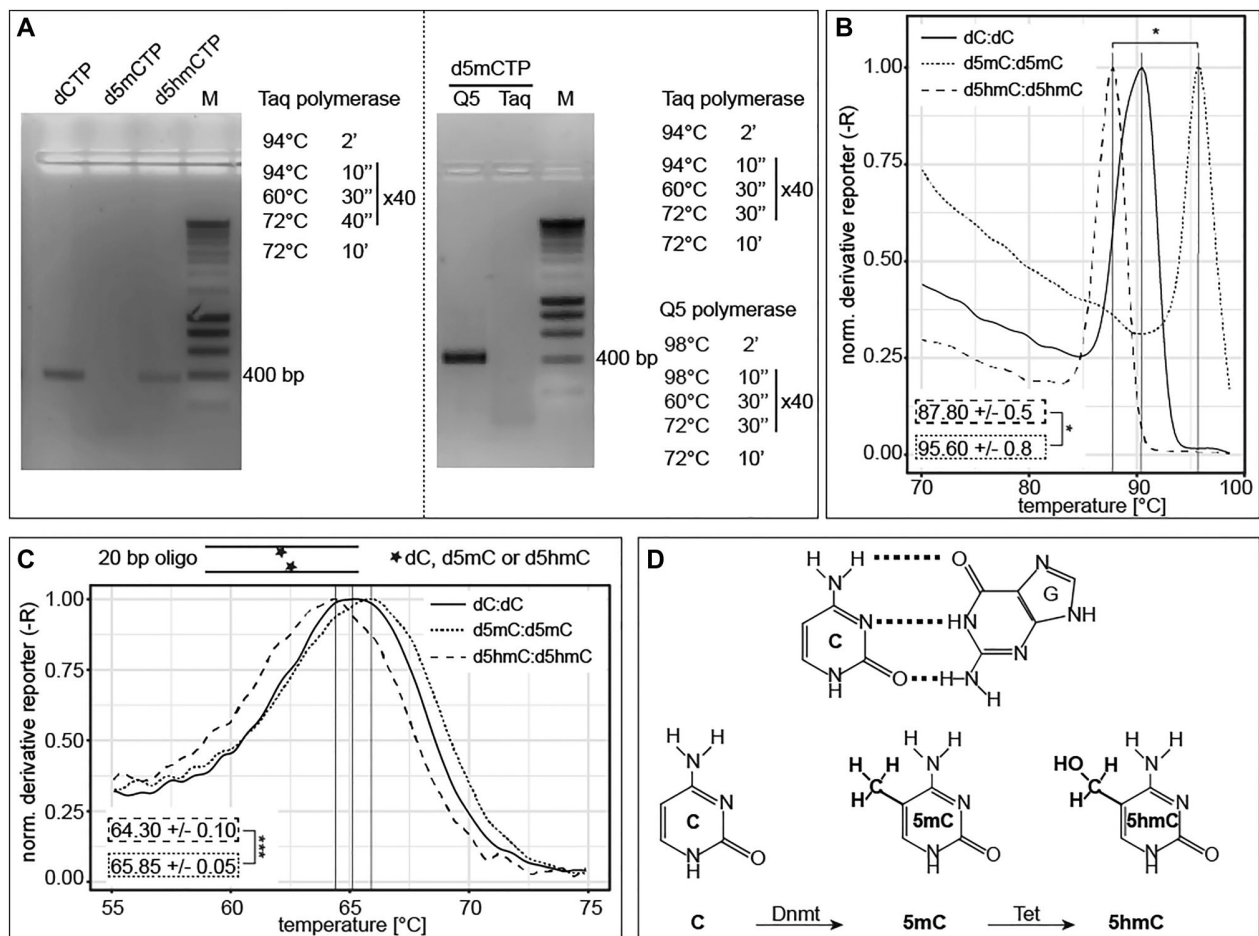
via 3' to 5' exonuclease activity, we first analyzed the synthesis efficiency of the 3 $\rightarrow$ 5' *exo*<sup>-</sup> Taq polymerase. Therefore, PCR reactions containing modified cytosine triphosphates were performed. All reactions contained deoxy-adenosine (dATP), deoxy-thymidine (dTTP) and deoxy-guanosine (dGTP) triphosphates and either deoxy-cytidine (dCTP), deoxy-5-methylcytidine (d5mCTP) or deoxy-5-hydroxymethylcytidine (d5hmCTP) triphosphates. Cycling conditions included denaturing steps at 94°C, as using higher denaturing temperatures decreased Taq polymerase activity/stability and consequently also PCR product yield (Supplementary Figure S8A). PCR reactions containing the standard dNTP mix (dATP, dTTP, dGTP and dCTP) resulted in one specific amplicon of 375 bp. Replacing dCTP with d5mCTP did not yield a PCR amplification product (Figure 1A, left), although Taq polymerase is able to incorporate d5mCTPs at higher denaturing temperatures (Supplementary Figure S8B). Several bacterial and archaeal polymerases have been shown to efficiently integrate d5mCTP (52), and family A polymerases like Taq and Klenow polymerase were shown to have similar incorporation rates for cytosines with small C5 substitutions and unmodified cytosine (53). These results demonstrate that the observed difference in PCR product amount between dCTP and d5mCTP containing reactions is not due to altered (modified) nucleotide affinity of the DNA polymerase (54). The use of d5hmCTP, Taq polymerase and denaturation at 94°C, however, resulted in a PCR product (Figure 1A, left). An amplification fragment was also obtained using d5mCTPs when the Taq polymerase (3' $\rightarrow$ 5' *exo*<sup>-</sup>) was replaced by Q5 polymerase (3' $\rightarrow$ 5' *exo*<sup>+</sup>), and the denaturing temperature in the cycling conditions was increased to 98°C (Figure 1A, right), which did not impair the activity of Q5 (Supplementary Figure S8A). Moreover, amplification using the Q5 polymerase at 98°C resulted in different yields of amplicon DNA depending on the cytosine modification (Supplementary Figure S8C and Supplementary Table S8). All modified nucleotides were also stably integrated in the PCR products as tested by slot blot analysis with antibodies specific to the respective base modifications (Supplementary Figure S8D). Altogether, these data suggest that cytosine modifications might themselves affect the denaturing temperature of the DNA. To directly test this, we measured the thermal double-stranded DNA stability using our established high-resolution melting (HRM) temperature assay (47). We first analyzed modified 375 bp long human LINE1 promoter PCR fragments (15) containing 222 cytosines (191 cytosines not included in the primer binding sites, i.e., modified) with 22 CpGs. We measured the highest melting temperatures for methylated PCR fragments followed by unmodified cytosine and d5hmC-containing amplicons (Figure 1B). d5caC-containing DNA fragments showed similar results to d5hmC, whereas d5fC showed the lowest melting temperatures (Supplementary Figure S8E and Supplementary Table S8). Similar results were obtained for HRM analysis of the 377 bp long MINX amplicon, containing 201 cytosines and 46 CpGs (188 cytosines and 42 CpGs not included in the primer regions) (Supplementary Figure S8F and Supplementary Table S8). Next, we analyzed the melting temperature of 20 bp DNA oligonucleotides containing a single central (modified) CpG

(salt corrected melting temperature of 64.2°C). We also measured a significantly lower melting temperature for the 5hmCpG containing oligonucleotide in comparison to the 5mCpG containing one (Figure 1C). Furthermore, an even lower melting temperature for the 5fCpG oligonucleotide was observed (Supplementary Figure S8G), which is in accordance with our previous PCR amplicon HRM analysis. These results indicate that DNA modifications affect DNA double helix thermal stability *in vitro* in the way that methylated cytosines stabilize the DNA double helix, whereas d5hmC, d5caC and especially d5fC negatively influence DNA duplex stability. Our results ( $T_m[\text{d5hmC}] < T_m[\text{dC}] < T_m[\text{d5mC}]$ ) are in line with previous studies analyzing melting temperature of longer oligonucleotides with several modifications (20). However, 10mer duplexes showed lowest melting temperatures for unmodified DNA (55).

### Modified cytosine levels can be manipulated in cells

In cells, levels of DNA base modifications can be manipulated by either changing the expression levels of the respective writer and/or modifier enzymes or introducing modified bases into the cell nucleus that will subsequently be incorporated into newly synthesized DNA. In mammalian cells, methylated cytosines are substrates of Tet dioxygenases ((9,13,51), Figure 1D and Supplementary Figure S8G). To increase d5hmC levels in DNA, mouse myoblast cells were transfected with the mcherry-tagged catalytic domain of Tet1 (Tet1CD), and 5hmC levels were quantified 24 h later using a 5hmC specific antibody. As shown before, further oxidized forms of d5mC have similarly helix destabilizing effects to d5hmC (Figure 1B and C, Supplementary Figures S8E–G and (20,25,26,55)) and, thus, would have similar effects on DNA strand separation. Previous studies showed that the Tet1 protein levels in low overexpressing cells are comparable to the endogenous Tet1 level in mES cells (49). We, therefore, restricted our analyses to these cells, as they represent a physiological relevant system. In these low Tet1CD-overexpressing cells, d5hmC levels significantly increased proportionally to Tet1CD protein levels, while the mcherry tag alone did not alter the d5hmC levels (Figure 2A and Supplementary Table S7). Importantly, contrary to Tet1 full-length (Tet1f), Tet1CD lacks the large N-terminal regulatory domain, including the CXXC zinc finger. This allows Tet1CD binding not only to unmethylated but also to hypermethylated CGIs, thereby inducing significant global genomic d5mC oxidation. Additionally, Tet1CD catalytic activity is not biased to specific genomic regions and can act randomly throughout the methylated genome (56). Consequently, highly methylated pericentromeric heterochromatin (chromocenters) are targeted by Tet1CD and showed increased d5hmC levels in transfected cells (Figure 2A).

Since overexpression of Tet1CD leads to significantly increased levels of Tet1 protein in the cell nucleus and on chromatin, we selected a different additional approach to enhance the levels of modified cytosine bases in genomic DNA, relying on deoxynucleotide triphosphates (dNTPs) incorporation upon cell electroporation. Genomic d5mC and d5hmC levels significantly increased 24 h after cell transfection. Additionally, cell transfection with dCTP did



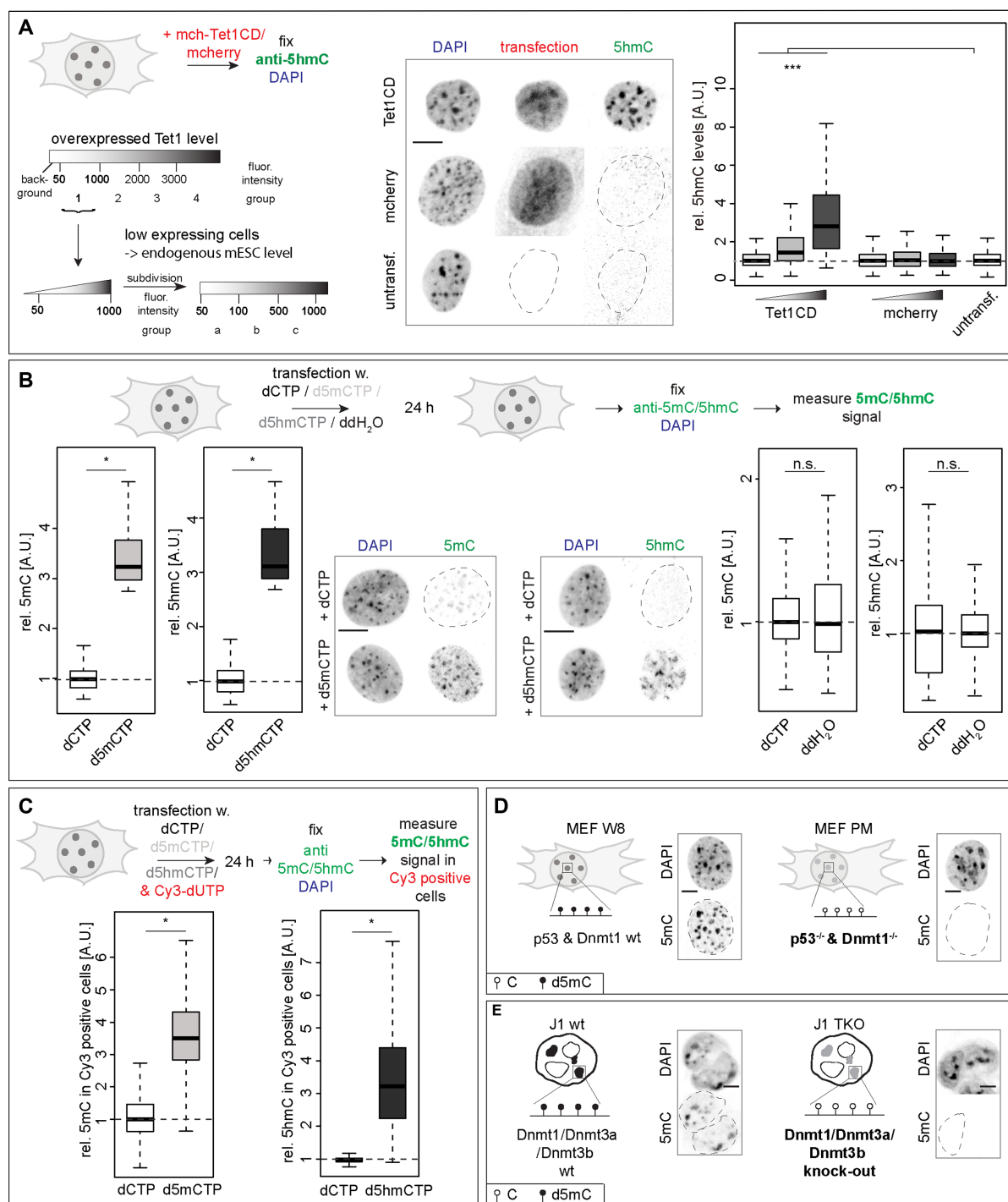
**Figure 1.** Effect of cytosine modifications on DNA double helix stability *in vitro*. (A) Agarose gel electrophoretic analysis of LINE1 PCR amplification products obtained with Taq polymerase, denaturing at 94°C and using dCTP, d5mCTP or d5hmCTP (left). Agarose gel electrophoretic analysis of LINE1 PCR amplification products obtained with Taq or Q5 polymerase, denaturing at 94 or 98°C and using d5mCTP (right). Complete PCR conditions are indicated. M stands for DNA size markers. (B and C) High-resolution melting temperature (HRM) analysis of LINE1 PCR fragments containing dC, d5mC or d5hmC, respectively, (B) and of 20 bp oligonucleotides containing one central (modified) CpG dinucleotide (CpG, 5mCpG or 5hmCpG) (C). Shown are the corresponding derivatives of the normalized mean SYBR green fluorescence intensity values of technical replicates over increasing temperature. Independent experiments were repeated in duplicates with three (B) or four (C) technical replicates, respectively, and one representative result is depicted. (D) Cytosine–guanine base pairings and (hydroxy)methylated cytosine bases with the respective writer enzymes are depicted. *P*-values were determined by Student's *t*-test; \**P* < 0.05 and \*\*\**P* < 0.001.

not change d5mC or d5hmC levels. Co-transfection of dCTP, d5mCTP or d5hmCTP and Cy3-labeled dUTPs (Cy3-dUTP) resulted in d5mC and d5hmC increase in Cy3 positive cells (Figure 2B and C; Supplementary Table S7). The aforementioned approaches lead to an increase of the respective cytosine bases; hence, we also opted for loss of function experiments with concomitant lowering of the cytosine modification. Therefore, we chose *Dnmt1*<sup>-/-</sup> mouse embryonic fibroblasts (MEF PM) and *Dnmt* triple knockout (TKO) mouse embryonic stem (ES) cells that are depleted of all cytosine base modifications (Figure 2D and E and (35,57)).

In summary, with the above strategies we can efficiently up and down modulate cytosine modification levels in cells by interfering with levels of Tet1 and Dnmts (Supplementary Figure S9A and B) or by introducing modified cytosine bases into the cells (Supplementary Figure S9C).

### DNA base modifications affect double helix stability in cells

The next step was to analyze the DNA thermal stability differences in the above-mentioned cellular systems. We, therefore, made use of genomic DNA denaturation in cells via heat treatment followed by fluorescence *in situ* hybridization (FISH). We employed a DNA probe specific for pericentromeric heterochromatic major satellite DNA (MaSat, chromocenters), which contains high levels of d5mC (38,58). We first tested the temperature range in which *in situ* denaturation of the genomic DNA and, subsequently, annealing of the probe were possible. We, therefore, generated MaSat probes via PCR including biotin-conjugated dUTPs, combined them with fixed and permeabilized C2C12 mouse myoblast cells, denatured DNA at different temperatures (60, 70 and 80°C) and annealed the probe at 37°C. The highest MaSat signals were obtained at the optimal denaturing temperature of 80°C and



**Figure 2.** Manipulation of modified cytosine levels in cells. (A) Experimental setup to determine d5hmC levels in mch-Tet1CD overexpressing cells and representation of the cell grouping and binning approach applied. Only cells with Tet1 expression levels corresponding to group 1 (fluorescence intensity 50–1000, low Tet1 expressing cells) were analyzed and further subdivided in three subgroups (fluorescent intensities group a: 50–100, group b: 100–500 and group c: 500–1000) during plotting. Relative 5hmC sum fluorescence intensities in C2C12 cells, 24 h after transfection are shown. (B and C) Relative 5mC and 5hmC sum intensities within the whole (Cy3 positive) C2C12 nucleus 24 h after transfection with dCTP, d5mCTP, d5hmCTP or ddH<sub>2</sub>O (B) or co-transfection with Cy3-dUTPs (C) as indicated. (D and E) Schematic representation of the genotype and cytosine modification state in MEF W8 (wild-type) and MEF PM (*p53*<sup>-/-</sup> & *Dnmt1*<sup>-/-</sup>) cells (D) and in J1 mouse ES (wild-type) and J1 TKO (*Dnmt1*<sup>-/-</sup>, *Dnmt3a*<sup>-/-</sup> and *Dnmt3b*<sup>-/-</sup>) ES cells (E) and representative spinning disk confocal images of immunofluorescent detection of 5mC in wt and KO cells. Independent experiments were done in duplicates (B–E) or in triplicates (A) and *P* (determined by Student's *t*-test) and *n*-values are summarized in Supplementary Table S7. All fluorescence signals are plotted as a ratio to the values of the respective control cells. All boxes and whiskers represent 25–75 percentiles and 1.5 times the IQD (interquartile distance), respectively, and the center line depicts the median. Dotted lines represent nuclear contours; A.U.: arbitrary units; scale bar = 5 μm; \* *P* < 0.05, \*\*\* *P* < 0.001 and n.s.: non-significant.

MaSat signal decreased with decreasing denaturing temperatures (Supplementary Figure S10A and Supplementary Table S8). Nuclear and chromocentric MaSat FISH signal intensity increased significantly in Tet1CD but not in mcherry transfected cells when DNA was denatured at 60°C (Figure 3A, Supplementary Figure S10B and Supplementary Tables S7 and S8). Similar results were observed when DNA was denatured at 65 and 70°C (Supplementary Figure S10B and Supplementary Table S8).

To validate that our findings are not C2C12 mouse myoblast specific, we performed the FISH based assay also in Tet1CD transfected mouse embryonic fibroblast (MEF W8) cells, which yielded the same outcome of increased MaSat signal at low denaturation temperatures (65, 70 and 75°C) in Tet1CD transfected cells (Supplementary Figure S11A–C and Supplementary Table S8).

Previous studies found that cells enriched for d5hmC showed large and decondensed nuclei as well as decompacted heterochromatin (8,59,60). Since a more decondensed constitutive heterochromatin results in a more homogeneous DAPI staining throughout the nucleus and, in turn, to lower DAPI standard deviations, we measured the latter as a proxy for DNA decondensation. Indeed, increased levels of Tet1CD led to chromocenter decondensation (Supplementary Figure S10C and Supplementary Table S8). Therefore, we investigated whether the latter FISH results were due to Tet1CD induced chromatin decondensation facilitating DNA denaturation. Hence, we treated cells with trichostatin A (TSA), a potent histone deacetylase inhibitor. After 72 h of treatment, a significant increase in histone H3 acetylation and concomitantly chromatin decondensation was measured, which is in agreement with previous studies (38). d5hmC levels *per se* were not affected by TSA treatment and, importantly, TSA-treated cells did not show increased MaSat probe signal (Supplementary Figure S10D and Supplementary Table S8), indicating that chromatin decondensation does not facilitate thermal DNA denaturation.

To further rule out that we are measuring effects due to changes in overall chromatin structure, composition and accessibility as a consequence of Tet1CD overexpression, we performed the denaturation assays in cells transfected with nucleotides. In comparison to control-treated cells, increased d5mC levels led to significantly lower MaSat signals, whereas increased genomic d5hmC content increased the amount of probe annealing (Figure 3B and Supplementary Table S7).

Lastly, we performed the denaturation assays in the Dnmt loss of function systems (MEF PM and J1 TKO), where we observed a significant increase in MaSat probe signals (Figure 3C and D; Supplementary Table S7). This result is consistent with methylated DNA double helix being more stable than DNA containing unmodified cytosines.

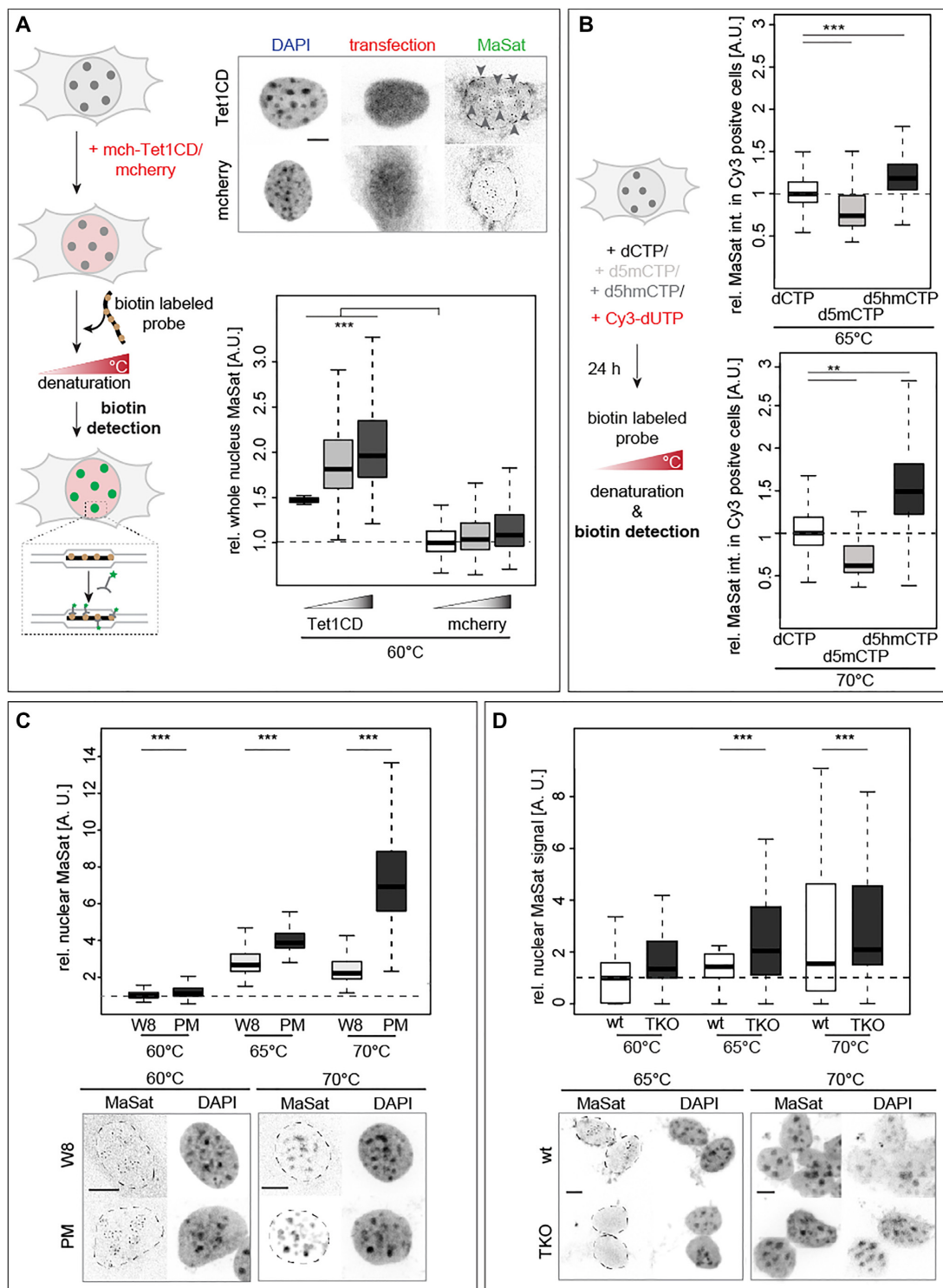
Altogether these results indicate that modified cytosines influence DNA double helix stability, both negatively and positively. Methylated cytosines render the double helix more stable and, therefore, DNA denaturation needs more energy, i.e., higher temperatures, whereas the presence of d5hmC leads to an easier strand separation, due to a less stable DNA helix. Since global genome processes, such as DNA replication, transcription and repair, all require un-

winding of the DNA double helix, alterations of DNA base-pairing stability could fundamentally influence these processes. We, therefore, aimed to analyze the impact of cytosine base modifications on RNA and DNA polymerization, as well as on DNA unwinding by helicase enzymes.

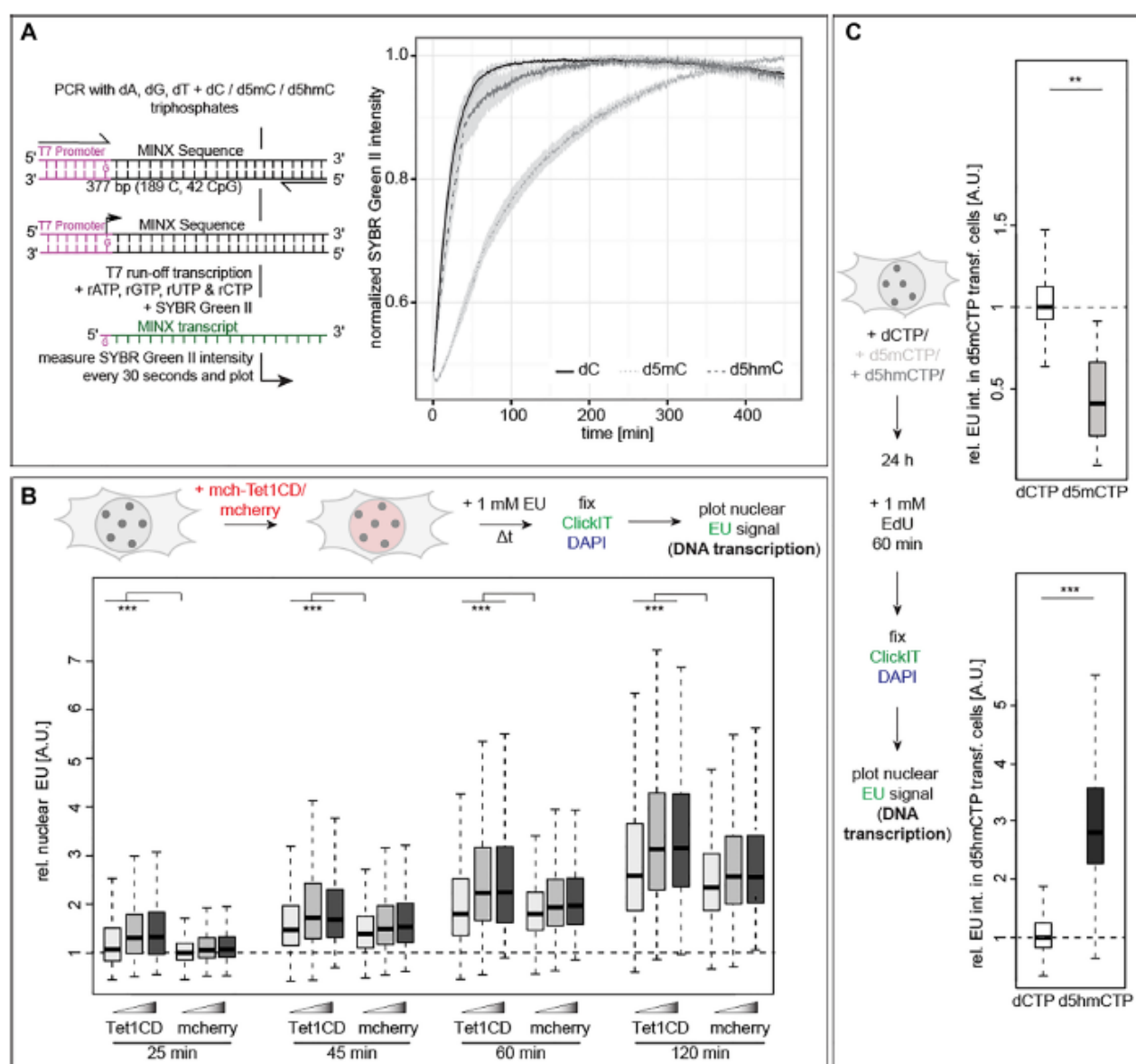
### DNA base modifications affect DNA transcription

One of the major DNA metabolic processes within the cell nucleus is DNA transcription. As this ubiquitous process also depends on DNA double helix unwinding and the subsequent RNA polymerization by a DNA-dependent RNA polymerase, we addressed the effect of modified cytosines on RNA polymerase activity *in vitro*. We, therefore, performed an *in vitro* transcription assay using *in vitro* modified DNA templates containing an unmodified T7 promoter to avoid gene silencing in d5mC containing templates (61–65), followed by a MINX sequence and purified T7 RNA polymerase. The transcription activity of T7 RNA polymerase was detected in real-time with unmodified, methylated and hydroxy-methylated templates (Figure 4A). Importantly, high structural, functional and enzymatic similarities were reported among organisms for RNA polymerases I and II (66,67). We found that methylated templates resulted in significantly lower transcript amounts compared to unmodified or hydroxymethylated templates (Figure 4A). Since RNA polymerase elongation effects could be a concern in the *in vitro* transcription (IVT) assays, we performed similar experiments on (un)modified single-stranded (ssDNA) templates containing a double-stranded (dsDNA) T7 promoter region. Similarly to the dsDNA templates, we observed SYBR Green II fluorescence for the ssDNA templates, although the overall signal intensity, i.e. the overall amount of RNA transcripts, was lower than for dsDNA templates (Supplementary Figure S12A). Additionally, we measured no significant difference in template elongation between the unmodified and 5mdC- or 5hmC-containing substrates. This is in agreement with previous studies revealing similar GTP incorporation rates opposite of dC, d5mC and d5hmC (68,69). We, therefore, conclude that the differences in transcription measured with the dsDNA template are due to changes in DNA strand separation, i.e. DNA melting, and not due to differences in elongation rates.

To further extend the *in vitro* transcription assays to eukaryotic enzymes, we analyzed *Saccharomyces cerevisiae* DNA-dependent RNA polymerase I (Pol I) transcription on (un)modified templates. The majority of the RNA produced in cells is produced by RNA polymerase I and, hence, this enzyme better represents the measurements *in vivo* (see below). To allow nonspecific RNA polymerase initiation, we generated templates containing a 3'-overhang (3'-tailed templates) based on the nicking enzyme Nb.BsmI (46). Slot blot analyses with antibodies specifically recognizing 5mC and 5hmC, respectively, confirmed the modification of the templates (Supplementary Figure S13A). *In vitro* transcription assays contained endogenous 14-subunit *S. cerevisiae* Pol I (44), and synthesized RNA was visualized and quantified by detecting the incorporated radiolabeled nucleotides, as a proxy for DNA transcription (Supplementary Figure S13A). We observed a clear increase of RNA products with increasing incubation time. The use of similar amounts of



**Figure 3.** Effect of cytosine modifications on DNA double helix stability *in vivo*. (A) Major satellite (MaSat) DNA probes were used for fluorescence *in situ* hybridization (FISH) on transfected C2C12 mouse myoblast cells with genomic DNA denaturation performed at 60°C. Representative spinning disk confocal microscopy images with MaSat probe hybridization detection and DNA counterstaining are shown. Relative quantifications of the nuclear MaSat sum intensity are shown. (B) Relative major satellite DNA FISH sum intensities within the whole C2C12 nucleus of Cy3 positive cells at 65 or 70°C denaturation temperature 24 h after co-transfection with dCTP, d5mCTP, d5hmCTP and Cy3-dUTP as indicated. (C and D) Relative major satellite DNA FISH sum fluorescent intensities in MEF W8 and MEF PM nuclei (C) and in J1 wt and J1 TKO nuclei (D) at 60, 65 and 70°C denaturation; scale bars: 5  $\mu$ m. Independent experiments were done twice (C and D) or in triplicates (A) and P (determined by Student's *t*-test) and *n*-values (number of cells) are summarized in Supplementary Table S7. All boxes and whiskers are as in Figure 2. Dotted lines represent the nuclear or chromocenter contours. A.U.: arbitrary units; \*  $P < 0.05$ , \*\*  $P < 0.005$  and \*\*\*  $P < 0.001$ .



**Figure 4.** Effect of cytosine modifications on RNA polymerase activity. (A) Experimental setup for *in vitro* T7 run-off transcription. Template DNA contained 100% cytosine (dC), methylcytosine (d5mC) or hydroxymethylcytosine (d5hmC) and T7 promoter sequence was unmodified. Normalized mean SYBR green II intensities ( $\pm$  standard deviation), i.e., the amount of RNA transcripts, over time are shown. (B) Schematic representation of the experimental setup for *in vivo* ribonucleotide incorporation analysis. Cells were incubated with 1 mM EU 24 h after transfection with the constructs as indicated, EU was detected and the normalized nuclear EU sum intensities are plotted. (C) Schematic representation of the experimental setup for *in vivo* ribonucleotide incorporation analysis in dCTP, d5mCTP or d5hmCTP transfected cells. Cells were incubated with 1 mM EU 24 h after transfection with the nucleotides as indicated, EU was detected and the normalized nuclear EU sum intensities are plotted. Independent experiments were done twice (A and C) or in triplicates (B). All boxes and whiskers are as in Figure 2 and *P* and *n*-values are summarized in Supplementary Table S7; \*\* *P* < 0.005 and \*\*\* *P* < 0.001.

methylated templates, however, resulted in a clear trend indicating partial inhibition of transcription compared to unmodified templates. In contrast, d5hmC-containing DNA templates showed a reversion of transcription impairment and synthesized RNA amounts were similar to the dC containing control template (Supplementary Figure S13A). The fact that these effects can be detected even using budding yeast RNA polymerase, which in the natural environment of a yeast cell does not face modified DNA, denotes a high degree of evolutionary conservation of the enzymes

involved in genome metabolism. Whereas the effect of mC-containing templates on Pol I transcription is subtle, a previous study showed that DNA methylation contributes to Pol I inhibition *in vivo* (70). Furthermore, Pol I transcription can overcome stronger opposing forces than Pol II *in vitro* (71), rationalizing the scale of the effects observed.

Next, we investigated the influence of cytosine base modifications on RNA polymerases and transcription rates *in vivo*. To this end, we incubated cells for various periods of time with EU (5-ethynyl-uridine), an uridine derivative that

is incorporated into nascent RNA and can be detected via Click chemistry. The results showed a time dependent increase of EU signal (labeled RNA). Additionally, Tet1CD-transfected cells showed significantly higher EU levels than mcherry expressing cells, for all analyzed time points (Figure 4B and Supplementary Table S7). Since the nucleoli are subnuclear compartments where rDNA transcription by DNA dependent RNA polymerase I takes place, we additionally analyzed the EU levels in these specialized nuclear regions. Similar to whole nuclear EU results, we found elevated RNA levels in Tet1CD transfected cells (Supplementary Figure S12B and Supplementary Table S8). To relate these *in vivo* observations to the *in vitro* transcription assays, we also measured EU intensities in high Tet1CD expressing cells. Interestingly, and in accordance with the single enzyme reactions, we see similar levels of RNA in control and cells with high d5hmC levels (Supplementary Figure S13B).

We note, however, that the increased transcription in cells with high d5hmC levels may, in addition to the helix destabilizing effect, be due to the release of transcriptional inhibition by cytosine methylation and associated reader proteins (72). To bypass potential effects caused by ectopically overexpressing genes, we analyzed DNA transcription in nucleotide transfected cells. Nuclear EU levels decreased upon incorporation of d5mCTP into the genomic DNA, whereas the transcription rate increased upon elevating genomic d5hmC levels (Figure 4C and Supplementary Table S7). These results are in agreement with previous reports showing that DNA methylation inhibits transcription, whereas transcription is re-initiated upon oxidation of the d5mC. Moreover, this is also consistent with our findings that DNA metabolic processes are slowed upon DNA methylation due to helix stability increase whereas oxidation of d5mC to d5hmC results in a less stable DNA duplex facilitating DNA unwinding and subsequent DNA/RNA polymerization reactions.

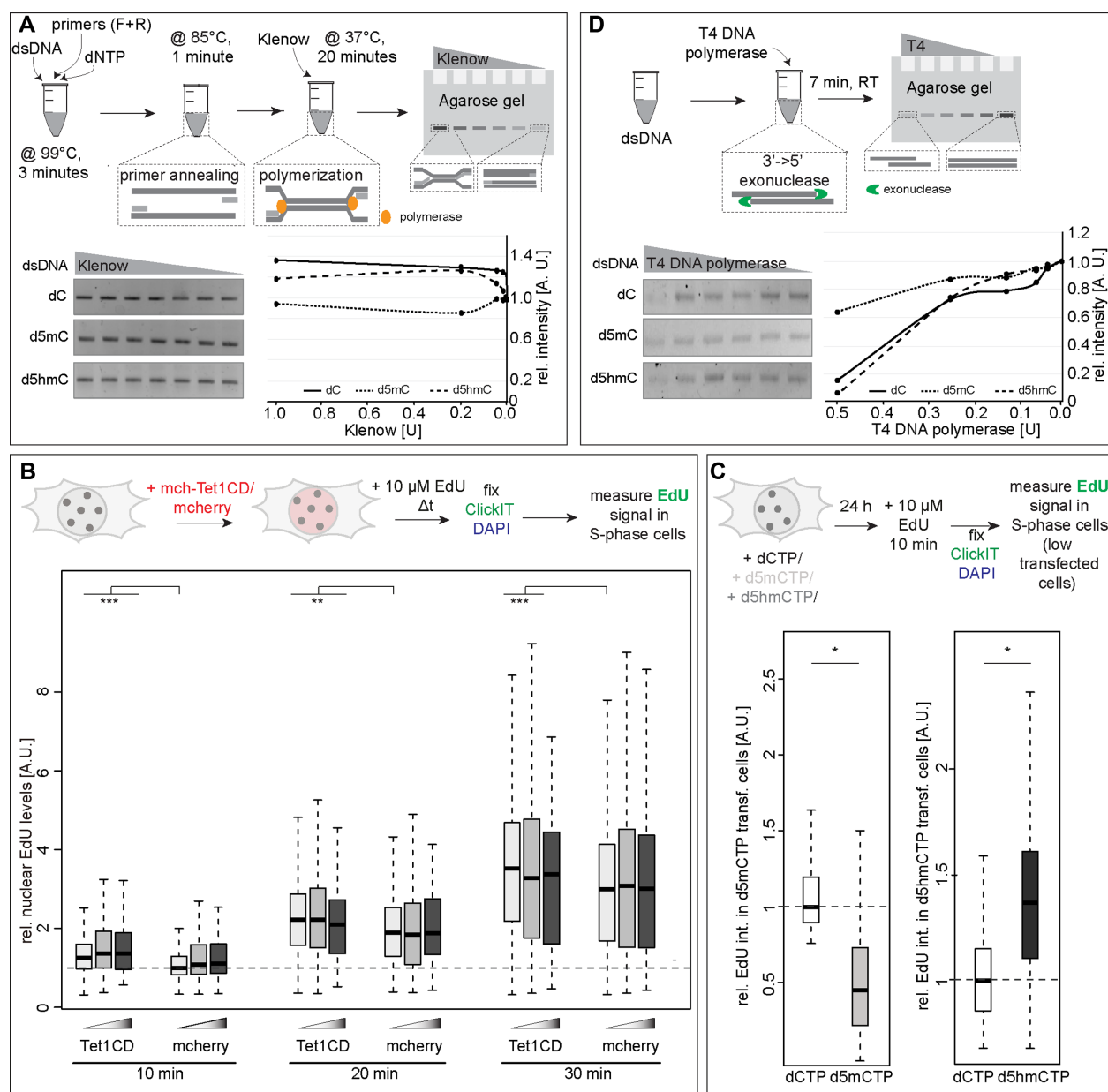
Altogether these *in vitro* and *in vivo* data indicate that DNA transcription rates decrease in the presence of stabilizing d5mC with d5hmC reverting this inhibition and leading to increased synthesis of RNA.

#### DNA base modifications affect DNA polymerase and exonuclease activities

Besides RNA synthesis, DNA polymerization during DNA replication and repair is an essential nuclear process to ensure complete and error-free genome duplication. We, thus, aimed to investigate the effect of DNA base modifications on DNA polymerase activity. To test this, we performed *in vitro* DNA replication assays using *in vitro* modified DNA templates and *in vitro* purified Klenow or Taq DNA polymerases. The function of DNA polymerases in DNA synthesis is conserved across species and the basic reaction mechanism of DNA double helix opening does not differ between prokaryotic and eukaryotic organisms (73,74). Modified dsDNA template, dNTPs and primers were heated to denature DNA, cooled down to allow primer annealing and incubated at 37 or 72°C while adding Klenow or Taq polymerase, respectively. Strand displacement and DNA polymerization were analyzed by

agarose gel electrophoresis. The amount of DNA synthesized, increased with increasing polymerase concentrations, and d5mC modified templates resulted in less DNA amplification compared to d5hmC and unmodified cytosine (Figure 5A; Supplementary Figure S14A and Supplementary Tables S7 and S8). These results indicate that modified cytosines influence the *in vitro* DNA polymerization rate in a way that correlates with the helix stabilization effect of the modified nucleotide present in the template DNA. Similarly to RNA polymerases, several archaeal (KOD), bacterial (Klenow  $exo^-$ ) and human (polymerase  $\delta$  and  $\eta$ ) DNA-dependent DNA polymerases were also shown to discriminate barely, if any, between dGTP incorporation opposite unmodified and modified cytosines and very similar dGTP incorporation rates were found for C, d5mC, d5hmC and d5fC (75–78). Consequently, our results most probably reflect differing helix stability rather than different polymerase specificities or elongation effects.

To extend our *in vitro* enzymatic assay results to an eukaryotic system, we performed processive soluble replication reactions with the purified *Saccharomyces cerevisiae* replisome components that are required for replication *in vitro* as described in (42). Double-stranded 3 kb linear template DNA contained an unmodified central ARS306 yeast origin of replication sequence with two opposing ORC-binding sites. This allows loading of an MCM helicase double hexamer and gives rise to bidirectional replication. DNA templates additionally contained 3' and 5' (un)modified sequences flanking the origin (Supplementary Figure S15A). To further analyze the influence of different amounts of modified cytosines and since 100% modified templates are unlikely to appear at high frequencies in cells, we included DNA templates containing various amounts of d5mC or d5hmC. To this end, PCR reactions contained either 100% (un)modified dCTPs, 50% modified and 50% dCTPs or 12.5% modified and 87.5% unmodified dCTPs (Supplementary Figure S15A). Indeed, the slot blot analysis revealed a clear decrease in signal intensity correlating with the decrease in the amounts of modified nucleotides present in the PCR reaction (Supplementary Figure S15A). *In vitro* replication assays contained replisome components enabling leading and lagging strand synthesis at comparable rates of *in vivo* replication (42), and synthesized DNA was visualized and quantified by detecting the incorporated radiolabeled nucleotides, as a proxy for DNA synthesis rates (Supplementary Figure S15A). While the yeast-based assay is no longer a single enzyme experiment, the use of purified proteins, yet, allows a very defined molecular composition of the reaction. Similarly to our results obtained with bacterial enzymes described above (Figure 5A and Supplementary Figure S14A), we observed less DNA replication for methylated templates compared to unmodified or hydroxymethylated templates (Supplementary Figure S15A). This effect was gradually reverted with decreasing d5mC levels in the template DNA. Interestingly, replication of 100% d5hmC-containing DNA was also lower, and reducing d5hmC amounts in the templates led to DNA replication rates similar to the unmodified dC template. The reversal of the d5hmC replication inhibition by decreasing modified nucleotide amounts was more pronounced than for the d5mC samples. Although we also measured this ef-



**Figure 5.** Effect of cytosine modifications on DNA polymerase activities. (A and D) Experimental setup for *in vitro* polymerization (A) and *in vitro* exonuclease (D) assays. Template DNA contained 100% cytosine (dC), methylcytosine (d5mC) or hydroxymethylcytosine (d5hmC) (except primer sequences). Exemplary agarose gel images for Klenow polymerase assay (A) and for T4 DNA polymerase exonuclease assay (D) and the quantification of the corresponding sum band intensities as a proxy for polymerization (A) and exonuclease activity (D), normalized to the DNA input (0 U enzyme) are shown. (B) Schematic representation of the experimental setup for *in vivo* nucleotide incorporation analysis. Cells were incubated with 10 μM EdU for the pulse labeling times indicated 24 h after transfection, EdU was detected and normalized nuclear EdU sum fluorescence intensities of S-phase cells are plotted. (C) Schematic representation of the experimental setup for *in vivo* nucleotide incorporation analysis in dCTP, d5mCTP or d5hmCTP transfected cells. Cells were incubated with 10 μM EdU for the pulse labeling times indicated 24 h after nucleotide transfection; EdU was detected and normalized nuclear EdU sum fluorescence intensities of S-phase cells are plotted. Independent experiments were done twice (A, C and D) or in triplicates (B) and *P* and *n*-values are summarized in Supplementary Table S7. Full gels are shown in Supplementary Figure S6. All boxes and whiskers are as in Figure 2; \* *P* < 0.05, \*\* *P* < 0.005 and \*\*\* *P* < 0.001.

fect in the single enzyme reactions (Figure 5A), the difference between dC and d5hmC was much smaller than in the multi-complex replisome experiment (Supplementary Figure S15A). As mentioned above for the yeast RNA polymerase assay, the effects measured denote a high conservation of the enzymatic machinery involved in genome metabolism across species even the ones, such as budding yeast, with no natural DNA modifications.

We next tested the impact of cytosine modification on mammalian DNA replication *in vivo*. As it is not feasible to study DNA polymerization in living cells in the absence of DNA unwinding by the DNA helicase complex, we extended our analysis to the whole replisome, thereby analyzing the effect of d5hmC on the activity of the coupled helicase and polymerase complex. We incubated C2C12 cells 24 h after transfection for defined time periods with the thymidine analog EdU (5-ethynyl-2'-deoxyuridine) that is incorporated into newly synthesized DNA and can be detected via Click chemistry to visualize sites of ongoing DNA replication. We found EdU levels to increase with increasing labeling time (Figure 5B; Supplementary Figure S14B, Supplementary Tables S7 and S8). Low Tet1CD transfected cells incorporated more EdU in comparison to mcherry transfected cells during all three labeling times. However, when analyzing high Tet1CD expressing cells, we observed significantly decreasing levels of EdU incorporation (Supplementary Figure S15B and Supplementary Table S8). This observation is similar to the trend observed in the yeast *in vitro* replication experiments (Supplementary Figure S15A). Previous studies showed that DNA decondensation upon TSA induced histone hyperacetylation leads to slower DNA synthesis rates (38). Low Tet1CD-expressing cells showed DNA decondensation (Supplementary Figure S10C and Supplementary Table S8); however, instead of lower EdU incorporation they showed higher EdU incorporation rates (Figure 5B; Supplementary Figure S14B, Supplementary Tables S7 and S8). This indicates that the increased d5hmC levels and the accompanying less stable DNA double helix rather than DNA decondensation influence the speed of the replisome resulting in a higher nucleotide incorporation rate.

As high DNA methylation levels are a hallmark of pericentric heterochromatin and simultaneously constitute a Tet1 oxidation substrate, we further analyzed EdU incorporation at major satellite repeats. In agreement with our previous findings, we found higher incorporation of EdU in Tet1CD-transfected cells (Supplementary Figure S14B and Supplementary Table S8).

Next, we increased modified cytosine levels in the genomic DNA by d5mCTP or d5hmCTP transfection. Whereas d5mCTP incorporation led to a significantly decreased EdU signal, d5hmCTP resulted in higher EdU incorporation rates (Figure 5C and Supplementary Table S7). This is consistent with the Tet1 overexpression data and the decrease/increase of EdU incorporation reflects the DNA helix stabilizing/destabilizing effect of d5mC/d5hmC, respectively.

Lastly, we expanded the analysis to test whether and how DNA base modifications in the template DNA affect other enzymatic activities of DNA polymerases. Hence, we incubated MINX PCR amplicons with T4 DNA polymerase.

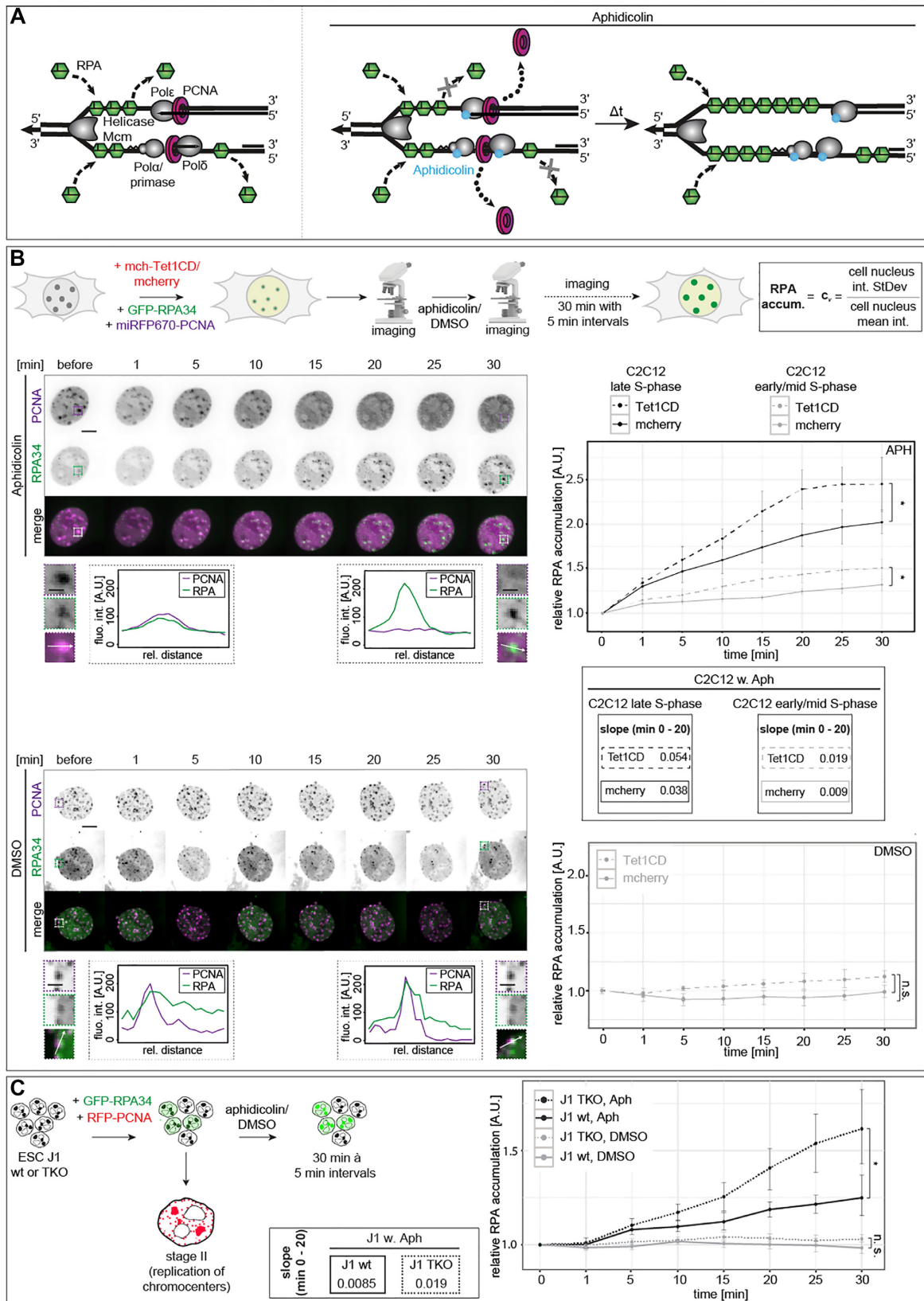
In the absence of nucleotides, T4 DNA polymerase, similar to other DNA polymerases, exhibits 3'→5' exonuclease activity. Indeed, we observed decreasing DNA amounts with increasing enzyme amounts for all three DNA templates. Exonuclease activity was, however, reduced on methylated templates (Figure 5D and Supplementary Table S7).

Taken together, these results show that DNA replication is negatively influenced by methylcytosine, whereas the oxidized methylcytosine variants revert and increase DNA polymerization speed in a dose-dependent manner.

#### DNA base modifications affect DNA helicase unwinding in cells

A prerequisite for DNA replication is DNA double helix unwinding by DNA helicases into two single strands (79,80). We reasoned that changes in the mechanical stability of the DNA double strand might influence helicase speed and processivity and, thus, analyzed helicase unwinding *in vivo*.

To test the influence of modified cytosine bases on the helicase unwinding activity *in vivo*, we made use of a system we developed before, based on the decoupling of the replisome via a drug treatment (81). For this purpose, we treated cells with the tetracyclic antibiotic aphidicolin that reversibly inhibits eukaryotic DNA polymerases  $\alpha$ ,  $\delta$  and  $\epsilon$  (82–85). For DNA replication to take place, the DNA double helix needs to be unwound by DNA helicases. The single-stranded DNA generated (ssDNA) is immediately covered with single-stranded DNA-binding proteins (RPA complex) to avoid DNA damage and/or re-annealing of the two DNA strands. The DNA synthesis complex, comprising DNA polymerases and PCNA (proliferating nuclear antigen), uses the ssDNA strands as template to synthesize a complementary daughter strand, while simultaneously displacing RPA from the ssDNA (Figure 6A and (80)). With the aphidicolin treatment, we specifically inhibit the DNA polymerase without inhibiting the helicase activity (81). As previously published (81), the DNA polymerase gets uncoupled from the DNA helicase, leading to a continuous unwinding of the DNA double helix by the helicase, while the DNA polymerization is stalled by aphidicolin. The ssDNA generated is continually covered with the single-stranded binding protein RPA. Due to the inhibition of the DNA polymerase, RPA is not displaced from the DNA, leading to an accumulation at DNA replication sites. Furthermore, the polymerase inhibition via aphidicolin leads to the disassembly of proteins involved in DNA replication elongation from replication sites (e.g., PCNA, Figure 6A and (81)). We made use of the uncoupling of helicase and polymerase to analyze the effect of modified cytosine bases and the concomitant change in double helix stability on the helicase unwinding speed, by measuring the RPA accumulation on ssDNA at DNA replication sites at defined time points after aphidicolin addition. C2C12 mouse myoblast cells were transiently transfected with expression constructs coding for fluorescently marked proteins, namely GFP-tagged RPA34 to determine the helicase unwinding speed via RPA accumulation over time, miRFP670-tagged PCNA to control for complete polymerase inhibition, indicated by dissociation of PCNA from replication foci and



**Figure 6.** Effect of cytosine modifications on DNA helicase activity *in vivo*. (A) In the absence of aphidicolin, the DNA double helix is unwound by DNA helicases, ssDNA is covered with ssDNA binding proteins (RPA) and the DNA synthesis complex displaces RPA during DNA polymerization.

mcherry-tagged Tet1CD or mcherry. Twenty-four hours after transfection, cells were subjected to live cell confocal microscopy and only cells showing a clear S-phase pattern in the GFP channel (RPA) and/or miRFP670 (PCNA) channel were taken into consideration. As a pretreatment control, z-stacks of the cells were acquired before adding aphidicolin or DMSO and after the addition of aphidicolin (DMSO for the controls), cells were imaged every 5 min over a period of 30 min (Figure 6B). Already after 5 min of drug treatment, a clear accumulation of RPA was observed at sites of ongoing DNA replication (Figure 6B). Complete inhibition of the polymerase by aphidicolin was, on the one hand, verified by PCNA dissociation from replication foci (Figure 6B) and, on the other hand, by incubating the cells with the thymidine analog BrdU (bromodeoxyuridine) after 30 min of aphidicolin treatment, with no BrdU incorporation being detected (Supplementary Figure S16A). Control cells that were treated with DMSO, neither showed RPA accumulation nor PCNA dissociation from the replisome and incubation of the cells with BrdU showed replication foci pattern colocalizing with the RPA signal (Supplementary Figure S16A). Comparison of the RPA and PCNA signals before and after DMSO treatment showed S-phase progression during the 30 min of imaging (Figure 6B). These results demonstrate that we were able to completely block DNA synthesis in S-phase C2C12 mouse myoblast cells by aphidicolin treatment.

To measure and analyze the RPA accumulation, the nuclei of S-phase cells were segmented, the GFP intensity inside each nucleus was measured for each time point (Supplementary Figure S5), normalized to the GFP channel values at time point 0 (pretreatment) and the coefficient of variation of the GFP-channel intensities was plotted for each time point. In addition, we validated that the GFP-RPA levels did not vary significantly between the mcherry-tagged Tet1CD and mcherry expressing cells (Supplementary Figure S16B and Supplementary Table S8).

DNA replication during S-phase is organized in a distinct spatio-temporal manner that, in somatic cells, follows the chromatin compaction state, starting with replication of euchromatin (early S-phase), facultative heterochromatin (mid S-phase) and constitutive heterochromatin (late S-phase) (Supplementary Figure S1C and reviewed in (80)). The latter type of chromatin is mainly transcriptionally inactive and marked by high levels of cytosine methylation, and, in mouse cells, forms prominent clusters visible as bright foci in DAPI counterstaining. Immunofluores-

cent detection of d5mC and d5hmC in mcherry-transfected cells showed low 5hmC levels, which is consistent with the low Tet1-3 expression levels in C2C12 cells (86). 5mC signals were mostly, but not exclusively, visible as clustered and compact structures corresponding to chromocenters. In Tet1CD-transfected cells, 5hmC levels increased significantly and correlated with 5mC signals (Figure 2A and Supplementary Figure S17A). Highest 5hmC increases were observed in the normally highly methylated constitutive heterochromatic genomic regions. Additionally, colocalization analyses via line intensity profiles and Pearson's correlation factors showed the strict dependency of 5hmC on 5mC (Supplementary Figure S17A). According to the genome-wide analyses of Jin *et al.* (56), Tet1CD catalytic activity is not biased to specific methylated genomic regions and can act throughout the genome. Hence, 5hmC generation and distribution 'follow' 5mC genome-wide. Considering the fundamental prerequisite of the presence of 5mC to allow 5hmC generation, hydroxymethylation is prominently observed in pericentromeric heterochromatin (chromocenters) of Tet1CD-transfected cells.

In view of the abundant chromocentric 5hmC levels in Tet1CD-transfected cells, we analyzed RPA accumulation in the helicase unwinding assays in late S-phase cells, i.e. the time of S-phase when chromocenters are replicated, and early/mid S-phase cells separately (Figure 6B). Correlation analyses revealed RPA accumulation at these highly modified regions (Supplementary Figure S17B). To confirm the presence of 5hmC in pericentromeric heterochromatin in GFP-RPA and mcherry-Tet1CD transfected cells treated with aphidicolin/DMSO, we co-stained the latter for 5mC and 5hmC and performed line profile colocalization analyses (Supplementary Figure S17B). Late S-phase cells depicted clear GFP-RPA accumulation at 5mC-stained genomic regions, a hallmark for pericentromeric heterochromatin. As previously published (81), aphidicolin treatment led to the accumulation of GFP-RPA at sites of ongoing DNA replication. Tet1CD-transfected cells additionally showed clear 5hmC signals, which, due to the strict prerequisite of 5mC for 5hmC oxidation, overlap with the immunofluorescently detected 5mC.

Detection and analysis of the helicase activity and speed in replicating C2C12 cells showed a time-dependent accumulation of RPA at replication sites of S-phase cells treated with aphidicolin (Figure 6B; Supplementary Figure S16C, Supplementary Tables S7 and S8). Cells transfected with Tet1CD, however, showed a significantly higher amount of

Upon addition of aphidicolin, DNA polymerase, but not DNA helicase activity, is inhibited. This allows for continued RPA loading on the unwound ssDNA and the lack of RPA displacement by the DNA polymerases, leading to RPA accumulation at DNA replication sites. (B) Live cell microscopy setup for helicase activity measurements. Cells were triple transfected as indicated 24 h before imaging. RPA accumulation at replication foci, as a proxy for helicase speed, was calculated as depicted and normalized to the respective pretreatment control. Representative spinning disk confocal time-lapse microscopy images of aphidicolin/DMSO-treated cells over 30 min, showing RPA accumulation and PCNA dissociation from replication foci and the progression of S-phase, respectively (left). Dashed boxes represent the selected magnified ROIs for line intensity profile analysis (arrows) before and 30 min after aphidicolin/DMSO addition. Line intensity plots of PCNA (magenta) and RPA (green) through a replicating chromocenter before and 30 min after aphidicolin/DMSO addition are shown. Graphs showing the normalized average RPA accumulation ( $c_v \pm$  standard deviation) over 30 min at 5 min interval imaging for aphidicolin-treated cells and for 30 min imaging for DMSO-treated cells (bottom). Aphidicolin cells were grouped into early/mid and late C2C12 S-phase cells. Slopes of the RPA accumulation curves were calculated for aphidicolin treated cells as a ratio of the rise (difference of the y-coordinates) over the run (difference of the x-coordinates) of the respective linear regression line of the average RPA accumulation curves between 0 and 20 min of drug treatment. (C) J1 wt and TKO cells were double transfected as indicated 24 h before imaging stage II S-phase cells (chromocenter replication). Imaging, analysis and representation were performed as described above; scale bars: 5  $\mu$ m (whole nuclei images) and 2.5  $\mu$ m (magnifications). *P* and *n*-values are summarized in Supplementary Table S7; \**P* < 0.05, \*\*\**P* < 0.001, n.s.: non-significant.

RPA accumulation at replication sites than mcherry transfected cells. Potential reduction of DNA helicase activity as a consequence of aphidicolin addition would be similar in both cases, thus, allowing us to conclude that the effects observed are Tet1CD activity dependent. Interestingly, increased RPA accumulation in Tet1CD-expressing cells was observed for early and mid S-phase substages; late S-phase cells, however, showed a higher accumulation after 30 min of drug treatment (Figure 6B).

As the main increase of RPA accumulation took place during the first 10 min after aphidicolin addition, we specifically analyzed the RPA accumulation during this period by imaging the triple transfected cells at a higher frame rate with one image every 10 s for 12 min. As expected, Tet1CD-transfected cells showed a rapid and steady increase of RPA accumulation, whereas a much slower RPA accumulation was observed for the control cells (Supplementary Figure S16C). No RPA accumulation could be measured for DMSO control-treated cells, although Tet1CD-transfected cells showed slightly higher RPA levels after 30 min (Figure 6B). Similar results of increased RPA accumulation in Tet1CD transfected cells were obtained when the replisome uncoupling was performed in triple transfected MEF W8 cells (Supplementary Figure S16D and E and Supplementary Table S8).

To further validate the effect of d5mC on the DNA unwinding activity, we used the reverse cellular model of loss of cytosine modification by deletion of all three DNA methyltransferases (Dnmt1, Dnmt3a and Dnmt3b, [TKO]) in ES cells or the maintenance DNA methyltransferase Dnmt1 in fibroblasts. We, therefore, co-transfected J1 wt and J1 TKO ES cells and MEF W8 (wild-type) and MEF *dnmt1*<sup>-/-</sup> (PM) fibroblasts with GFP-RPA and RFP-PCNA expression constructs to identify S-phase cells. Akin to the results we obtained with overexpression experiments, we observed in wild-type and knockout cells a time-dependent RPA accumulation, PCNA dissociation from replication foci and the lack of BrdU incorporation after aphidicolin addition. RPA accumulation was higher in J1 TKO and MEF PM cells compared to wild-type J1 and MEF W8 cells. DMSO-treated cells showed normal S-phase progression, BrdU incorporation in newly synthesized DNA as well as no significant RPA accumulation (Figure 6C and Supplementary Figure S18A–C). These results are in line with increased helicase unwinding in d5mC-deficient cells, which is associated with a decrease in helix stability and more efficient helicase unwinding.

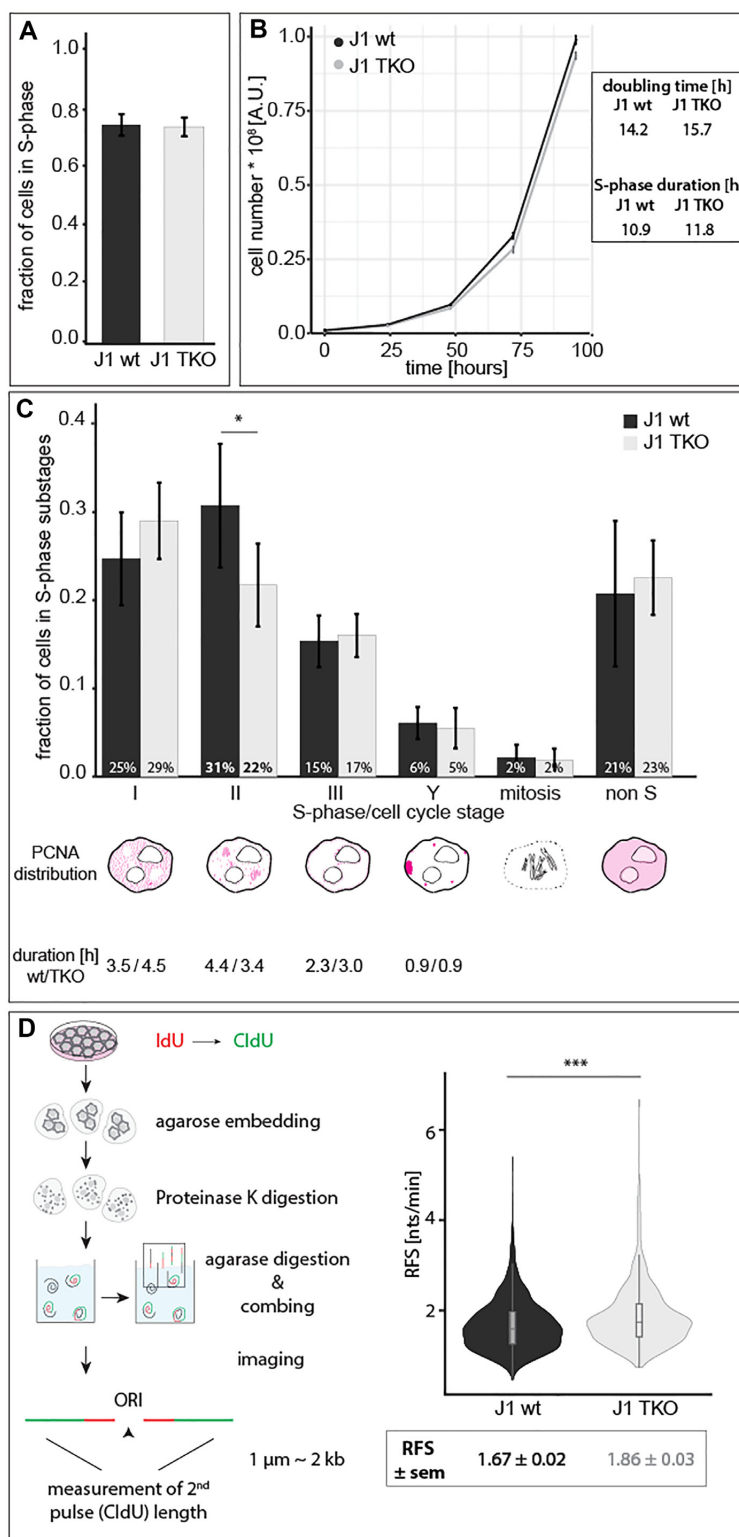
As Tet1 overexpression as well as Dnmt1 deletion could deeply impact chromatin structure, which in turn might influence genomic processes like DNA unwinding prior to DNA replication, we investigated potential changes in chromatin accessibility and nucleosome composition by micrococcal nuclease (MNase) digestion assay. We found that the ratio of monomeric nucleosomal DNA to the total DNA for each MNase concentration did not vary significantly (Supplementary Figure S19A), indicating that nucleosome positioning and number were not significantly influenced by Tet1CD overexpression. Similar results were obtained for MEF W8 and MEF PM cells (Supplementary Figure S19B).

In summary, methylated DNA is unwound by helicases slower than unmodified DNA, whereas oxidation of the methyl group reverts this effect and even enhances helicase unwinding, at least in part due to its DNA duplex destabilizing effect. Accordingly, we show that the lack of methylated cytosines and consequently the less stable DNA helix, also increases the unwinding speed of DNA helicases *in vivo*. This is in agreement with our previous results that the addition of a methyl group to cytosine bases leads to a more stable DNA helix and in turn higher energy and/or more time is needed to separate the helix in two single-stranded DNA strands.

### Loss of DNA methylation affects S-phase progression and increases replication fork speed in mouse embryonic stem cells

Since helicase-mediated DNA unwinding and DNA polymerization are the key enzymatic processes during genome duplication and in view of our results described above, we aimed to further clarify the effect of loss of DNA methylation on cell cycle and S-phase progression in general as well as DNA replication at the molecular level. Loss of all three DNA methyltransferases completely abolished d5mC levels in J1 TKO cells (Figure 2E and Supplementary Figure S9C). Doubling time and the percentage of cells present in S-phase of J1 TKO was not different compared to wild-type ES cells (Figure 7A and B; Supplementary Table S7). Importantly, loss of d5mC did not change the temporal order of DNA replication patterns of ES cells ((39) and Supplementary Figure S3). However, when we compared the frequency of the different replication substages in J1 wt and TKO cells, we observed a significant decrease of TKO cells at stage II (Figure 7C and Supplementary Table S7), indicating that duplication of normally highly methylated pericentromeric heterochromatin required less time in TKO than in wild-type cells (39). In general, faster S-phase progression can be achieved either by the activation of additional origins of replication or by increasing the replication fork speed (RFS). We, thus, measured RFS of J1 wt and J1 DNMT TKO cells by molecular combing. ES cells lacking global DNA methylation exhibited a significant increase in DNA replication fork speed (Figure 7D and Supplementary Table S7).

Previous studies in Dnmt1-deficient fibroblasts showed that the loss of DNA methylation was accompanied by an increase in histone acetylation, chromatin decondensation and reduced replication fork speed (38). In ES cells, however, no significant differences in histone acetylation or methylation were observed between wild-type and knockout cells (Supplementary Figure S20, Supplementary Table S8 and (57)), arguing against a possible effect of histone modifications on the observed differences in RFS. Furthermore, in somatic cells increased histone acetylation resulted in decreased fork speed, which is opposite to what we found here. All in all, in asynchronous J1 Dnmt TKO cell cultures, we counted fewer stage II cells (~10%) and concomitantly calculated a shorter stage II duration, indicating that Dnmt TKO cells require less time for the duplication of pericentromeric heterochromatin. Together with the difference in replication fork speed (~10%), our results suggest that the loss of DNA methylation results in faster replication fork



**Figure 7.** Effect of loss of cytosine modifications in mouse embryonic stem cells on S-phase progression and replication fork speed. **(A)** Bar plot showing the mean percentage of cells in S-phase in J1 wt and TKO ES cells. **(B)** Line plots showing the population doubling time of mouse J1 and TKO ES cells over five consecutive days with the ensuing doubling times depicted. S-phase duration was derived from the doubling times and the percentage of S-phase cells. **(C)** Bar plot showing the mean percentage of cells within the different cell cycle and S-phase substages (S-phase I - Y, G1/G2 (non S) and the ensuing S-phase substage durations). **(D)** Schematic representation of the molecular combing technique. Violin plots showing the replication fork speed (RFS) of J1 wt and TKO cells. Error bars represent the standard deviation, all boxes and whiskers are as in Figure 2 and independent experiments were performed in quadruplicates (A–C) or in duplicates (D). *P* (determined by Mann–Whitney–Wilcoxon test) and *n*-values are summarized in Supplementary Table S7; \* *P* < 0.05 and \*\*\* *P* < 0.001.

progression, specifically along DNA that is usually highly methylated under wild-type conditions.

## DISCUSSION

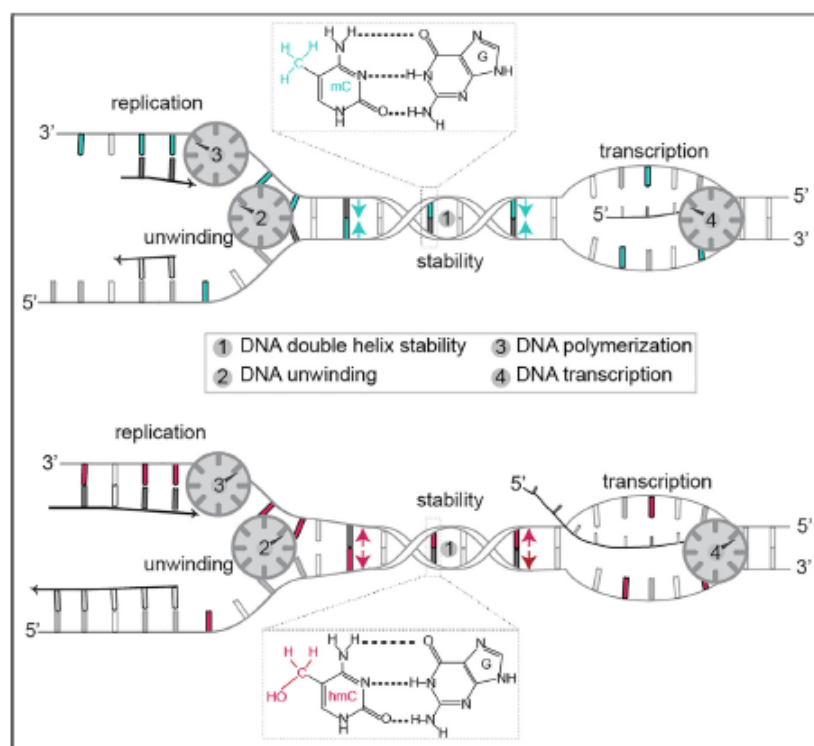
DNA base modifications greatly expand genome diversity and provide additional mechanisms to regulate gene expression. Consequently, the tight regulation of their establishment, maintenance and removal is essential for proper cell survival and development. In mammalian cells, the most extensively studied DNA modification is methylation of cytosines by Dnmts (87). More recently, the Tet enzyme family was found to oxidize d5mC (8,9) and the function and regulation thereof are still poorly understood. Whereas DNA methylation has been mainly linked to gene silencing and chromatin remodeling (88), Tet-mediated oxidation of d5mC is, on the one hand, a mechanism for active loss of DNA methylation (10) and, on the other hand, the oxidized derivatives of d5mC are reported to be stably maintained in the genome (89,90). Specific readers of all cytosine base modifications have been reported (14,91) and their influence on gene expression and chromatin regulation is the subject of extensive investigations. Whether and how cytosine base modifications can however directly influence the DNA double helix and, as a consequence, fundamental genomic processes such as DNA replication or transcription are far less understood (reviewed in (17)).

Here, we show *in vitro* and *in vivo* that the presence of methylated cytosines increases the melting temperature of the DNA double helix, demonstrating that the addition of the methyl group stabilizes the double helix structure. Incorporation of oxidized variants of d5mC, however, leads to lower melting temperatures, hinting toward a helix destabilizing effect of the hydroxy, formyl and carboxyl groups (Figure 8). Previous *in vitro* studies showed increased sequence dependent duplex stability upon cytosine methylation (20,25,50,55). Additionally, calculations of enthalpy and entropy changes upon cytosine modifications also reflected the thermal stability effects measured (92). Analysis of hydrogen bonding and base stacking interactions reinforced the idea of an increase in stability upon cytosine methylation and a destabilizing effect arising from hydroxymethyl, formyl and carboxyl groups (28,93,94). The influence on base stacking would to some extent also explain the previously observed sequence dependency. While increasing the size of the cytosine modification leads to less structural fluctuation between base pairs and, thus, a more rigid and stable structure, this can be counteracted by adding polar groups as in hydroxymethylcytosine. This is explained by the fact that the hydrophobic methyl group is less favorable for interaction with water molecules, resulting in a higher rigidity and, consequently, altogether in higher thermostability of 5mC-G base pairs, whereas the more hydrophilic and polar hydroxymethyl group has higher affinity for water molecules, leading to higher fluctuations in 5hmC-G base pairs (28). Furthermore, electron-withdrawing abilities of the formyl and carboxyl groups reduce hydrogen bonding in 5fC-G and 5caC-G base pairs, destabilizing the DNA duplex (93). These *in vitro* and *in silico* data are in line with our findings and provide a mechanistic explanation for our results.

The local changes in DNA helix structure may also represent a specific recognition platform for reader proteins. The weaker d5fC and d5caC pairing abilities could facilitate TDG base flipping during DNA demethylation (93). In addition, it was shown that DNase I cleavage was enhanced adjacent to methylated CpGs, presumably due to the minor groove narrowing effect exerted by the bulky methyl group (95). DNA shape variations arising from the presence of methylated DNA bases augment the negative electrostatic potential, which in turn can more effectively attract basic side chains of DNA-binding proteins (95–97). Interestingly, the chromatin remodeler proteins MeCP2, Mbd1 and Mbd2 identify 5mCpGs via two conserved positively charged, arginine-rich fingers (98–100), possibly exploiting changes in helix structure and stability for binding site recognition.

In our *in vivo* analyses, we made use of several different approaches to manipulate cytosine modification levels within cells (Supplementary Figure S9). To increase modified cytosine levels, on the one hand, we changed expression levels of enzymes modifying cytosines (Dnmts and Tet1; Supplementary Figure S9A and B). On the other hand, we chose a nucleotide-based approach (Supplementary Figure S9C). Both approaches were performed in undifferentiated C2C12 mouse myoblast cells, as these cells show low endogenous Tet1–3 and Mbd expression (15,86,101). The lack of the N-terminal CXXC DNA-binding motif in the Tet1CD variant allowed us to achieve global, random and unbiased methylcytosine oxidation. In contrast to Tet1 full-length (Tet1fl), Tet1CD has no preference for specific genomic regions and massive conversion of DNA methylation is observed in CpG islands (CGIs), non-CGIs, upstream, promoter, exon, intron, downstream and intergenic DNA regions (56). To study the loss of cytosine modifications on DNA structure and subsequent genomic processes, we used Dnmt1/Dnmt3a/Dnmt3b knockout mouse embryonic stem cells (57) and Dnmt1 knockout mouse embryonic fibroblasts (35), both of which are depleted of cytosine modifications (Supplementary Figure S9B).

We show that DNA helicase, DNA exonuclease and DNA-dependent DNA and RNA polymerase speed are regulated by the cytosine variants present in the DNA templates (Figure 8). In agreement with the duplex stabilizing effect of cytosine methylation observed in our *in vitro* melting experiments, helicase and polymerase speeds were decreased with increasing d5mC level in cells. Importantly, DNA replication initiation is not dependent on the presence of cytosine variants, as Dnmt knockout cells actively replicate their genome within a similar time frame to the wild-type cells. Hence, cytosine variants are unlikely to influence the amount of active replicons. In contrast, oxidation of d5mC to d5hmC (and d5fC or d5caC) resulted in helix destabilizing effects and increased *in vivo* helicase and polymerase speed/activity. Interestingly, this process is not dependent on chromatin decondensation or global nucleosomal remodeling effects. Oxidation of cytosine methylation therefore *de facto* reverts DNA duplex stability without the actual need for removal of the methyl group by DNA repair-based processes. It is thus tempting to speculate that cytosine modifications could locally regulate DNA unwinding, DNA replication and gene expression by changing DNA



**Figure 8.** Graphical summary of the effect of cytosine modifications on DNA double helix stability and metabolism. DNA methylation (blue) increases DNA duplex stability (1) and reduces in turn helicase (2) and DNA/RNA polymerase (3/4) speed. Oxidation of methylated cytosine to hydroxymethylation (red) results in decreased helix stability (1) and increased helicase (2) as well as DNA/RNA polymerase (3 and 4) speed.

(electro)mechanical properties. This is further corroborated and expanded by our replication fork speed analysis in TKO ES cells (Figure 7D). Importantly, the effect we measured was greater in normally heavily methylated genomic regions (Figures 6B and C, 7C; Supplementary Figure S17A). Since we previously mapped the exact positions of methylated CpGs in major satellite repeats and in view of the unbiased and global cytosine oxidation of Tet1CD, this observation underlines the decrease in DNA helix stability by the loss of cytosine methylation. In contrast to results from somatic cells (38), we did not observe changes in histone acetylation upon Dnmt and concomitant d5mC depletion, indicating that the increased fork speed was dependent on DNA helix stability rather than on chromatin decondensation. In line with this, it was shown that bacteriophage T7 DNA helicase unwinding rate is correlated to the local DNA sequence and unzipping force (102). Genome-wide mapping of DNA destabilizing oxidized cytosine modifications revealed enrichment of d5hmC and d5fC within gene bodies of actively transcribed genes and within CpG islands in promoters of actively transcribed genes (103,104). Cytosine methylation, however, stabilizes the DNA helix and plays a role in transcriptional repression and gene silencing (65). This is consistent with our data that oxidation of d5mC increased the transcription rate *in vitro* and *in vivo*.

We now propose an additional role of DNA base modifications in the regulation of genome processes by directly modulating DNA double helix stability to impede (cytosine methylation) or facilitate (methylation oxidation) access and unwinding of double-stranded DNA. Hence, the

variegated modifications along the DNA molecule could play a role in the specific regulations of its own local trans- actions.

#### DATA AVAILABILITY

All data and custom written scripts are available from the Omero open microscopy environment public repository (<http://cc-omero.bio.tu-darmstadt.de/webclient/userdata/?experimenter=-1>) and TUDatalib (<https://doi.org/10.48328/tudatalib-551>). All renewable biological materials will be made available upon request from the lead corresponding author M. Cristina Cardoso ([cardoso@bio.tu-darmstadt.de](mailto:cardoso@bio.tu-darmstadt.de)).

#### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

#### ACKNOWLEDGEMENTS

We are indebted to Anne Lehmkuhl, Diana Imblan and Manuela Mildner for excellent technical assistance and to Anna Höfler for helping with the MEF W8 overexpressing assays. We thank Michael Pils and Herbert Tschochner for help and discussions. We thank Heinrich Leonhardt for helpful discussions and Annina Scholl, Kathrin S. Heinz, Anne K. Ludwig and Alexander Rapp for experimental advice and comments on the manuscript.

*Author contributions:* C.R., P.Z. and M.C.C. conceived the project. C.R. generated final figures and C.R. and M.C.C.

wrote the manuscript. All authors revised, commented and agreed on the manuscript. C.R. performed and analyzed all cell-based assays and contributed to *in vitro* PCR, HRM and run-off transcription experiments. P.Z. performed and analyzed *in vitro* PCR and HRM experiments. F.H. performed and analyzed *in vitro* run-off transcription experiments. P.W. contributed to the ES cell cycle distribution, S-phase progression experiments and performed molecular combing experiments. C.C.-D. and G.C. performed and analyzed *in vitro* yeast replication assays. J.L.D. and C.E. performed and analyzed *in vitro* yeast transcription assays.

## FUNDING

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 393547839 – SFB 1361, CA 198/9-2 and CA 198/12-1 to M.C.C. and by SFB 960 (TP-A8) and the Emmy-Noether Program (DFG grant no. EN 1204/1-1) to C.E.

*Conflict of interest statement.* None declared.

## REFERENCES

- Gommers-Ampt, J.H. and Borst, P. (1995) Hypermodified bases in DNA. *FASEB J.*, **9**, 1034–1042.
- Hotchkiss, R.D. (1948) The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J. Biol. Chem.*, **175**, 315–332.
- Goll, M.G. and Bestor, T.H. (2005) Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.*, **74**, 481–514.
- Hermann, A., Gowher, H. and Jeltsch, A. (2004) Biochemistry and biology of mammalian DNA methyltransferases. *Cell. Mol. Life Sci.*, **61**, 2571–2587.
- Suzuki, M.M. and Bird, A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.*, **9**, 465–476.
- Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.*, **16**, 6–21.
- Li, E., Beard, C. and Jaenisch, R. (1993) Role for DNA methylation in genomic imprinting. *Nature*, **366**, 362–365.
- Kriaucionis, S. and Heintz, N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, **324**, 929–930.
- Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L. *et al.* (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930–935.
- Kohli, R.M. and Zhang, Y. (2013) TET enzymes, TDG and the dynamics of DNA demethylation. *Nature*, **502**, 472–479.
- Maiti, A. and Drohat, A.C. (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.*, **286**, 35334–35338.
- Müller, U., Bauer, C., Siegl, M., Rottach, A. and Leonhardt, H. (2014) TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic Acids Res.*, **42**, 8592–8604.
- Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C. and Zhang, Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, **333**, 1300–1303.
- Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W., Bauer, C., Munzel, M., Wagner, M., Müller, M., Khan, F. *et al.* (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell*, **152**, 1146–1159.
- Ludwig, A.K., Zhang, P., Hastert, F.D., Meyer, S., Rausch, C., Herce, H.D., Müller, U., Lehmkuhl, A., Hellmann, I., Trummer, C. *et al.* (2017) Binding of MBD proteins to DNA blocks Tet1 function thereby modulating transcriptional noise. *Nucleic Acids Res.*, **45**, 2438–2457.
- Zhang, P., Ludwig, A.K., Hastert, F.D., Rausch, C., Lehmkuhl, A., Hellmann, I., Smets, M., Leonhardt, H. and Cardoso, M.C. (2017) L1 retrotransposition is activated by ten-eleven-translocation protein 1 and repressed by methyl-CpG binding proteins. *Nucleus*, **8**, 548–562.
- Rausch, C., Hastert, F.D. and Cardoso, M.C. (2019) DNA modification readers and writers and their interplay. *J. Mol. Biol.*, **432**, 1731–1746.
- Raiber, E.A., Murat, P., Chirgadze, D.Y., Beraldi, D., Luisi, B.F. and Balasubramanian, S. (2015) 5-Formylcytosine alters the structure of the DNA double helix. *Nat. Struct. Mol. Biol.*, **22**, 44–49.
- Ngo, T.T., Yoo, J., Dai, Q., Zhang, Q., He, C., Aksimentiev, A. and Ha, T. (2016) Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat. Commun.*, **7**, 10813.
- Lopez, C.M., Lloyd, A.J., Leonard, K. and Wilkinson, M.J. (2012) Differential effect of three base modifications on DNA thermostability revealed by high resolution melting. *Anal. Chem.*, **84**, 7336–7342.
- Hardwick, J.S., Ptchelkine, D., El-Sagheer, A.H., Tear, I., Singleton, D., Phillips, S.E.V., Lane, A.N. and Brown, T. (2017) 5-Formylcytosine does not change the global structure of DNA. *Nat. Struct. Mol. Biol.*, **24**, 544–552.
- Renciuk, D., Blacque, O., Vorlickova, M. and Spingler, B. (2013) Crystal structures of B-DNA dodecamer containing the epigenetic modifications 5-hydroxymethylcytosine or 5-methylcytosine. *Nucleic Acids Res.*, **41**, 9891–9900.
- Acosta-Silva, C., Branchadell, V., Bertran, J. and Oliva, A. (2010) Mutual relationship between stacking and hydrogen bonding in DNA. Theoretical study of guanine-cytosine, guanine-5-methylcytosine, and their dimers. *J. Phys. Chem. B*, **114**, 10217–10227.
- Carvalho, A.T., Gouveia, L., Kanna, C.R., Warmlander, S.K., Platts, J.A. and Kamerlin, S.C. (2014) Understanding the structural and dynamic consequences of DNA epigenetic modifications: computational insights into cytosine methylation and hydroxymethylation. *Epigenetics*, **9**, 1604–1612.
- Severin, P.M., Zou, X., Gaub, H.E. and Schulten, K. (2011) Cytosine methylation alters DNA mechanical properties. *Nucleic Acids Res.*, **39**, 8740–8751.
- Severin, P.M., Zou, X., Schulten, K. and Gaub, H.E. (2013) Effects of cytosine hydroxymethylation on DNA strand separation. *Biophys. J.*, **104**, 208–215.
- Lercher, L., McDonough, M.A., El-Sagheer, A.H., Thalhammer, A., Kriaucionis, S., Brown, T. and Schofield, C.J. (2014) Structural insights into how 5-hydroxymethylation influences transcription factor binding. *Chem. Commun.*, **50**, 1794–1796.
- Wanunu, M., Cohen-Karni, D., Johnson, R.R., Fields, L., Benner, J., Peterman, N., Zheng, Y., Klein, M.L. and Drndic, M. (2011) Discrimination of methylcytosine from hydroxymethylcytosine in DNA molecules. *J. Am. Chem. Soc.*, **133**, 486–492.
- Sporbert, A., Gahl, A., Ankerhold, R., Leonhardt, H. and Cardoso, M.C. (2002) DNA polymerase clamp shows little turnover at established replication sites but sequential de novo assembly at adjacent origin clusters. *Mol. Cell*, **10**, 1355–1365.
- Sporbert, A., Domaing, P., Leonhardt, H. and Cardoso, M.C. (2005) PCNA acts as a stationary loading platform for transiently interacting Okazaki fragment maturation proteins. *Nucleic Acids Res.*, **33**, 3521–3528.
- Shcherbakova, D.M., Baloban, M., Emelyanov, A.V., Brenowitz, M., Guo, P. and Verkhusha, V.V. (2016) Bright monomeric near-infrared fluorescent proteins as tags and biosensors for multiscale imaging. *Nat. Commun.*, **7**, 12405.
- Taylor, M.T.G. and Yeeles, J.T.P. (2018) The initial response of a eukaryotic replisome to DNA damage. *Mol. Cell*, **70**, 1067–1080.
- Yaffe, D. and Saxel, O. (1977) Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature*, **270**, 725–727.
- Peters, A.H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schofer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A. *et al.* (2001) Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell*, **107**, 323–337.
- Lande-Diner, L., Zhang, J., Ben-Porath, I., Amariglio, N., Keshet, I., Hecht, M., Azuara, V., Fisher, A.G., Rechavi, G. and Cedar, H. (2007) Role of DNA methylation in stable gene repression. *J. Biol. Chem.*, **282**, 12194–12200.

24 *Nucleic Acids Research*, 2021

36. Graham, F.L., Smiley, J., Russell, W.C. and Nairn, R. (1977) Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J. Gen. Virol.*, **36**, 59–74.
37. Becker, A., Zhang, P., Allmann, L., Meilinger, D., Bertulat, B., Eck, D., Hofstaetter, M., Bartolomei, G., Hottiger, M.O., Schreiber, V. et al. (2016) Poly(ADP-ribosyl)ation of methyl CpG binding protein 2 regulates chromatin structure. *J. Biol. Chem.*, **291**, 9382.
38. Casas-Delucchi, C.S., van Bommel, J.G., Haase, S., Herce, H.D., Nowak, D., Meilinger, D., Stear, J.H., Leonhardt, H. and Cardoso, M.C. (2012) Histone hypoacetylation is required to maintain late replication timing of constitutive heterochromatin. *Nucleic Acids Res.*, **40**, 159–169.
39. Rausch, C., Weber, P., Prorok, P., Hörl, D., Maiser, A., Lehmkuhl, A., Chagin, V.O., Casas-Delucchi, C.S., Leonhardt, H. and Cardoso, M.C. (2020) Developmental differences in genome replication program and origin activation. *Nucleic Acids Res.*, **48**, 12751–122777.
40. Weber, P., Rausch, C., Scholl, A. and Cardoso, M.C. (2019) Repli-FISH (fluorescence in situ hybridization): application of 3D-(Immuno)-FISH for the study of DNA replication timing of genetic repeat elements. *OBM Genetics*, **3**, 31.
41. Zillmann, M., Zapp, M.L. and Berget, S.M. (1988) Gel electrophoretic isolation of splicing complexes containing U1 small nuclear ribonucleoprotein particles. *Mol. Cell. Biol.*, **8**, 814–821.
42. Yeeles, J.T.P., Janska, A., Early, A. and Diffley, J.F.X. (2017) How the eukaryotic replisome achieves rapid and efficient DNA replication. *Mol. Cell*, **65**, 105–116.
43. Yeeles, J.T.P., Deegan, T.D., Janska, A., Early, A. and Diffley, J.F.X. (2015) Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature*, **519**, 431–435.
44. Engel, C., Sainsbury, S., Cheung, A.C., Kostrewa, D. and Cramer, P. (2013) RNA polymerase I structure and transcription regulation. *Nature*, **502**, 650–655.
45. Pils, M. and Engel, C. (2020) Structural basis of RNA polymerase I pre-initiation complex formation and promoter melting. *Nat. Commun.*, **11**, 1206.
46. Merkl, P.E., Pils, M., Fremter, T., Schwank, K., Engel, C., Längst, G., Milkereit, P., Griesenbeck, J. and Tschochner, H. (2020) RNA polymerase I (Pol I) passage through nucleosomes depends on Pol I subunits binding its lobe structure. *J. Biol. Chem.*, **295**, 4782–4795.
47. Zhang, P., Hastert, F.D., Ludwig, A.K., Breitwieser, K., Hofstätter, M. and Cardoso, M.C. (2017) DNA base flipping analytical pipeline. *Biol. Meth. Protoc.*, **2**, bpx010.
48. Leonhardt, H., Rahn, H.P., Weinzierl, P., Sporbert, a., Cremer, T., Zink, D. and Cardoso, M.C. (2000) Dynamics of DNA replication factories in living cells. *J. Cell Biol.*, **149**, 271–280.
49. Zhang, P., Rausch, C., Hastert, F.D., Boneva, B., Filatova, A., Patil, S.J., Nuber, U.A., Gao, Y., Zhao, X. and Cardoso, M.C. (2017) Methyl-CpG binding domain protein 1 regulates localization and activity of Tet1 in a CXXC3 domain-dependent manner. *Nucleic Acids Res.*, **45**, 7118–7136.
50. Lefebvre, A., Mauffret, O., el Antri, S., Monnot, M., Lescot, E. and Fermanian, S. (1995) Sequence dependent effects of CpG cytosine methylation. A joint 1H-NMR and 31P-NMR study. *Eur. J. Biochem.*, **229**, 445–454.
51. Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C. and Zhang, Y. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, **466**, 1129–1133.
52. Sawai, S., Nagashima, J., Kuwahara, M., Kitagata, R., Tamura, T. and Matsui, I. (2007) Differences in substrate specificity of C(5)-substituted or C(5)-unsubstituted pyrimidine nucleotides by DNA polymerases from Thermophilic bacteria, archaea, and phages. *Chem. Biodiver.*, **4**, 1979–1995.
53. Hottin, A. and Marx, A. (2016) Structural insights into the processing of nucleobase-modified nucleotides by DNA polymerases. *Acc. Chem. Res.*, **49**, 418–427.
54. Howard, M.J., Foley, K.G., Shock, D.D., Batra, V.K. and Wilson, S.H. (2019) Molecular basis for the faithful replication of 5-methylcytosine and its oxidized forms by DNA polymerase beta. *J. Biol. Chem.*, **294**, 7194–7201.
55. Thalhammer, A., Hansen, A.S., El-Sagheer, A.H., Brown, T. and Schofield, C.J. (2011) Hydroxylation of methylated CpG dinucleotides reverses stabilisation of DNA duplexes by cytosine 5-methylation. *Chem. Commun.*, **47**, 5325–5327.
56. Jin, C., Lu, Y., Jelinek, J., Liang, S., Estecio, M.R.H., Barton, M.C. and Issa, J.P.J. (2014) TET1 is a maintenance DNA demethylase that prevents methylation spreading in differentiated cells. *Nucleic Acids Res.*, **42**, 6956–6971.
57. Tsumura, A., Hayakawa, T., Kumaki, Y., Takebayashi, S., Sakaue, M., Matsuoka, C., Shimotohno, K., Ishikawa, F., Li, E., Ueda, H.R. et al. (2006) Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells*, **11**, 805–814.
58. Miller, O.J., Schnedl, W., Allen, J. and Erlanger, B.F. (1974) 5-Methylcytosine localised in mammalian constitutive heterochromatin. *Nature*, **251**, 636–637.
59. Kobayakawa, S., Miike, K., Nakao, M. and Abe, K. (2007) Dynamic changes in the epigenomic state and nuclear organization of differentiating mouse embryonic stem cells. *Genes Cells*, **12**, 447–460.
60. Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F. and Leonhardt, H. (2010) Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.*, **38**, e181.
61. Curradi, M., Izzo, A., Badaracco, G. and Landsberger, N. (2002) Molecular mechanisms of gene silencing mediated by DNA methylation. *Mol. Cell. Biol.*, **22**, 3157–3173.
62. Jones, P.A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.*, **13**, 484–492.
63. Klose, R.J. and Bird, A.P. (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.*, **31**, 89–97.
64. Miranda, T.B. and Jones, P.A. (2007) DNA methylation: the nuts and bolts of repression. *J. Cell. Physiol.*, **213**, 384–390.
65. Attwood, J.T., Yung, R.L. and Richardson, B.C. (2002) DNA methylation and the regulation of gene transcription. *Cell. Mol. Life Sci.*, **59**, 241–257.
66. Ebright, R.H. (2000) RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J. Mol. Biol.*, **304**, 687–698.
67. Heiss, F.B., Daiß, J.L., Becker, P. and Engel, C. (2021) Conserved strategies of RNA polymerase I hibernation and activation. *Nat. Commun.*, **12**, 758.
68. Kellinger, M.W., Song, C., Chong, J., Lu, X., He, C. and Wang, D. (2012) 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.*, **19**, 831–833.
69. You, C., Ji, D., Dai, X. and Wang, Y. (2014) Effects of Tet-mediated oxidation products of 5-methylcytosine on DNA transcription in vitro and in mammalian cells. *Sci. Rep.*, **4**, 7052.
70. Chen, Z.J. and Pikaard, C.S. (1997) Epigenetic silencing of RNA polymerase I transcription: a role for DNA methylation and histone modification in nucleolar dominance. *Genes Dev.*, **11**, 2124–2136.
71. Lisica, A., Engel, C., Jahnel, M., Roldán, E., Galburt, E.A., Cramer, P. and Grill, S.W. (2016) Mechanisms of backtrack recovery by RNA polymerases I and II. *Proc. Natl Acad. Sci.*, **113**, 2946–2951.
72. Zhu, H., Wang, G. and Qian, J. (2016) Transcription factors as readers and effectors of DNA methylation. *Nat. Rev. Genet.*, **17**, 551–565.
73. Bernad, A., Zaballos, A., Salas, M. and Blanco, L. (1987) Structural and functional relationships between prokaryotic and eukaryotic DNA polymerases. *EMBO J.*, **6**, 4219–4225.
74. Hubscher, U. (1983) DNA polymerases in prokaryotes and eukaryotes: mode of action and biological implications. *Experientia*, **39**, 1–25.
75. Huber, C., Von Watzdorf, J. and Marx, A. (2016) 5-Methylcytosine-sensitive variants of *Thermococcus kodakaraensis* DNA polymerase. *Nucleic Acids Res.*, **44**, 9881–9890.
76. Von Watzdorf, J., Leitner, K. and Marx, A. (2016) Modified nucleotides for discrimination between cytosine and the epigenetic marker 5-methylcytosine. *Angew. Chem.*, **55**, 3229–3232.
77. Howard, M.J., Foley, K.G., Shock, D.D., Batra, V.K. and Wilson, S.H. (2019) Molecular basis for the faithful replication of 5-methylcytosine and its oxidized forms by DNA polymerase beta. *J. Biol. Chem.*, **294**, 7194–7201.
78. Shibutani, T., Ito, S., Toda, M., Kanao, R., Collins, L.B., Shibata, M., Urabe, M., Koseki, H., Masuda, Y., Swenberg, J.A. et al. (2014) Guanine-5-carboxylcytosine base pairs mimic mismatches during DNA replication. *Sci. Rep.*, **4**, 5220.

79. Burgers, P.M.J. and Kunkel, T.A. (2017) Eukaryotic DNA replication fork. *Annu. Rev. Biochem.*, **86**, 417–438.
80. Chagin, V.O., Stear, J.H. and Cardoso, M.C. (2010) Organization of DNA replication. *Cold Spring Harb. Perspect. Biol.*, **2**, a000737.
81. Gorisch, S.M., Sporb, A., Stear, J.H., Grunewald, I., Nowak, D., Warbrick, E., Leonhardt, H. and Cardoso, M.C. (2008) Uncoupling the replication machinery: replication fork progression in the absence of processive DNA synthesis. *Cell Cycle*, **7**, 1983–1990.
82. Baranovskiy, A.G., Babayeva, N.D., Suwa, Y., Gu, J., Pavlov, Y.I. and Tahirov, T.H. (2014) Structural basis for inhibition of DNA replication by aphidicolin. *Nucleic Acids Res.*, **42**, 14013–14021.
83. Bambara, R.A. and Jessee, C.B. (1991) Properties of DNA polymerases delta and epsilon, and their roles in eukaryotic DNA replication. *Biochim. Biophys. Acta*, **1088**, 11–24.
84. Byrnes, J.J. (1984) Structural and functional properties of DNA polymerase delta from rabbit bone marrow. *Mol. Cell. Biochem.*, **62**, 13–24.
85. Lee, M.Y., Tan, C.K., Downey, K.M. and So, A.G. (1981) Structural and functional properties of calf thymus DNA polymerase delta. *Prog. Nucleic Acid Res. Mol. Biol.*, **26**, 83–96.
86. Zhong, X., Wang, Q.Q., Li, J.W., Zhang, Y.M., An, X.R. and Hou, J. (2017) Ten-Eleven Translocation-2 (Tet2) is involved in myogenic differentiation of skeletal myoblast cells in vitro. *Sci. Rep.*, **7**, 43539.
87. Denis, H., Ndlovu, M.N. and Fuks, F. (2011) Regulation of mammalian DNA methyltransferases: a route to new mechanisms. *EMBO Rep.*, **12**, 647–656.
88. Clouaire, T. and Stancheva, I. (2008) Methyl-CpG binding proteins: specialized transcriptional repressors or structural components of chromatin? *Cell. Mol. Life Sci.*, **65**, 1509–1522.
89. Bachman, M., Uribe-Lewis, S., Yang, X., Burgess, H.E., Iurlaro, M., Reik, W., Murrell, A. and Balasubramanian, S. (2015) 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.*, **11**, 555–557.
90. Bachman, M., Uribe-Lewis, S., Yang, X., Williams, M., Murrell, A. and Balasubramanian, S. (2014) 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat. Chem.*, **6**, 1049–1055.
91. Iurlaro, M., Ficiz, G., Oxley, D., Raiber, E.A., Bachman, M., Booth, M.J., Andrews, S., Balasubramanian, S. and Reik, W. (2013) A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.*, **14**, R119.
92. Wright, E.P., Abdelhamid, M.A.S., Ehiabor, M.O., Grigg, M.C., Irving, K., Smith, N.M. and Waller, Z.A.E. (2020) Epigenetic modification of cytosines fine tunes the stability of i-motif DNA. *Nucleic Acids Res.*, **48**, 55–62.
93. Dai, Q., Sanstead, P.J., Peng, C.S., Han, D., He, C. and Tokmakoff, A. (2016) Weakened N3 hydrogen bonding by 5-formylcytosine and 5-carboxylcytosine reduces their base-pairing stability. *ACS Chem. Biol.*, **11**, 470–477.
94. Song, Q., Qiu, Z., Wang, H., Xia, Y., Shen, J. and Zhang, Y. (2013) The effect of methylation on the hydrogen-bonding and stacking interaction of nucleic acid bases. *Struct. Chem.*, **24**, 55–65.
95. Lazarovici, A., Zhou, T., Shafer, A., Dantas Machado, A.C., Riley, T.R., Sandstrom, R., Sabo, P.J., Lu, Y., Rohs, R., Stamatoyannopoulos, J.A. et al. (2013) Probing DNA shape and methylation state on a genomic scale with DNase I. *Proc. Natl Acad. Sci.*, **110**, 6376–6381.
96. Rohs, R., West, S.M., Sosinsky, A., Liu, P., Mann, R.S. and Honig, B. (2009) The role of DNA shape in protein-DNA recognition. *Nature*, **461**, 1248–1253.
97. Dantas Machado, A.C., Zhou, T., Rao, S., Goel, P., Rastogi, C., Lazarovici, A., Bussemaker, H.J. and Rohs, R. (2015) Evolving insights on how cytosine methylation affects protein-DNA binding. *Brief. Funct. Genom.*, **14**, 61–73.
98. Ohki, I., Shimotake, N., Fujita, N., Jee, J., Ikegami, T., Nakao, M. and Shirakawa, M. (2001) Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell*, **105**, 487–497.
99. Liu, K., Xu, C., Lei, M., Yang, A., Loppnau, P., Hughes, T.R. and Min, J. (2018) Structural basis for the ability of MBD domains to bind methyl-CG and TG sites in DNA. *J. Biol. Chem.*, **293**, 7344–7354.
100. Ho, K.L., McNae, I.W., Schmiedeberg, L., Klose, R.J., Bird, A.P. and Walkinshaw, M.D. (2008) MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Mol. Cell*, **29**, 525–531.
101. Brero, A., Easwaran, H.P., Nowak, D., Grunewald, I., Cremer, T., Leonhardt, H. and Cardoso, M.C. (2005) Methyl CpG-binding proteins induce large-scale chromatin reorganization during terminal differentiation. *J. Cell Biol.*, **169**, 733–743.
102. Johnson, D.S., Bai, L., Smith, B.Y., Patel, S.S. and Wang, M.D. (2007) Single-molecule studies reveal dynamics of DNA unwinding by the ring-shaped T7 helicase. *Cell*, **129**, 1299–1309.
103. Raiber, E.A., Beraldi, D., Ficiz, G., Burgess, H.E., Branco, M.R., Murat, P., Oxley, D., Booth, M.J., Reik, W. and Balasubramanian, S. (2012) Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. *Genome Biol.*, **13**, R69.
104. Wu, H., D'Alessio, A.C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y.E. and Zhang, Y. (2011) Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev.*, **25**, 679–684.

Rausch et al.

## **Cytosine base modifications regulate DNA duplex stability and metabolism**

Cathia Rausch<sup>1</sup>, Peng Zhang<sup>1,2,\*</sup>, Corella S. Casas-Delucchi<sup>3,#</sup>, Julia L. Daiß<sup>4,#</sup>, Christoph Engel<sup>4</sup>, Gideon Coster<sup>3</sup>, Florian D. Hastert<sup>1</sup>, Patrick Weber<sup>1</sup> and M. Cristina Cardoso<sup>1,\*</sup>

<sup>1</sup> Cell Biology and Epigenetics, Department of Biology, Technical University of Darmstadt, 64287 Darmstadt, Germany.

<sup>2</sup> Center for Tissue Engineering and Stem Cell Research, Guizhou Medical University, Guiyang, Guizhou, China, 550004.

<sup>3</sup> The Institute of Cancer Research, Chester Beatty Laboratories, London SW3 6JB, United Kingdom.

<sup>4</sup> Regensburg Center for Biochemistry, University of Regensburg, 93053 Regensburg, Germany.

# These authors contributed equally to this work.

\* To whom correspondence should be addressed: M. Cristina Cardoso: Tel: +49 6151 16 21882; Fax: +49 6151 16 21880; Email: [cardoso@bio.tu-darmstadt.de](mailto:cardoso@bio.tu-darmstadt.de).  
Correspondence may also be addressed to: Peng Zhang: Tel/Fax: +86 85180988028; Email: [peng12zhang@outlook.com](mailto:peng12zhang@outlook.com).

Present address: Florian D. Hastert, Paul-Ehrlich-Institut, Bundesinstitut für Impfstoffe und biomedizinische Arzneimittel, 63225 Langen, Germany.

Rausch et al.

## SUPPLEMENTARY TABLES, FIGURES AND FIGURE LEGENDS

Supplementary Table 1: Oligonucleotide characteristics.

Name	Sequence [5' - 3']	Application	Reference
Agel-miRFP-fw	ATC CAC CGG TCG CCA CCA TGG TA	cloning	this study
BsrGI-miRFP-rev	AAA TGT ACA GGC TCT CAA GCG CGG TGA T	cloning	this study
MaSat-fw	AAA ATG AGA AAC ATC CAC TTG	FISH probe generation	(1)
MaSat-rev	CCA TGA TTT TCA GTT TTC TT	FISH probe generation	(1)
CG-up	ACT ACC ATC XGG ACC AGA AG*	HRM	(2)
CG-down	CTT CTG GTC XGG ATG GTA GT*	HRM	(2)
MINX-fw	CGG TAC CTA ATA CGA CTC ACT ATA GGG AGA	PCR fragment, replication & IVT template generation	(2)
MINX-5'P-fw	5'Phospho-CGG TAC CTA ATA CGA CTC ACT ATA	5' phosphorylated IVT template generation	this study
MINX-rev	GTG CCA AGC TTG CAT GC	PCR fragment & IVT template generation	(2)
LINE1-fw	ATC CCA CAC CTG GCT CAG AGG G	PCR fragment generation	(3)
LINE1-rev	GTC AGG GGT CAG GGA CCC ACT T	PCR fragment generation	(3)
GC497	GAT CCT TAA TTA ACC TCA GCT TGA CCA TGA CTC GAC TGC AAT CGC CCT CAG CGC GGC CGC CTG CA	cloning	this study
GC498	GGC GGC CGC GCT GAG GGC GAT TGC AGT CGA GTC ATG GTC AAG CTG AGG TTA ATT AAG	cloning	this study
pGC504-3kb-fw	TTC TAA ACC ATT GCC GCT TAC TC	yeast replication template generation	this study
pGC504-3kb-rev	CCG TAT CGT AGT TAT CTA CAC GAC	yeast replication template generation	this study
MINX-400bp-tailed-fw	ATC ACC TTA CCC TAT ACT TAC TCG CAT TCC CGG TAC CTA ATA CGA CTC ACT ATA	yeast IVT template generation	this study
MINX-400bp-tailed-rev	GTG CCA AGC TTG CAT GCC TGC	yeast IVT template generation	this study
Competitor oligonucleotide	CGA GTA AGT ATA GGG TAA GGT GAT	yeast IVT template generation	(4)

\* X: cytosine, 5-methyl, 5-hydroxymethyl, 5-formyl or 5-carboxyl cytosine.

Rausch et al.

**Supplementary Table 2: Plasmid characteristics.**

Name	pc number*	Fluorophore	Gene species	Promoter	Reference
mch-mTet1CD	2547	mcherry	Mus musculus	CAG	(5)
mcherry	2387	mcherry	Discosoma sp.	CMV	(5)
mRFP-hPCNA	1054	mRFP	Homo sapiens	CMV	(6)
miRFP-hPCNA	3385	miRFP670	Homo sapiens	CMV	this study
GFP-hRPA34	624	GFP	Homo sapiens	CMV	(7)
pUC18-MINX-M3	3902	-	-	T7	(8)
pGC504 ARS306	4625	-	Saccharomyces cerevisiae	-	this study

\* pc number: plasmid collection number.

**Supplementary Table 3: Cell line characteristics.**

Name	Species	Type	Genotype	Reference
C2C12	Mus musculus	myoblast	wildtype	(9)
MEF W8	Mus musculus	embryonic fibroblast	wildtype	(10)
MEF PM	Mus musculus	embryonic fibroblast	p53 <sup>-/-</sup> & Dnmt1 <sup>-/-</sup>	(11)
HEK 293	Homo sapiens	embryonic kidney	wildtype	(12)
J1 wt	Mus musculus	embryonic stem cell	wildtype	(13)
J1 TKO	Mus musculus	embryonic stem cell	Dnmt1 <sup>-/-</sup> & Dnmt3a <sup>-/-</sup> & Dnmt3b <sup>-/-</sup>	(14)

Rausch et al.

Supplementary Table 4: Nucleotide and nucleoside characteristics.

Name	Application	Cat #	Company
Biotin-16-dUTP	labeling of FISH probes	-	selfmade (1)
EdU (5-ethynyl-2'-deoxyuridine)	labeling of nascent DNA	E10415	Thermo Fisher Scientific, Waltham, MA, USA
BrdU (5-bromo-2'-deoxyuridine)	labeling of nascent DNA	B5002	Sigma-Aldrich, St Louis, MO, USA
Thymidine	labeling of nascent DNA (during chase time)	T9250	Sigma-Aldrich, St Louis, MO, USA
EU* (5-ethynyl-uridine)	labeling of nascent RNA	E10345	Thermo Fisher Scientific, Waltham, MA, USA
dATP, dTTP & dGTP	PCR & <i>in vitro</i> replication	10297018	Thermo Fisher Scientific, Waltham, MA, USA
dCTP	PCR, manipulation of C modification level (transfection) & <i>in vitro</i> replication	10297018	Thermo Fisher Scientific, Waltham, MA, USA
d5mCTP	PCR & manipulation of C modification level (transfection)	NU-1125	Jena Bioscience, Jena, Germany
d5hmCTP	PCR & manipulation of C modification level (transfection)	NU-932	Jena Bioscience, Jena, Germany
d5fCTP	PCR	N-2064	Trilink, San Diego, CA, USA
d5caCTP	PCR	N-2063	Trilink, San Diego, CA, USA
$\alpha^{33}\text{P}$ -dATP, 3000 Ci/mmol, 20 mCi/ml	<i>in vitro</i> yeast replication assay	SRF-203H	Hartmann Analytic, Braunschweig, Germany
Cy3-dUTP	labeling of nascent DNA (transfection)	PA53022	GE Healthcare, Chicago, IL, USA
ATP, UTP, CTP & GTP*	<i>in vitro</i> T7 run-off transcription & <i>in vitro</i> yeast replication	18109-017	Thermo Fisher Scientific, Waltham, MA, USA
$\alpha^{32}\text{P}$ -CTP*, 400 Ci/mmol, 10 mCi/ml	<i>in vitro</i> transcription assay	SRP-109	Hartmann Analytic, Braunschweig, Germany
5-iodo-2'-deoxyuridine (IdU)	molecular DNA combing	I7125	Sigma-Aldrich, St Louis, MO, USA
5-chloro-2'-deoxyuridine (CldU)	molecular DNA combing	C6891	Sigma-Aldrich, St Louis, MO, USA

\* ribonucleotides for RNA synthesis.

Rausch et al.

**Supplementary Table 5: Primary and secondary antibody characteristics.**

Reactivity	Host	Dilution	Application	Cat #	Company/Reference
anti-PCNA (PC-10)	mouse	1:100/1:1000	IF*/WB**	M0879	Dako, Hamburg, Germany
anti-5mC (33D3)	mouse	1:150/1:1000	IF/Slot Blot	39649	Active Motif, La Hulpe, Belgium
anti-5hmC	rabbit	1:250/1:5000	IF/Slot Blot	39769	Active Motif, La Hulpe, Belgium
anti-BrdU	rabbit	1:500	IF	600-401-C29	Rockland Immunochemicals Inc., Limerick, PA, USA
anti-H3 acetyl	rabbit	1:200	IF	06-599	Upstate, Lake Placid, NY, USA
anti-H3K9ac (1B10)	mouse	1:500	IF	61251	Active Motif, La Hulpe, Belgium
anti-H3K18ac (EP959Y)	rabbit	1:200	IF	ab40888	Abcam, Cambridge, UK
anti-H3K27ac	rabbit	1:200	IF	4353	Cell Signaling Technology, Danvers, MA, USA
anti-H3K56ac (EPR996Y)	rabbit	1:100	IF	ab76307	Abcam, Cambridge, UK
anti-H4K8ac	rabbit	1:100	IF	06-760	Upstate, Lake Placid, NY, USA
anti-H4K16ac	rabbit	1:200	IF	39167	Active Motif, La Hulpe, Belgium
anti-H3K9m3	rabbit	1:200	IF	39161	Active Motif, La Hulpe, Belgium
anti-H4K20m3	rabbit	1:500	IF	ab9053	Abcam, Cambridge, UK
anti-RFP (5F8)	rat	1:10	IF/FISH	-	(15)
anti-5fC	rabbit	1:2000	Slot Blot	61223	Active Motif, La Hulpe, Belgium
anti-5caC	rabbit	1:2000	Slot Blot	61225	Active Motif, La Hulpe, Belgium
anti-BrdU/IdU (B44)	mouse	1:200	molecular combing IF	347580	Becton Dickinson, Franklin Lakes, NJ, USA
anti-BrdU/CldU (BU1/75 (ICR1))	rat	1:200	molecular combing IF	OBT003 0CX	Biorad, Puchheim, Germany
anti-ssDNA (16-19, IgG2a)	mouse	1:200	molecular combing IF	MAB303 4	Millipore, Burlington, MA, USA
anti-mouse IgG Chromeo 488	goat	1:1000	IF (fluorescent secondary)	15031	Active Motif, La Hulpe, Belgium
anti-rabbit IgG AMCA	donkey	1:00	IF (fluorescent secondary)	715-155-151	The Jackson Laboratory, Bar Harbor, ME, USA
anti-rabbit IgG Chromeo 488	goat	1:1000	IF (fluorescent secondary)	15041	Active Motif, La Hulpe, Belgium
anti-mouse IgG Cy3	donkey	1:300	IF (fluorescent secondary)	715-165-151	The Jackson Laboratory, Bar Harbor, ME, USA
anti-rat IgG Cy3	donkey	1:300	IF/FISH (fluorescent secondary)	712-165-153	The Jackson Laboratory, Bar Harbor, ME, USA
anti-rabbit IgG Cy3	donkey	1:300	IF (fluorescent secondary)	711-165-152	The Jackson Laboratory, Bar Harbor, ME, USA
anti-rabbit IgG Cy5	donkey	1:300	IF (fluorescent secondary)	711-175-152	The Jackson Laboratory, Bar Harbor, ME, USA
anti-mouse IgG HRP	sheep	1:5000	Slot blot (HRP*** conj. secondary)	NA 931V	GE Healthcare, Chicago, IL, USA
anti-rabbit IgG HRP	goat	1:5000	Slot blot (HRP conj. secondary)	A0545	Sigma-Aldrich, St Louis, MO, USA
anti-mouse IgG Chromeo 546	goat	1:200	molecular combing IF (fluorescent secondary)	15033	Active Motif, La Hulpe, Belgium
anti-rat IgG AlexaFluor 488	donkey	1:200	molecular combing IF (fluorescent secondary)	712-545-153	The Jackson Laboratory, Bar Harbor, ME, USA
anti-mouse IgG2a AlexaFluor 647	goat	1:200	molecular combing IF (fluorescent secondary)	A-21241	Fisher Scientific GmbH, Hampton, NH, USA

\* IF: immunofluorescence, \*\* WB: western blot; \*\*\* HRP: horseradish peroxidase.

Rausch et al.

Supplementary Table 6: Imaging system characteristics.

Microscope /Company	Lasers/lamps	Filters (ex. & em. [nm])*	Objectives/lenses	Detection system	Incubation system	Application
Ultra-View VoX spinning disk on an inverted Nikon Ti-E microscope/PerkinElmer Life Sciences, UK	solid state diode lasers (405 nm, 488 nm, 561 nm, 640 nm)	405/488/568/640** 405: 415–475 488: 505–549 561: 580–650 640: 664–754	oil immersion 60x Plan-Apochromat (NA 1.45)	cooled 14-bit Hamamatsu® C9100-50 EMCCD	closed live-cell microscopy chamber (ACU control, Olympus) for time-lapse microscopy	time-lapse microscopy & confocal z-stack imaging
Operetta high content screening microscopy/PerkinElmer Life Sciences, UK	Xenon fiber-optic light source, 300 W, 360 – 640 nm continuous spectrum	405: 360-400 & 410–480 488: 460-490 & 500–550 561: 560-580 & 590–640	20x or 40x air (0.45 NA and 0.95 NA) long WD***	14-bit Jenoptik CMOS	-	high content screening microscopy
Widefield Axiovert 200 /Zeiss, Germany	HBO100 mercury lamp	488: 473-491 & 506-534 561: 550-580 & 590-650 640: 590-650 & 663-738	oil immersion 63x Plan-Apochromat (1.4 NA)	12-bit AxioCam mRM	-	molecular combing imaging
Amersham AI600 imager/GE Healthcare, Chicago, IL, USA	UV transillumination light: 312 nm	EtBr: 312 & 585-625	large aperture f/0.85 FUJINON™	16-bit Peltier cooled Fujifilm Super CCD	-	EtBr stained gel & HRP stained blot imaging
Typhoon FLA-9500/GE Healthcare, Chicago, IL, USA	635 nm (red LD laser)	IP BP390	-	bi-alkali photomultiplier tubes (PMT), 16-bit digitization	-	radioactive gel imaging
Personal Molecular Imager™ System/ Biorad, Hercules, CA, USA	635 nm	BP390	-	photomultiplier tubes (PMT), 16-bit digitization	-	radioactive gel imaging

\*ex.: excitation &amp; em.: emission, \*\* dichroic specification, \*\*\* WD: working distance.

Rausch et al.

Supplementary Table 7: Plot statistics (main figures).

Figure	Sample	n <sup>a</sup>	Median	Mean	StDev <sup>a</sup>	95% CI <sup>a</sup>	p-value <sup>a</sup>
2A	Tet1CD 1	3318	0.93	1.04	0.51	1.04 - 1.06	5.02E-05
	Tet1CD 2	1142	1.33	1.92	1.47	1.83 - 2.01	< 2.2E-16
	Tet1CD 3	128	2.61	3.21	1.3	2.98 - 3.44	< 2.2E-16
	mcherry 1	3292	1	1.1	0.54	1.08 - 1.12	0.652
	mcherry 2	2809	1.04	1.28	0.97	1.24 - 1.32	0.06399
	mcherry 3	295	0.98	1.17	0.99	1.06 - 1.28	0.6548
2B	untransf.	4460	1	1.07	0.66	1.05 - 1.09	-
	dCTP 5mC	2710	1	1.19	0.35	0.98 - 1.1	-
	d5mCTP 5mC	2274	3.3	3.13	0.41	2.98 - 3.01	0.045
	dCTP 5hmC	2564	1	1.08	0.31	0.97 - 1.0	-
	d5hmCTP 5hmC	2283	3.09	3.25	0.5	3.0 - 3.2	0.039
	dCTP 5mC	2710	1	1.06	0.29	0.99 - 1.1	-
	ddH <sub>2</sub> O 5mC	2566	0.99	1.04	0.38	0.9 - 1.1	0.351
dCTP 5hmC	2564	1	0.98	0.48	0.99 - 1.12	-	
ddH <sub>2</sub> O 5hmC	2081	1.01	0.99	0.42	0.95 - 1	0.471	
2C	dCTP 5mC	1935	1	1.08	0.24	0.98 - 1.02	-
	d5mCTP 5mC	2051	3.43	3.42	0.39	3.39 - 3.43	0.035
	dCTP 5hmC	2146	1	1.2	0.21	1.01 - 1.21	-
	d5hmCTP 5hmC	2314	3.17	3.09	0.41	3.3 - 3.41	0.014
3A	Tet1CD 1	300	1.47	1.47	0.05	1.47 - 1.48	3.3E-06
	Tet1CD 2	1477	1.81	2.02	1.53	1.94 - 2.09	< 2.2E-16
	Tet1CD 3	809	1.96	2.28	1.4	2.18 - 2.38	< 2.2E-16
	mcherry 1	190	1	1.11	0.45	1.04 - 1.17	-
	mcherry 2	3134	1.04	1.29	4.24	1.15 - 1.44	0.251
	mcherry 3	253	1.08	1.39	1.39	1.33 - 1.44	0.0857
3B	dCTP 65°C	4783	1	1.08	0.52	1.07 - 1.1	-
	d5mCTP 65°C	4443	0.74	0.94	0.7	0.92 - 0.96	0.000214
	d5hmCTP 65°C	4401	1.18	1.24	0.4	1.23 - 1.26	< 2.2E-16
	dCTP 70°C	4329	1	1.11	0.72	1.09 - 1.13	-
	d5mCTP 70°C	4140	0.62	0.96	1.32	0.92 - 1	< 2.2E-16
	d5hmCTP 70°C	3982	1.49	1.61	0.81	1.59 - 1.64	< 2.2E-16
3C	W8 60°C	2827	1	1.08	0.32	1.07 - 1.09	-
	W8 65°C	3150	1.52	1.33	0.66	1.31 - 1.35	-
	W8 70°C	2846	2.23	3.06	1.8	2.99 - 3.13	-
	PM 60°C	3102	2.57	3.06	1.72	1.08 - 1.1	8.70E-06
	PM 65°C	4572	6.93	2.55	1.02	2.52 - 2.58	< 2.2E-16
	PM 70°C	4514	2.22	7.69	2.28	7.62 - 7.76	< 2.2E-16
3D	wt 60°C	27	1	1.12	1.2	1.06 - 1.08	-
	TKO 60°C	36	1.35	1.28	0.9	1.11 - 1.15	0.06524
	wt 65°C	22	1.44	1.43	0.8	0.87 - 0.98	-
	TKO 65°C	26	2.05	1.52	1.31	1.2 - 1.35	0.0000268
	wt 70°C	23	1.55	2.08	1.4	1.26 - 1.47	-
	TKO 70°C	22	2.59	2.19	1.25	1.24 - 1.31	< 2.2E-16
4B	Tet1CD 25 min 1	2768	1.08	1.26	0.59	1.24 - 1.28	< 2.2E-16
	Tet1CD 25 min 2	1153	1.31	1.49	0.69	1.45 - 1.53	< 2.2E-16
	Tet1CD 25 min 3	293	1.34	1.51	0.66	1.43 - 1.58	< 2.2E-16
	mcherry 25 min 1	2283	1	1.09	0.38	1.07 - 1.1	-
	mcherry 25 min 2	3687	1.06	1.19	0.53	1.17 - 1.21	0.354
	mcherry 25 min 3	1500	1.08	1.24	0.8	1.2 - 1.28	0.0215
	Tet1CD 45 min 1	2680	1.47	1.68	0.75	1.65 - 1.7	< 2.2E-16
	Tet1CD 45 min 2	751	1.73	1.99	1.09	1.91 - 2.06	< 2.2E-16
	Tet1CD 45 min 3	147	1.68	1.95	0.86	1.81 - 2.09	1.32E-8
	mcherry 45 min 1	2553	1.39	1.52	0.62	1.5 - 1.54	-
	mcherry 45 min 2	2784	1.5	1.7	0.89	1.67 - 1.74	5.20E-09
	mcherry 45 min 3	1314	1.54	1.73	0.84	1.71 - 1.8	0.0479
	Tet1CD 60 min 1	2293	1.8	1.75	1.33	2.02 - 2.1	3.05E-11
	Tet1CD 60 min 2	1343	2.24	2.06	1.68	2.49 - 2.64	< 2.2E-16
	Tet1CD 60 min 3	257	2.24	2.62	0.75	2.472 - 2.83	9.79E-10
	mcherry 60 min 1	2835	1.8	1.96	0.86	1.93 - 1.99	-
	mcherry 60 min 2	2795	1.94	2.14	0.87	2.11 - 2.18	0.145
	mcherry 60 min 3	1635	1.96	2.17	0.85	2.13 - 2.21	0.0568

Rausch et al.

Figure	Sample	n <sup>a</sup>	Median	Mean	StDev <sup>a</sup>	95% CI <sup>a</sup>	p-value <sup>a</sup>
	Tet1CD 120 min 1	2007	2.59	2.93	1.43	2.87 - 2.99	< 2.2E-16
	Tet1CD 120 min 2	1540	3.14	3.42	1.51	3.35 - 3.49	< 2.2E-16
	Tet1CD 120 min 3	348	3.15	3.41	1.52	3.25 - 3.57	< 2.2E-16
	mcherry 120 min 1	2527	2.35	2.58	1.02	2.54 - 2.62	-
	mcherry 120 min 2	2974	2.57	2.82	1.14	2.78 - 2.86	0.00743
	mcherry 120 min 3	1682	2.56	2.85	1.15	2.79 - 2.9	0.000485
4C	dCTP	3804	1	1.05	0.84	1.02 - 1.18	-
	d5mCTP	2864	0.41	0.39	0.99	0.35 - 0.48	0.0032
	dCTP	8619	1	1.2	0.54	0.9 - 1.31	-
	d5hmCTP	9835	2.73	2.83	1.12	2.54 - 292	0.00042
5A <sup>b</sup>	dC 0 U	2	-	1	-	-	-
	dC 1 U	2	-	1.36	-	-	-
	d5mC 0 U	2	-	1	-	-	-
	d5mC 1 U	2	-	1.18	-	-	0.00195
	d5hmC 0 U	2	-	1	-	-	-
	d5hmC 1 U	2	-	0.95	-	-	0.16
5B	Tet1CD 10 min 1	4125	1.26	1.47	0.71	1.45 - 1.49	< 2.2E-16
	Tet1CD 10 min 2	808	1.5	1.81	1.25	1.72 - 1.9	< 2.2E-16
	Tet1CD 10 min 3	650	1.64	1.97	1.21	1.88 - 2.06	3.82E-06
	mcherry 10 min 1	1950	1	1.21	0.58	1.18 - 1.24	-
	mcherry 10 min 2	798	1.21	1.48	0.8	1.42 - 1.53	0.022
	mcherry 10 min 3	127	1.15	1.4	0.7	1.28 - 1.52	0.323
	Tet1CD 20 min 1	1258	2.27	2.46	1.25	2.39 - 2.53	< 2.2E-16
	Tet1CD 20 min 2	1444	2.4	2.75	1.62	2.67 - 2.83	< 2.2E-16
	Tet1CD 20 min 3	184	2.25	2.45	1.37	2.25 - 2.65	0.002118
	mcherry 20 min 1	948	1.95	2.13	1.09	2.06 - 2.2	-
	mcherry 20 min 2	441	1.94	2.41	1.49	2.27 - 2.55	0.0004385
	mcherry 20 min 3	850	1.98	2.33	1.25	2.25 - 2.41	0.1832
	Tet1CD 30 min 1	1308	3.61	3.77	1.96	3.66 - 3.88	< 2.2E-16
	Tet1CD 30 min 2	1479	3.65	3.86	2.31	3.74 - 3.98	2.65E-15
	Tet1CD 30 min 3	155	3.58	3.81	2.36	3.44 - 4.18	0.0004385
mcherry 30 min 1	1216	3.32	3.35	2.01	3.24 - 3.46	-	
mcherry 30 min 2	3657	3.47	3.98	2.75	3.89 - 4.07	0.0236	
mcherry 30 min 3	578	3.55	4.08	2.92	3.84 - 4.32	5.32E-09	
5C	dCTP	8285	1	1.05	0.18	0.98 - 1.1	-
	d5mCTP	4861	0.45	0.38	0.36	0.35 - 0.48	0.024
	dCTP	9921	1	1.1	0.25	1 - 1.02	-
	d5hmCTP	4727	1.37	1.42	0.56	1.35 - 1.59	0.039
5D <sup>b</sup>	dC 0 U	2	-	1	-	-	-
	dC 1 U	2	-	0.182	-	-	-
	d5mC 0 U	2	-	1	-	-	-
	d5mC 1 U	2	-	0.65	-	-	0.315
	d5hmC 0 U	2	-	1	-	-	-
	d5hmC 1 U	2	-	0.08	-	-	0.902
6B	Tet1CD late 30 min	18	-	2.45	0.29	-	-
	mcherry late 30 min	18	-	2.02	0.12	-	0.0478
	Tet1CD early/mid 30 min	5	-	1.51	0.1	-	-
	mcherry early/mid 30 min	7	-	1.32	0.07	-	0.0143
	Tet1CD DMSO 30 min	21	-	1.06	0.04	-	-
	mcherry DMSO 30 min	18	-	0.99	0.05	-	0.173
6C	J1 wt DMSO	13	-	0.98	0.02	-	-
	J1 TKO DMSO	11	-	1.03	0.03	-	0.25
	J1 wt Aph	17	-	1.25	0.12	-	-
	J1 TKO Aph	20	-	1.61	0.21	-	0.0136
7A	J1 wt	1512	-	0.74	0.02	-	-
	J1 TKO	1647	-	0.72	0.02	-	0.356
7C <sup>c</sup>	Stage I wt	368	-	0.25	0.05	-	-
	Stage II wt	458	-	0.31	0.07	-	-
	Stage III wt	239	-	0.15	0.04	-	-
	Stage Y wt	91	-	0.06	0.02	-	-
	Mitosis wt	35	-	0.02	0.01	-	-

8

Rausch et al.

Figure	Sample	n <sup>a</sup>	Median	Mean	StDev <sup>a</sup>	95% CI <sup>a</sup>	p-value <sup>a</sup>
	Non-S wt	321	-	0.21	0.08	-	-
	Stage I TKO	484	-	0.29	0.04	-	0.6502
	Stage II TKO	365	-	0.22	0.05	-	0.0126
	Stage III TKO	326	-	0.17	0.05	-	0.6521
	Stage Y TKO	90	-	0.05	0.02	-	0.8322
	Mitosis TKO	30	-	0.02	0.01	-	0.7161
	Non-S TKO	379	-	0.23	0.04	-	0.9102
7D	J1 wt	731	1.59	1.67	0.57	0.38 – 0.84	-
	J1 TKO	641	1.75	1.86	0.67	0.56 – 1.42	0.000268

<sup>a</sup> n: number of cells of all replicates (if not stated otherwise), StDev: standard deviation, 95% CI: 95% confidence interval, p-value: calculated as stated in figure legends and material and methods section.

<sup>b</sup> n numbers refer to independent PCRs analyzed and p-value is calculated against the band intensities of the dC containing experiment.

<sup>c</sup> p-values are calculated for wt versus TKO in every stage.

Rausch et al.

Supplementary Table 8: Plot statistics (supplementary figures).

Figure	Sample	n <sup>a</sup>	Median	Mean	StDev <sup>a</sup>	95% CI <sup>a</sup>	p-value <sup>a</sup>
S8C <sup>b</sup>	dC	2	-	1	0.05	0.55 - 1.45	-
	d5mC	2	-	0.46	0.14	-0.8 - 1.72	0.02641
	d5hmC	2	-	1.35	0.37	-1.97 - 4.67	0.03969
S8E <sup>b</sup>	dC	3	-	87.93	0.74	86.09 - 89.77	-
	d5mC	3	-	93.72	0.25	93.1 - 94.34	0.0014
	d5hmC	3	-	86.56	0.26	85.91 - 87.21	0.046
	d5fC	3	-	83.75	0.18	83.3 - 84.2	0.03121
	d5caC	3	-	85.4	0.95	83.04 - 87.76	0.423
S8F <sup>b</sup>	dC	3	-	85.7	0.68	83.86 - 87.54	-
	d5mC	3	-	89.95	0.78	89.33 - 90.57	0.003652
	d5hmC	3	-	83.44	0.9	82.79 - 84.09	0.04702
	d5fC	3	-	82.62	0.55	82.17 - 83.07	0.003151
	d5caC	3	-	83.22	0.72	80.86 - 85.58	0.02314
S10A	60°C	114	1	1.33	0.61	1.22 - 1.44	-
	70°C	113	4.81	4.89	0.77	4.75 - 5.03	8.24E-07
	80°C	121	6.1	5.98	2.33	5.56 - 6.4	0.001492
S10B up	Tet1CD 65°C 1	545	2.15	2.3	0.65	2.24 - 2.35	< 2.2E-16
	Tet1CD 65°C 2	2601	2.54	2.88	1.3	2.83 - 2.93	< 2.2E-16
	Tet1CD 65°C 3	732	2.72	3.17	1.76	3.04 - 3.3	< 2.2E-16
	mcherry 65°C 1	353	1	1.09	0.6	1.02 - 1.15	-
	mcherry 65°C 2	3080	1.08	1.23	0.68	1.2 - 1.25	0.417
	mcherry 65°C 3	2145	1.16	1.38	1.42	1.32 - 1.44	0.415
	Tet1CD 70°C 1	520	2.56	2.73	0.8	2.66 - 2.8	< 2.2E-16
	Tet1CD 70°C 2	2622	2.92	3.29	1.64	3.23 - 3.35	< 2.2E-16
	Tet1CD 70°C 3	693	3.11	3.65	2.32	3.48 - 3.83	< 2.2E-16
	mcherry 70°C 1	1615	1.42	1.53	1.02	1.48 - 1.58	-
	mcherry 70°C 2	2924	1.52	1.77	2.32	1.69 - 1.85	0.1144
	mcherry 70°C 3	2770	1.61	1.87	2.15	1.79 - 1.95	0.0334
S10B down	Tet1CD 60°C 1	1002	1.68	2.01	1.05	1.94 - 2.08	< 2.2E-16
	Tet1CD 60°C 2	1189	1.69	1.92	1.62	1.83 - 2.01	< 2.2E-16
	Tet1CD 60°C 3	3347	1.75	2.72	2.16	2.65 - 2.79	< 2.2E-16
	mcherry 60°C 1	1043	1	1.19	0.63	1.15 - 1.23	-
	mcherry 60°C 2	2197	0.97	1.29	1.75	1.22 - 1.36	0.708
	mcherry 60°C 3	4203	1.06	1.4	1.99	1.34 - 1.46	0.2168
	Tet1CD 65°C 1	2190	2.31	2.21	0.82	2.18 - 2.24	< 2.2E-16
	Tet1CD 65°C 2	2664	2.15	2.44	1.1	2.4 - 2.48	< 2.2E-16
	Tet1CD 65°C 3	3709	2.65	3.9	2.9	3.81 - 3.99	< 2.2E-16
	mcherry 65°C 1	1010	1.02	1.11	0.65	1.07 - 1.15	-
	mcherry 65°C 2	2428	1.04	1.08	0.79	1.05 - 1.11	0.4578
	mcherry 65°C 3	3104	1.01	1.28	0.97	1.25 - 1.31	0.05469
S10C	Tet1CD 1	3318	0.48	0.41	0.14	0.41 - 0.42	< 2.2E-16
	Tet1CD 2	1142	0.5	0.45	0.17	0.44 - 0.46	< 2.2E-16
	Tet1CD 3	128	0.51	0.46	0.16	0.43 - 0.49	< 2.2E-16
	mcherry 1	3292	0.94	0.98	0.25	0.97 - 0.99	0.857
	mcherry 2	2809	0.97	1.04	0.35	1.03 - 1.05	0.7245
	mcherry 3	295	0.98	1.09	0.39	1.05 - 1.13	0.685
	untransf.	4460	1	1.03	0.33	1.02 - 1.04	-
S10D	acet. +TSA	2806	13.68	3.28	1.95	3.21 - 3.35	< 2.2E-16
	acet. -TSA	2714	1	1.05	0.56	1.03 - 1.07	-
	decond. +TSA	1342	0.39	0.42	0.32	0.4 - 0.44	< 2.2E-16
	decond. -TSA	2262	1	1.34	0.56	1.32 - 1.36	-
	5hmC +TSA	1342	1.01	1.06	0.87	1.01 - 1.11	0.5034
	5hmC -TSA	2262	1	1.03	0.63	1 - 1.06	-
	60°C +TSA	2540	0.96	1.32	0.54	1.3 - 1.34	0.554
	60°C -TSA	2351	1	1.56	0.36	1.55 - 1.57	-
	65°C +TSA	1956	1.12	1.23	0.98	1.19 - 1.27	0.079
	65°C -TSA	3976	1	1.54	1.02	1.51 - 1.57	-
	70°C +TSA	1845	1.04	1.52	0.98	1.48 - 1.56	0.208
	70°C -TSA	2630	1	1.21	0.65	1.19 - 1.23	-
	S11A	Tet1CD 1	1595	1.24	1.26	1.4	1.19 - 1.33

10

Rausch et al.

Figure	Sample	n <sup>a</sup>	Median	Mean	StDev <sup>a</sup>	95% CI <sup>a</sup>	p-value <sup>a</sup>
	Tet1CD 2	1326	1.94	2.47	1.98	2.36 - 2.58	1.14E-06
	Tet1CD 3	1550	3.54	4.47	1.6	4.39 - 4.55	1.03E-08
	mcherry 1	3688	1.26	1.44	0.8	1.41 - 1.47	0.221
	mcherry 2	1902	1.14	1.91	1.92	1.82 - 2	0.6627
	mcherry 3	1610	1.18	1.9	0.78	1.86 - 1.94	0.83
	untransf.	1950	1	1.22	1.1	1.17 - 1.27	-
S11B	60°C	129	1	1	0.01	0.99 - 1	-
	65°C	123	1.02	1.02	0.01	1.02 - 1.02	1.05E-07
	70°C	115	1.05	1.06	0.03	1.05 - 1.07	1.13E-11
	75°C	118	1.63	1.67	0.26	1.62 - 1.72	5.04E-10
	80°C	128	3.26	3.56	1.35	3.32 - 3.8	6.89E-11
S11C	Tet1CD 65°C	1116	1.38	1.32	0.03	1.31 - 1.32	2.74E-10
	mcherry 65°C	2168	1	1	0.01	0.99 - 1	-
	Tet1CD 70°C	1218	1.51	1.39	0.05	1.38 - 1.39	5.35E-07
	mcherry 70°C	2034	1.1	1.03	0.01	1.02 - 1.03	-
	Tet1CD 75°C	1326	2.32	2.33	0.14	2.32 - 2.33	8.84E-06
	mcherry 75°C	1929	1.36	1.19	0.12	1.18 - 1.19	-
	Tet1CD 80°C	2105	2.62	2.38	0.43	2.36 - 2.4	0.005725
	mcherry 80°C	2251	2.48	2.4	0.16	2.39 - 2.41	-
S12B	Tet1CD 60 min 1	2293	1.03	1.15	0.53	1.13 - 1.17	< 2.2E-16
	Tet1CD 60 min 2	1343	1.26	1.4	0.69	1.36 - 1.44	< 2.2E-16
	Tet1CD 60 min 3	257	1.26	1.45	0.87	1.34 - 1.55	9.58E-11
	mcherry 60 min 1	2835	1	1.08	0.4	1.06 - 1.09	-
	mcgerry60 min 2	2795	1.08	1.17	0.46	1.16 - 1.19	0.0975
	mcherry 60 min 3	1635	1.09	1.18	0.44	1.16 - 1.2	0.847
S13A <sup>c</sup>	dC 30 sec	12	0.08	0.1	0.02	0.04 - 0.05	-
	dC 60 sec	12	0.14	0.22	0.05	0.12 - 0.13	-
	dC 90 sec	12	0.25	0.35	0.05	0.24 - 0.25	-
	dC 120 sec	12	0.4	0.47	0.05	0.38 - 0.4	-
	dC 600 sec	12	1	1	0	0.9 - 1.0	-
	d5mC 30 sec	12	0.06	0.05	0.02	0.05 - 0.06	n.s.
	d5mC 60 sec	12	0.11	0.12	0.03	0.10 - 0.12	n.s.
	d5mC 90 sec	12	0.22	0.23	0.03	0.21 - 0.22	n.s.
	d5mC 120 sec	12	0.37	0.34	0.04	0.36 - 0.38	n.s.
	d5mC 600 sec	12	0.87	0.81	0.08	0.8 - 0.82	n.s.
	d5hmC 30 sec	12	0.06	0.03	0.03	0.06 - 0.07	n.s.
	d5hmC 60 sec	12	0.15	0.14	0.02	0.14 - 0.15	n.s.
	d5hmC 90 sec	12	0.28	0.27	0.03	0.27 - 0.29	n.s.
	d5hmC 120 sec	12	0.41	0.39	0.04	0.41 - 0.42	n.s.
	d5hmC 600 sec	12	0.97	0.92	0.08	0.96 - 0.98	n.s.
S13B	Tet1CD 1a	5768	1.08	1.26	0.59	1.2 - 1.22	< 2.2E-16
	Tet1CD 1b	1153	1.31	1.49	0.68	1.24 - 1.31	< 2.2E-16
	Tet1CD 1c	293	1.34	1.51	0.66	1.1 - 1.2	< 2.2E-16
	Tet1CD 2	159	1.38	1.49	0.69	1.01 - 1.2	5.59E-12
	Tet1CD 3	92	1.12	1.27	0.55	0.93 - 1.21	0.0017
	Tet1CD 4	213	1.1	1.25	0.6	0.81 - 0.94	0.00011
	mcherry 1a	9283	1	1.09	0.38	1.04 - 1.06	-
	mcherry 1b	4687	1.06	1.19	0.53	1.03 - 1.06	0.89
	mcherry 1c	1500	1.07	1.24	0.8	1.09 - 1.15	0.65
	mcherry 2	1020	1.14	1.32	1.23	1.02 - 1.08	0.056
	mcherry 3	315	1.07	1.17	1.06	1.04 - 1.22	0.36
	mcherry 4	409	1.08	1.21	1.28	0.9 - 1.01	0.078
S14A <sup>b</sup>	dC 0 U	2	-	1	-	-	-
	dC 2 U	2	-	1.49	-	-	-
	d5mC 0 U	2	-	1	-	-	-
	d5mC 2 U	2	-	0.9	-	-	0.0121
	d5hmC 0 U	2	-	1	-	-	-
	d5hmC 2 U	2	-	1.38	-	-	0.85
S14B	Tet1CD 10 min 1	4125	1.26	1.32	0.65	1.3 - 1.34	< 2.2E-16
	Tet1CD 10 min 2	808	1.46	1.67	0.98	1.6 - 1.74	< 2.2E-16
	Tet1CD 10 min 3	650	1.52	1.83	0.87	1.76 - 1.9	3.52E-06

Rausch et al.

Figure	Sample	n <sup>a</sup>	Median	Mean	StDev <sup>a</sup>	95% CI <sup>a</sup>	p-value <sup>a</sup>
	mcherry 10 min 1	1950	1	1.01	0.97	0.97-1.05	-
	mcherry 10 min 2	798	1.21	1.08	1.01	1.01 - 1.15	0.022
	mcherry 10 min 3	127	1.15	1.08	0.89	0.92 - 1.24	0.323
	Tet1CD 20 min 1	1258	1.97	2.32	1.04	2.26 - 2.38	< 2.2E-16
	Tet1CD 20 min 2	1444	2.02	2.1	1.02	2.05 - 2.15	< 2.2E-16
	Tet1CD 20 min 3	184	2.01	2.08	0.98	1.94 - 2.22	0.002748
	mcherry 20 min 1	948	1.87	2.03	1.02	1.96 - 2.1	-
	mcherry 20 min 2	441	1.98	2.1	1.12	2 - 2.2	0.0397
	mcherry 20 min 3	850	1.88	1.99	0.99	1.92 - 2.06	0.1531
	Tet1CD 30 min 1	1308	3.01	3.42	1.85	3.32 - 3.52	< 2.2E-16
	Tet1CD 30 min 2	1479	3.02	3.28	1.68	3.19 - 3.37	8.47E-16
	Tet1CD 30 min 3	155	3	3.18	1.45	2.95 - 3.41	0.01529
	mcherry 30 min 1	1216	2.85	3.08	1.04	3.02 - 3.14	-
	mcherry 30 min 2	3657	2.95	3.13	1.04	3.1 - 3.16	0.022
	mcherry 30 min 3	578	2.98	3.37	1.11	3.28 - 3.46	0.03161
S15A	dC	2	-	1	-	-	-
	12.5% d5hmC	2	-	0.92	-	-	-
	50% d5hmC	2	-	0.81	-	-	-
	100% d5hmC	2	-	0.41	-	-	-
	12.5% d5mC	2	-	0.47	-	-	-
	50% d5mC	2	-	0.38	-	-	-
	100% d5mC	2	-	0.32	-	-	-
S15B	Tet1CD 1a	1258	1.17	1.21	0.55	1.99 - 1.23	< 2.2E-16
	Tet1CD 1b	1444	1.18	1.28	0.69	1.24 - 1.31	< 2.2E-16
	Tet1CD 1c	184	1.11	1.48	0.59	1.06 - 1.24	0.0405
	Tet1CD 2	93	1.1	1.1	0.6	0.98 - 1.22	0.458
	Tet1CD 3	38	0.92	1.1	0.67	0.85 - 1.29	0.879
	Tet1CD 4	67	0.77	0.88	0.47	0.76 - 0.99	0.0047
	mcherry 1a	948	1	1.05	0.53	1.02 - 1.09	-
	mcherry 1b	441	0.97	1.05	0.57	0.99 - 1.1	0.847
	mcherry 1c	850	0.99	1.12	0.63	0.98 - 1.26	0.38
	mcherry 2	58	0.99	1.05	0.48	0.93 - 1.18	0.96
	mcherry 3	31	1.05	1.13	0.77	0.85 - 1.41	0.589
	mcherry 4	44	0.94	0.96	0.55	0.79 - 1.12	0.087
S16B	Tet1CD	35	0.98	0.98	0.03	0.97 - 0.99	0.341
	mcherry	31	1	0.99	0.06	0.97 - 1.01	-
S16C	Tet1CD 30 min	31	-	1.8	0.13	-	-
	mcherry 30 min	29	-	1.52	0.08	-	0.0489
	Tet1CD 12 min	35	-	1.33	0.09	-	-
	mcherry 12 min	30	-	1.16	0.08	-	< 2.2E-16
S16E	MEF Tet1CD 30 min	11	-	1.71	0.09	-	-
	MEF mcherry 30 min	19	-	1.45	0.07	-	0.048
	MEF Tet1CD DMSO 30 min	14	-	1.11	0.02	-	-
	MEF mcherry DMSO 30 min	10	-	1.12	0.02	-	0.0873
S17A	Tet1CD	21	0.78	0.81	0.1	0.73 - 0.84	-
	mcherry	18	0.25	0.23	0.08	0.2 - 0.3	2.76E-11
S18C	MEF W8 30 min	26	-	2.38	0.13	-	0.0144
	MEF PM 30 min	34	-	2.09	0.12	-	-
	MEF W8 DMSO 30 min	31	-	1.17	0.02	-	0.433
	MEF PM DMSO 30 min	31	-	1.13	0.04	-	-
S20A	Nuclear H3K9ac J1 wt	29	-	52.38	8.29	-	-
	Nuclear H3K9ac J1 TKO	34	-	47.21	3.21	-	0.92
	Chromocentric H3K9ac J1 wt	29	-	41.42	3.3	-	-
	Chromocentric H3K9ac J1 TKO	34	-	47.9	8.57	-	0.47
	Nuclear H3K18ac J1 wt	31	-	52.62	9.79	-	-
	Nuclear H3K18ac J1 TKO	20	-	51.1	10.2	-	0.505
	Chromocentric H3K18ac J1 wt	31	-	47.8	2.4	-	-
	Chromocentric H3K18ac J1 TKO	20	-	51.8	4.35	-	0.384
	Nuclear H3K27ac J1 wt	47	-	37.03	5.71	-	-
	Nuclear H3K27ac J1 TKO	23	-	54.3	5.04	-	0.0967

12

Rausch et al.

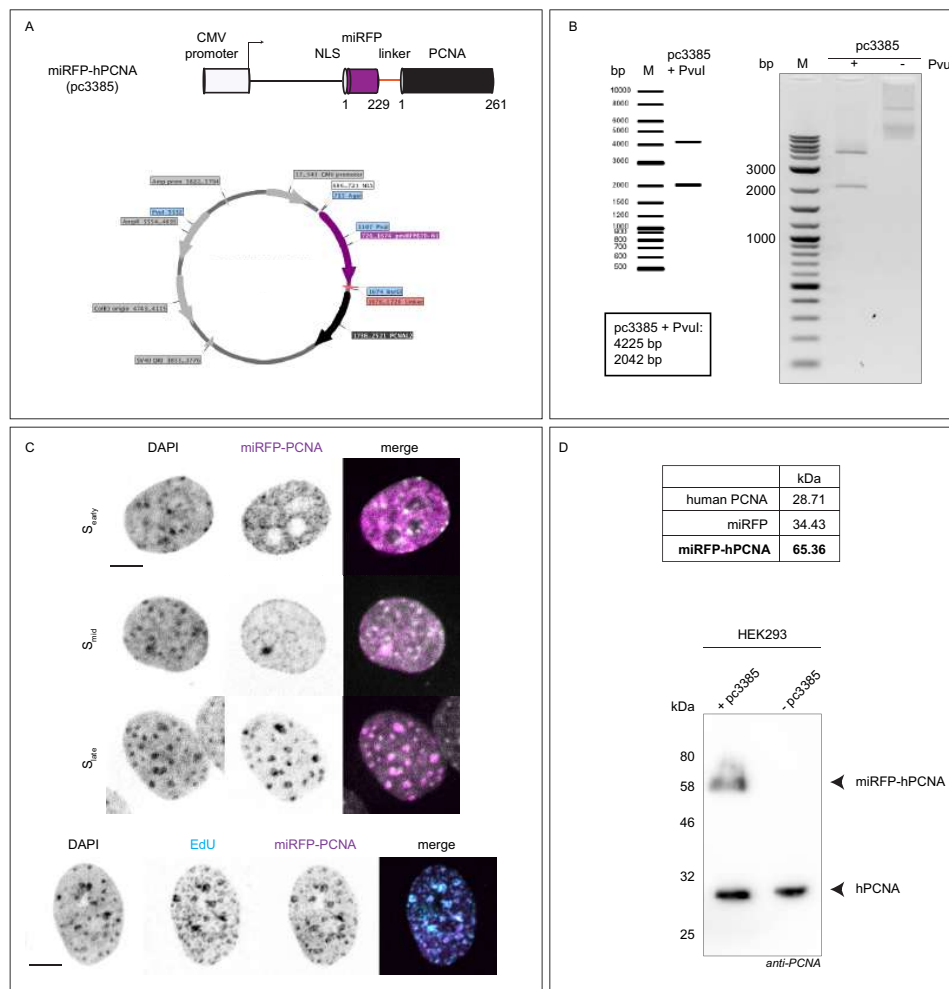
Figure	Sample	n <sup>a</sup>	Median	Mean	StDev <sup>a</sup>	95% CI <sup>a</sup>	p-value <sup>a</sup>
	Chromocentric H3K27ac J1 wt	47	-	38.52	3.6	-	-
	Chromocentric H3K27ac J1 wt	23	-	49.02	6.8	-	0.073
	Nuclear H3K56ac J1 wt	33	-	83.58	9.18	-	-
	Nuclear H3K56ac J1 TKO	21	-	87.87	1.67	-	0.97
	Chromocentric H3K56ac J1 wt	33	-	85.73	13.39	-	-
	Chromocentric H3K56ac J1 TKO	21	-	83.01	14.25	-	0.835
	Nuclear H4K8ac J1 wt	28	-	48.66	12.14	-	-
	Nuclear H4K8ac J1 TKO	22	-	52.25	8.93	-	0.523
	Chromocentric H4K8ac J1 wt	28	-	51.2	9.04	-	-
	Chromocentric H4K8ac J1 TKO	22	-	57.8	6.27	-	0.368
	Nuclear H4K16ac J1 wt	26	-	49.97	8.14	-	-
	Nuclear H4K16ac J1 TKO	23	-	38.19	10.76	-	0.832
	Chromocentric H4K16ac J1 wt	26	-	63.8	8.34	-	-
	Chromocentric H4K16ac J1 TKO	23	-	57.8	4.63	-	0.551
S20B	Nuclear H3K9m3 J1 wt	30	-	49.08	11.01	-	-
	Nuclear H3K9m3 J1 TKO	26	-	38.82	7.19	-	0.16
	Chromocentric H3K9m3 J1 wt	30	-	80.46	8.82	-	-
	Chromocentric H3K9m3 J1 TKO	26	-	65.32	5.97	-	0.132
	Nuclear H4K20m3 J1 wt	34	-	19.83	4.83	-	-
	Nuclear H4K20m3 J1 TKO	29	-	21.65	7.33	-	0.622
	Chromocentric H4K20m3 J1 wt	34	-	38.09	7.95	-	-
	Chromocentric H4K20m3 J1 TKO	29	-	33.36	9.64	-	0.403

<sup>a</sup> n: number of cells of all replicates (if not stated otherwise), StDev: standard deviation, 95% CI: 95% confidence interval, p-value: calculated as stated in figure legends and material and methods section.

<sup>b</sup> n numbers refer to independent PCRs analyzed and p-value is calculated against the band intensities of the dC containing experiment.

<sup>c</sup> 2 way ANOVA statistical analysis did not reveal any significant difference (n.s.) between dC, d5mC and d5hmC.

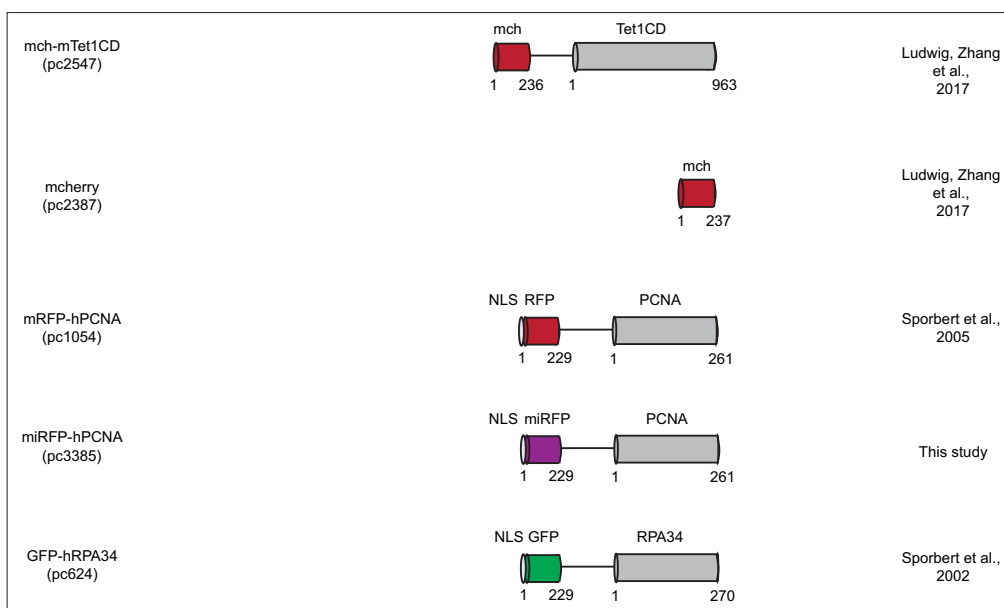
Rausch et al.



**Supplementary Figure 1 – Characterization of the pmIRFP-PCNA plasmid.** (A) miRFP was cloned into the RFP-hPCNA plasmid (pc1054) as described in the materials and methods section, to generate the miRFP-hPCNA plasmid (pc3385). The gene expression in mammalian cells is driven by a CMV promoter. Plasmid map shows all important vector components, including the restriction sites used for cloning (AgeI & BsrGI) and control digest (PvuI) of the recombinant plasmid. (B) Control digest of the miRFP-hPCNA construct using PvuI results in two bands of 4225 bp and 2042 bp, respectively. M stands for DNA size marker. (C) The subcellular localization of the miRFP-hPCNA construct was validated by microscopical analysis in transfected C2C12 mouse myoblast cells. The miRFP-hPCNA fusion protein showed the early, mid and late S-phase pattern described for somatic cells.  $S_{early}$  is characterized by a homogeneous distribution of replication throughout the nucleus, in  $S_{mid}$  replication signals are observed at the nucle(ol)ar periphery and  $S_{late}$  is marked by fewer but larger replication signals co-localizing with DAPI intense stained chromocenters. Labeling transfected cells with EdU resulted in a co-localization of the detection of nascent DNA (EdU) and the marked replisome component (miRFP-PCNA). (D) The expected miRFP-hPCNA fusion protein was detected via western blotting using a PCNA specific antibody. HEK293 cells were transfected with pc3385, lysed in loading dye and whole cell lysates were analyzed by SDS-PAGE gel electrophoresis followed by blotting onto a nitrocellulose membrane. Detection with anti-PCNA antibody showed bands at the size of the miRFP-hPCNA fusion protein (65 kDa) and the untagged endogenous PCNA (29 kDa), whereas the untransfected control cells (-pc3385) only showed signal for the endogenous PCNA. Scale bar = 5  $\mu$ m. pc numbers refer to the plasmid collection numbers.

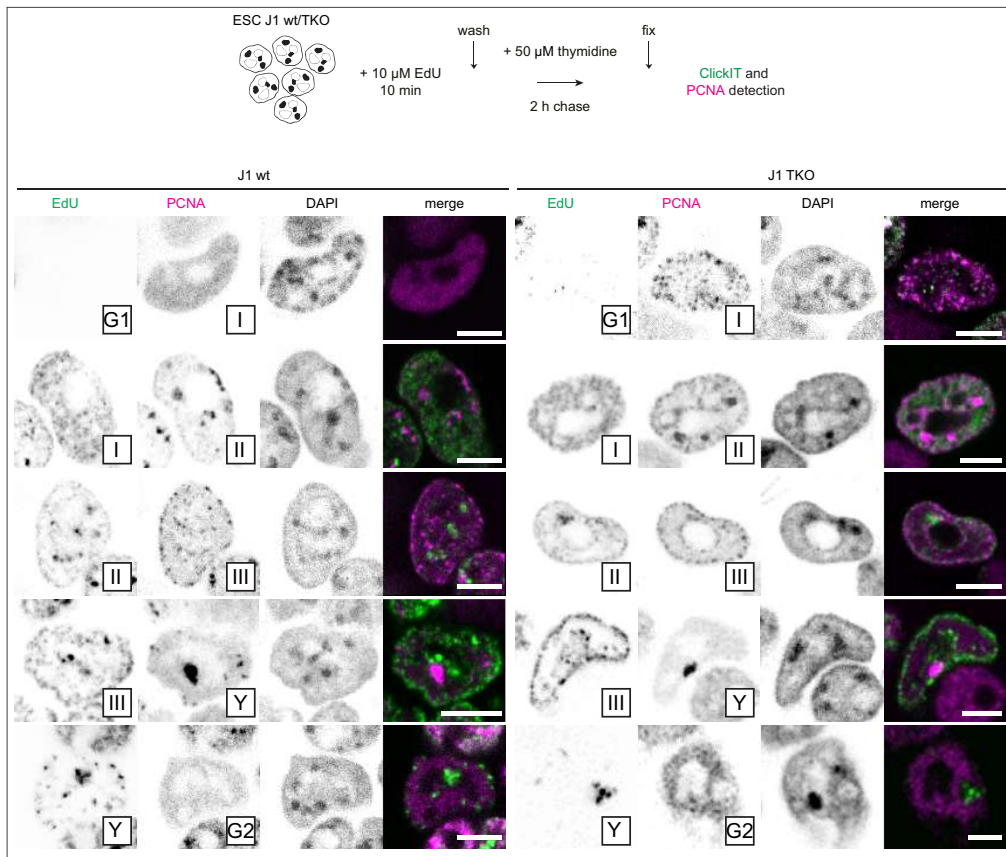
14

Rausch et al.



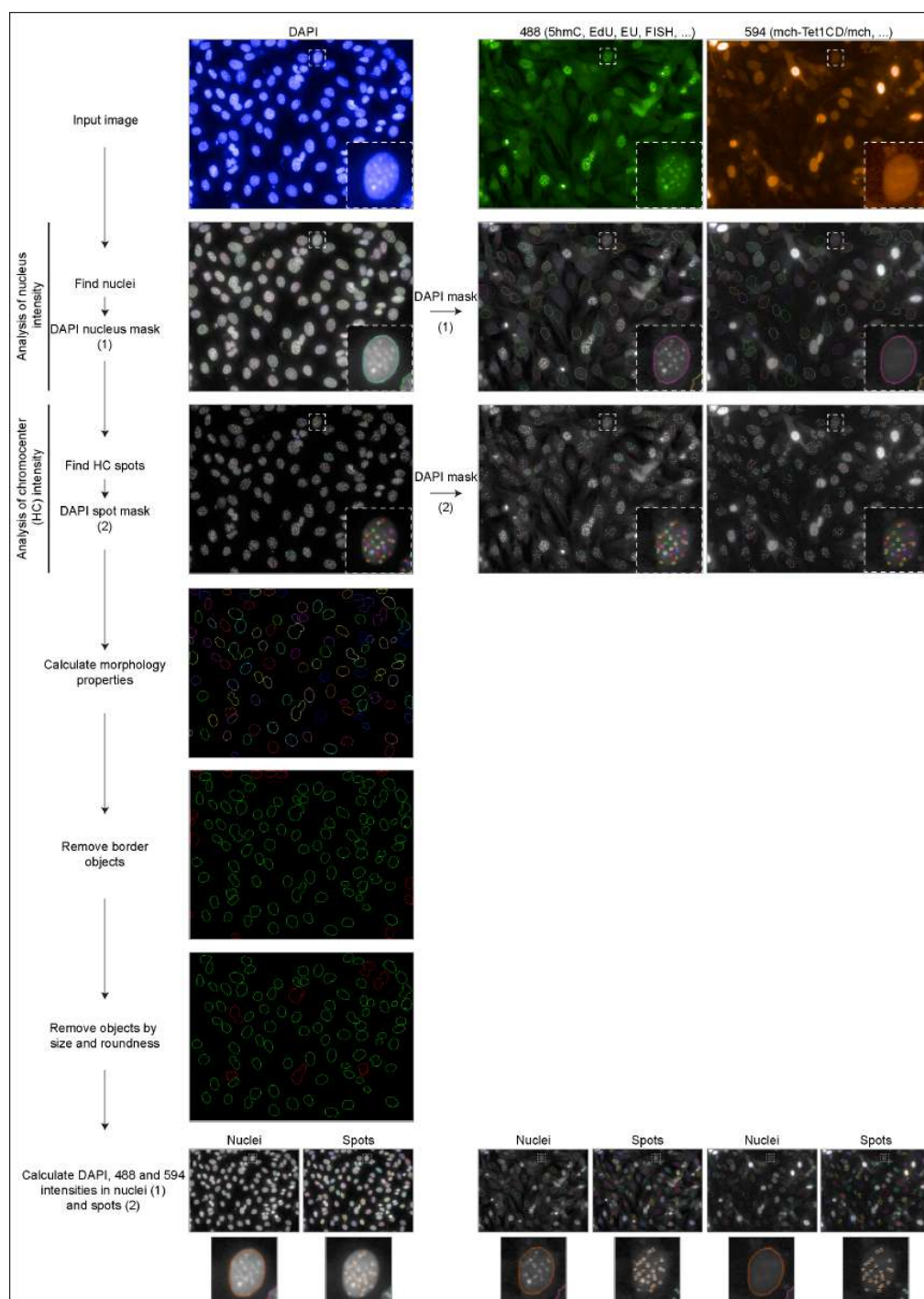
**Supplementary Figure 2 – Schematic illustration of the relevant protein domains encoded in the plasmids used.** Amino acid coordinates are indicated below every structure. pc numbers refer to the plasmid collection numbers. Respective publications are indicated. Plasmid construction details are described in the materials and methods part for the plasmids not published before. m: mouse, h: human, NLS: nuclear localization signal.

Rausch et al.



**Supplementary Figure 3 – Effect of loss of global DNA methylation on the spatio-temporal progression of DNA replication in ES cells.** J1 wt and TKO ES cells were pulse labeled with EdU and fixed after a chase period of 2 hours. (Immuno)fluorescent detection of the nucleotide analogue and endogenous PCNA allowed the analysis of the spatio-temporal progression of S-phase in fixed cells. Representative confocal images of the transitions from G1 to stage I, S-phase substage progressions and progression from stage Y to G2 are shown. Scale bar = 5  $\mu$ m.

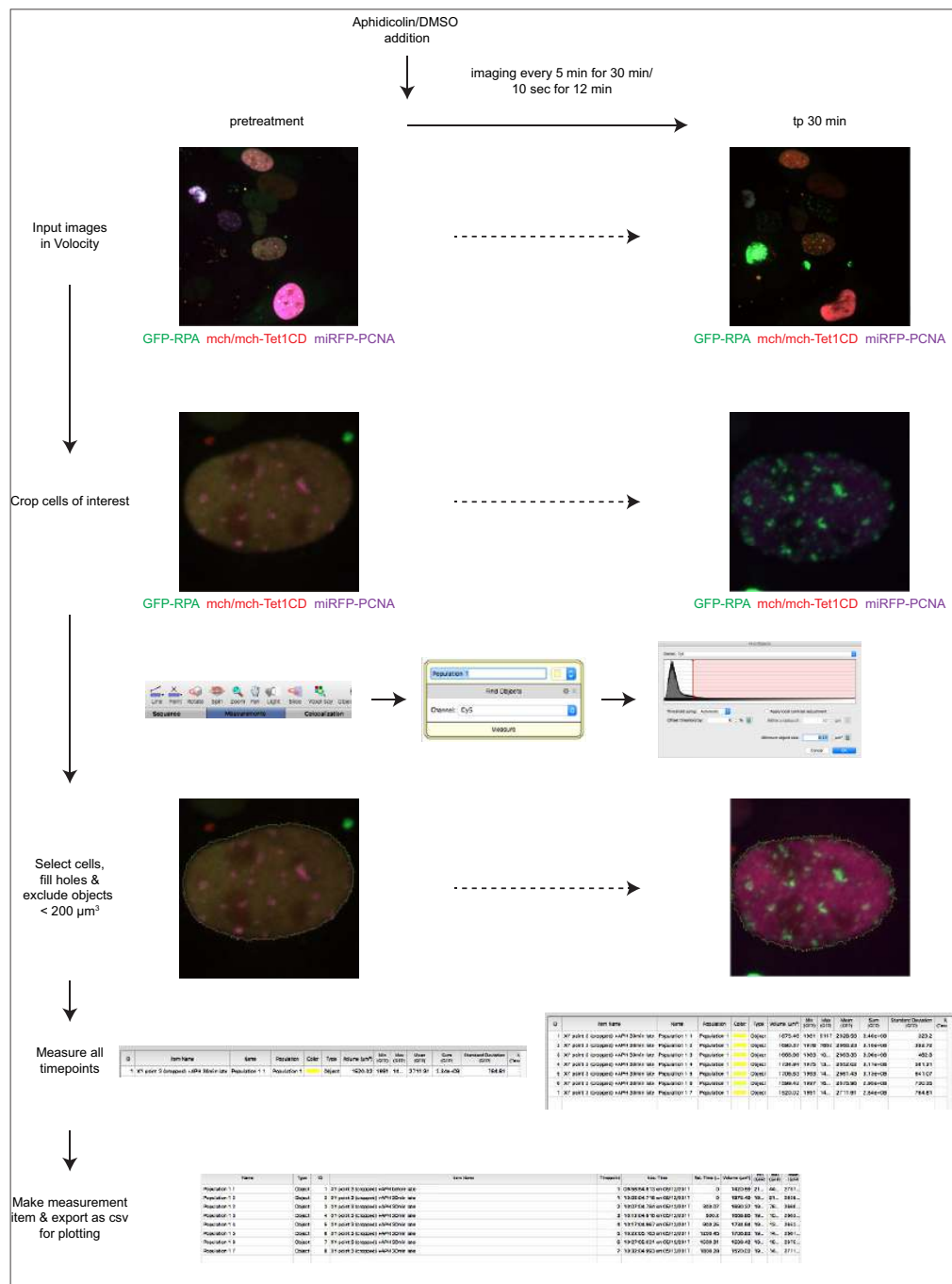
Rausch et al.



**Supplementary Figure 4 – High content screening microscopy analysis pipeline.**

Analysis of high content screening microscopy images was done using the PerkinElmer Harmony software. Nuclei and chromocenters (HC spots) were found and masked based on the DAPI input images. DAPI masks were used to segment 488 nm and 594 nm channel images. Morphology properties of segmented nuclei were calculated (roundness, area, ...), border objects were removed from the analysis and objects were further selected according to their size and roundness. Intensities of the DAPI, 488 and 594 channels were calculated for whole nuclei, chromocenter spots or nucleoli.

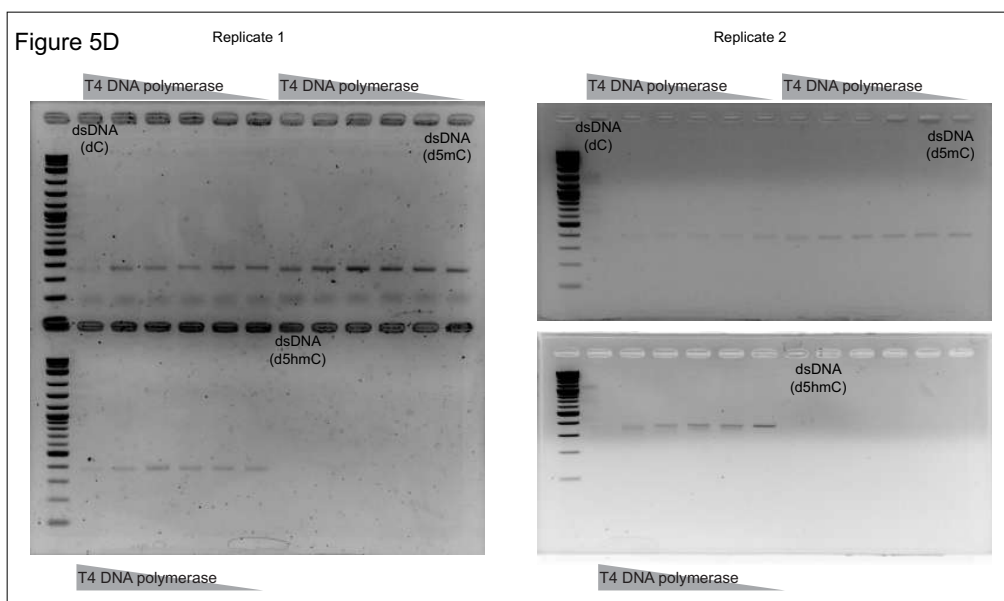
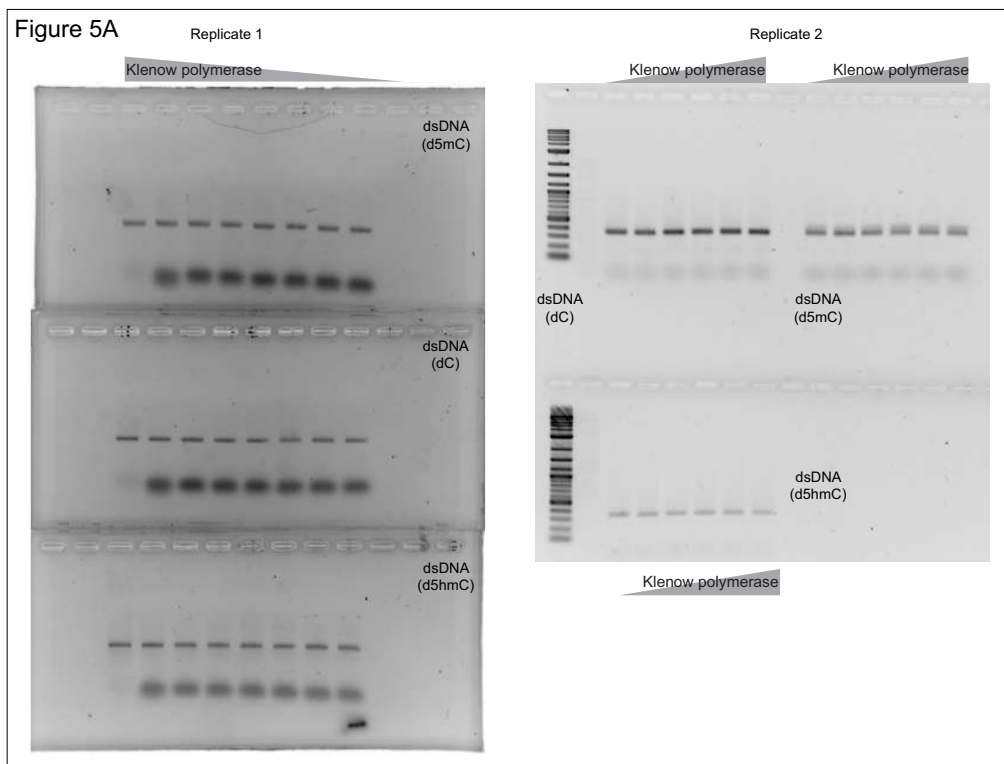
Rausch et al.



**Supplementary Figure 5 – Live cell RPA accumulation analysis pipeline.** Analysis of live cell microscopy images was done using the PerkinElmer Velocity software. Cells of interest were cropped in all images/timepoints. Cells were masked based on the GFP or miRFP signal. Objects smaller than  $200 \mu\text{m}^3$  were excluded and holes within the mask were filled. Signal intensities were measured in pretreatment and for all timepoints of treated cells. Measurements were grouped, saved and exported as csv files for further analysis and plotting as described in the Material and Methods section.

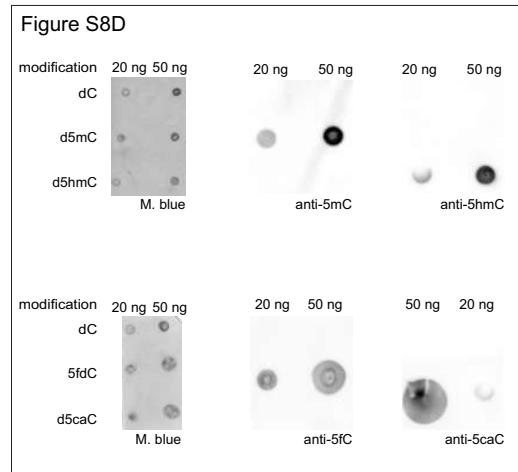
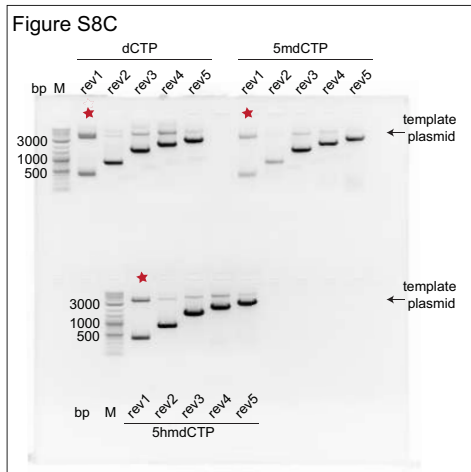
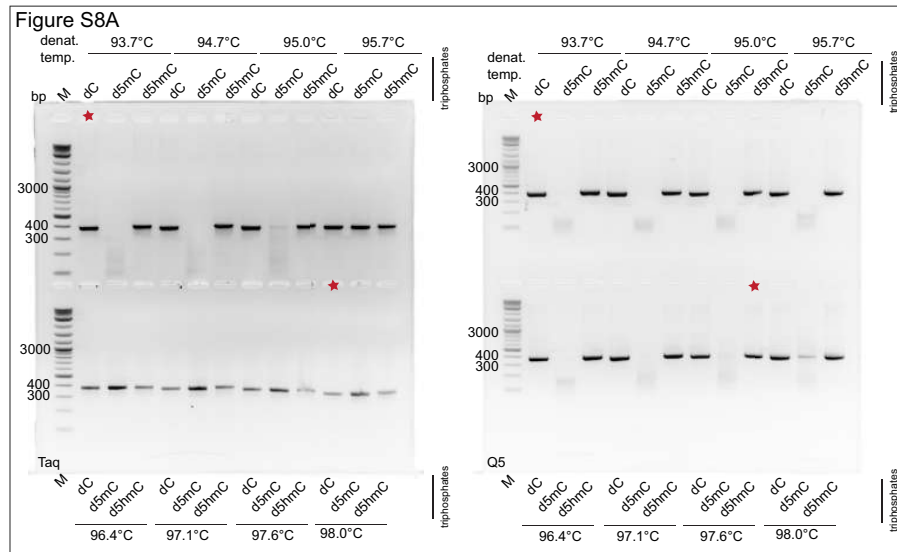
18

Rausch et al.



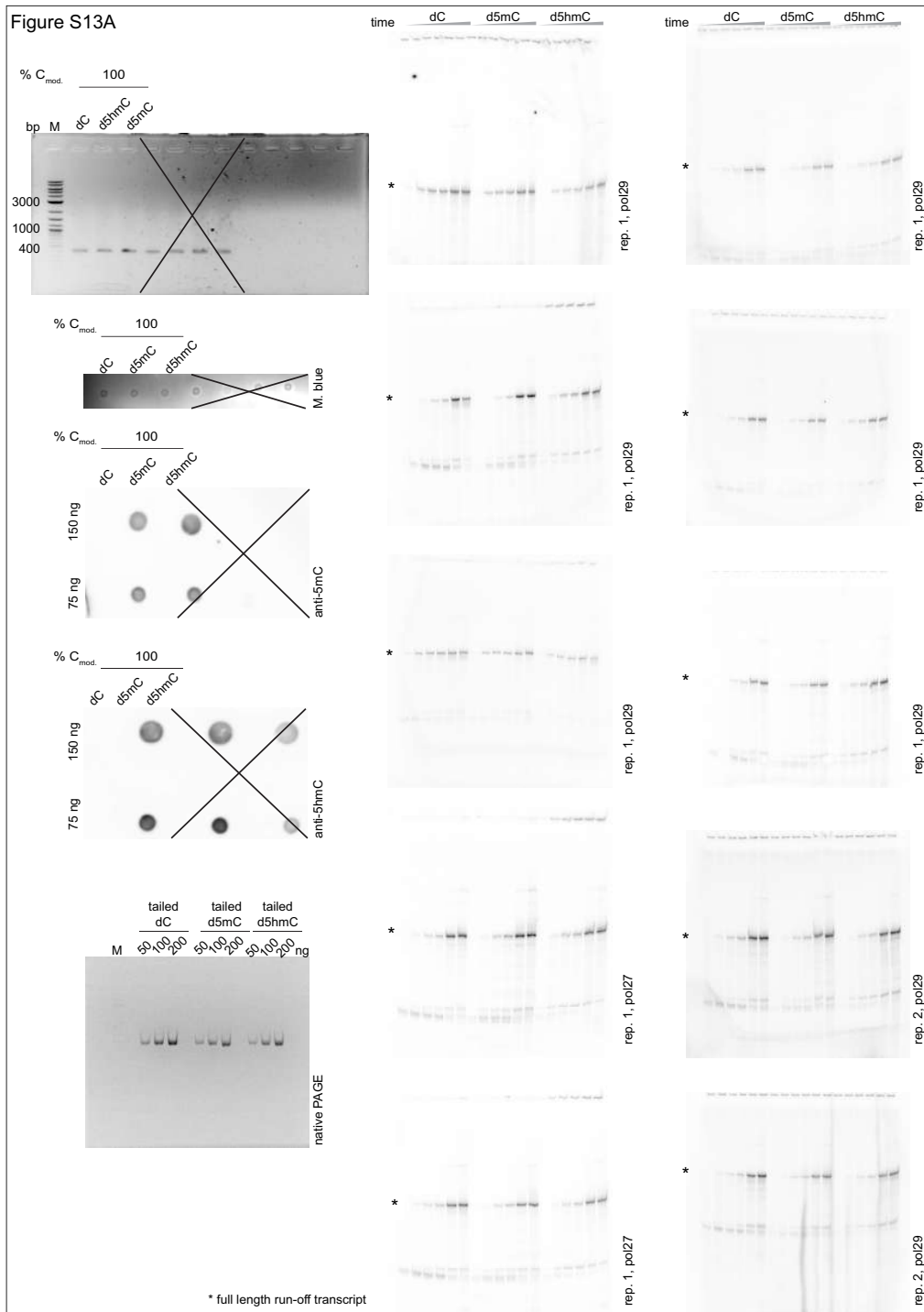
**Supplementary Figure 6 – Full gel images used for the analysis of the Klenow (Figure 5A) and T4 polymerase assays (Figure 5D). The two independent replicates are shown.**

Rausch et al.

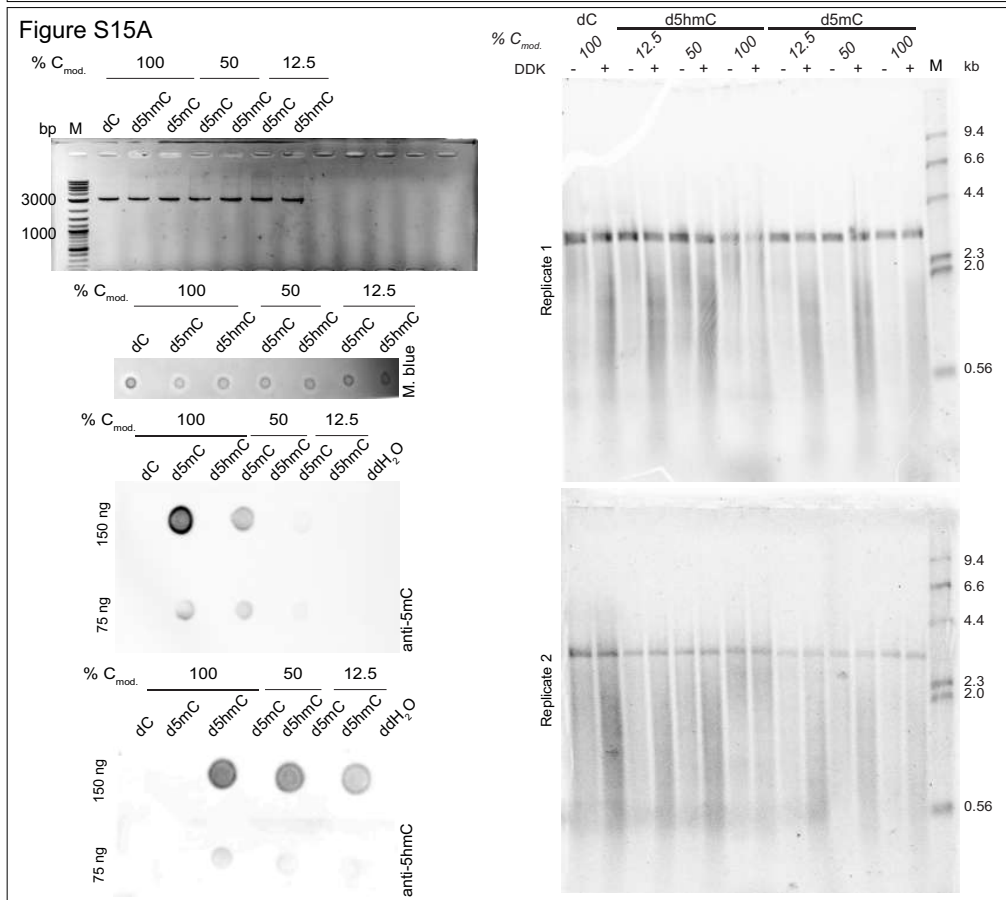
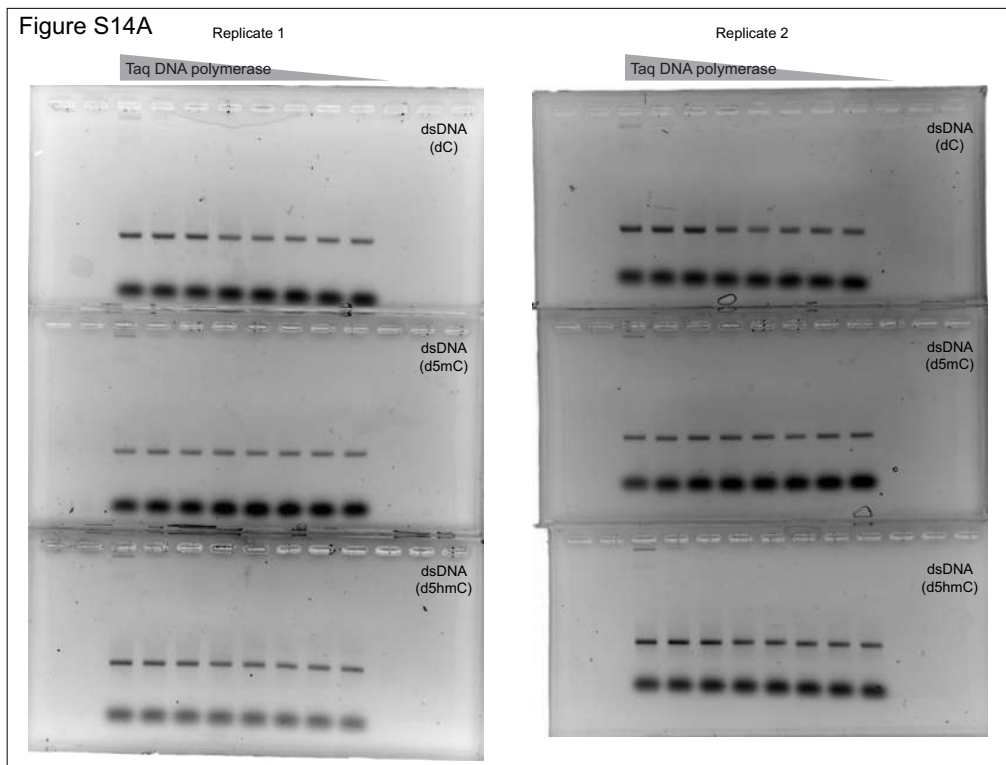


20

Rausch et al.



Rausch et al.

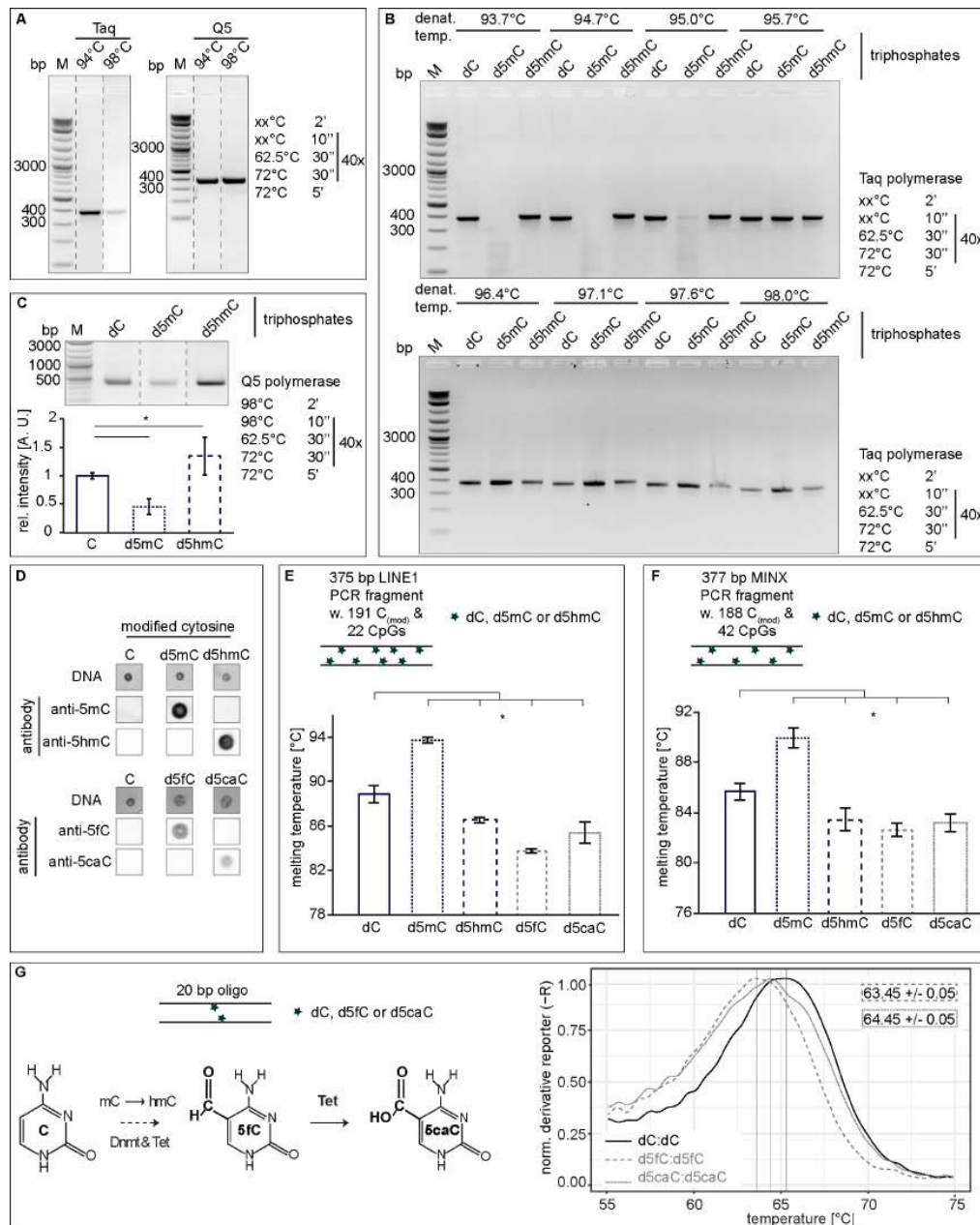


22

Rausch et al.

**Supplementary Figure 7 – Full gel images used for polymerase efficiency assays (Supplementary Figure 8A), PCR yield assays (Supplementary Figure 8C), slot blot analysis of modified PCR fragments (Supplementary Figure 8D), *in vitro* yeast RNA polymerization assays (Supplementary Figure 13A), Taq polymerase assays (Supplementary Figure 14A) and *in vitro* yeast replication assays (Supplementary Figure 15A).** The independent replicates of the assays are shown if applicable. Crossed out parts of gels and blots were not used in the manuscript.

Rausch et al.



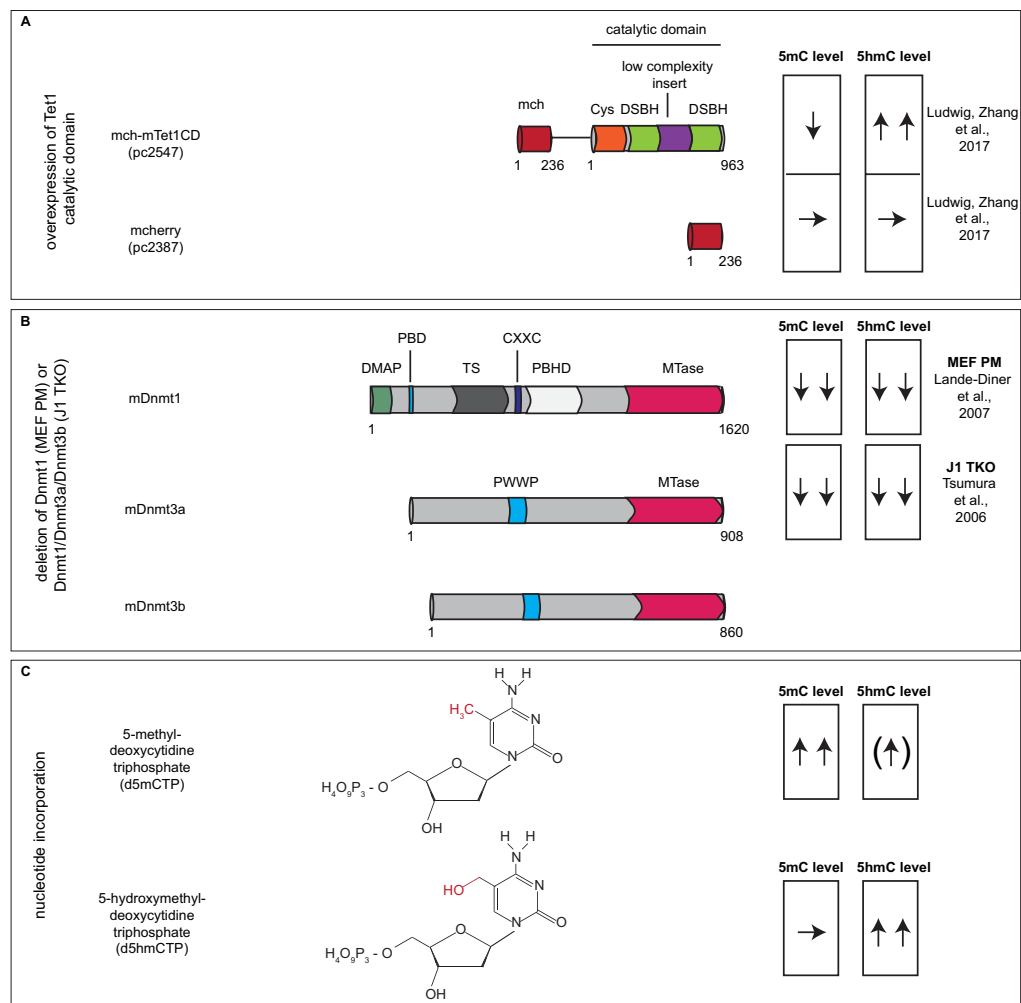
**Supplementary Figure 8 – Effect of cytosine modifications on DNA double helix stability *in vitro*.** (A) Agarose gel electrophoretic analysis of MINX PCR amplification efficiency obtained with Taq and Q5 polymerase with dCTP and using denaturing temperatures of 94°C and 98°C. (B) Agarose gel electrophoretic analysis of MINX PCR amplification products generated with Taq polymerase, denaturing temperatures between 93°C and 98°C and using dCTP, d5mCTP or d5hmCTP. (C) Agarose gel electrophoretic analysis of MINX PCR amplification product yields obtained with Q5 polymerase and using dCTP, d5mCTP or d5hmCTP and normalized quantification of PCR band intensities. (D) Slot blot analysis of dC, d5mC, d5hmC, d5fC and d5caC PCR amplicons. 20 ng of DNA were loaded and membranes were stained with methylene blue as a loading control (DNA). Membranes were incubated with anti-5mC, anti-5hmC, anti-5fC and anti-5caC antibodies, respectively and detected with an HRP conjugated secondary antibody. Exposure times and image contrast were adjusted for every antibody. (E-F), Melting temperature of LINE1 PCR

24

Rausch et al.

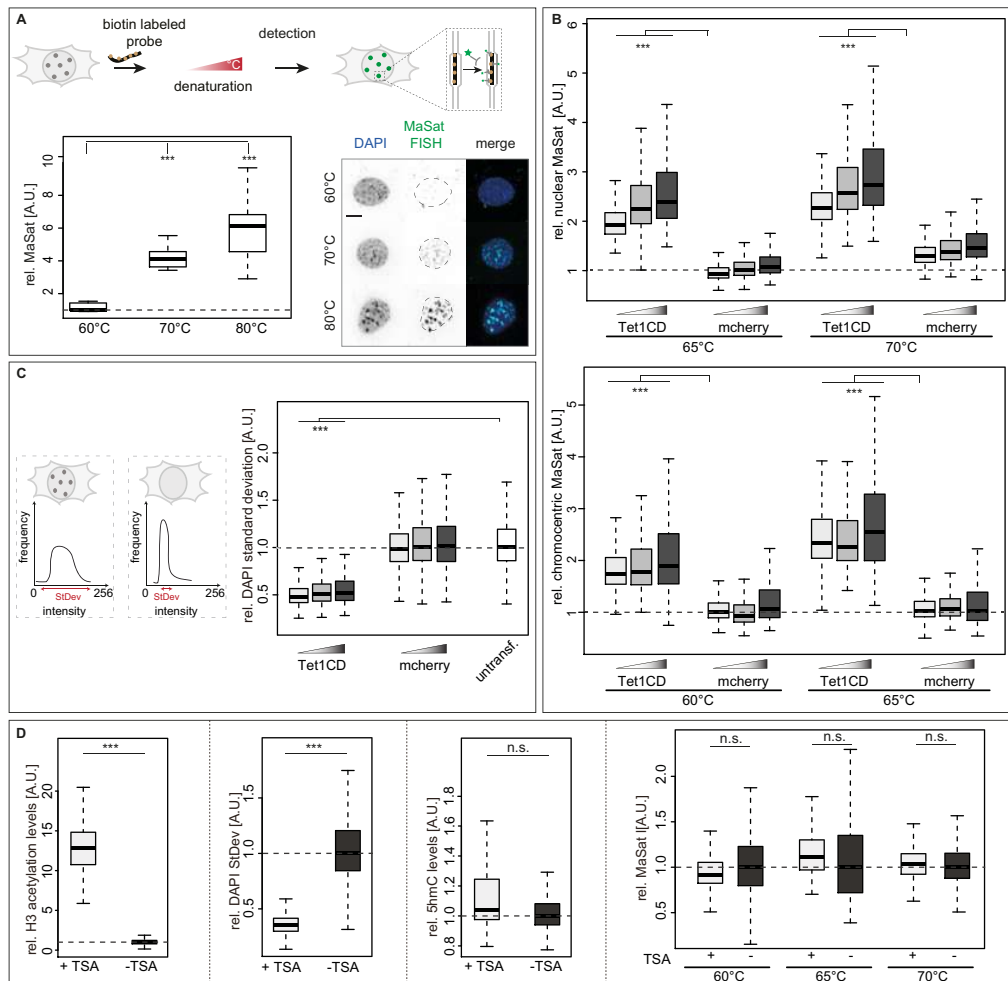
amplicons (**E**) and MINX PCR amplicons (**F**) containing 100% dC, d5mC, d5hmC, d5fC or d5caC, respectively (except primer sites), as determined by high resolution melting (HRM) temperature analysis. (**G**) Modified cytosine bases with the respective writer enzymes are depicted. High resolution melting temperature (HRM) analysis of 20 bp oligonucleotides containing one central (modified) CpG dinucleotide (CpG, 5fCpG or 5caCpG). Independent experiments were repeated in triplicates with four technical replicates (**E**, **F** and **G**), error bars represent standard deviation and *p* and *n*-values are summarized in Supplementary Table 8. Complete PCR conditions are indicated. M stands for DNA size markers. All quantifications of band intensities are depicted as a ratio to the respective controls. Full gels and blots are shown in Supplementary Figure 7. \* *p* < 0.05.

Rausch et al.



**Supplementary Figure S9 – Graphical summary of the strategies and outcomes for manipulation of cytosine levels in cells.** Amino acid coordinates are indicated below every structure. pc numbers refer to the plasmid collection numbers. **(A)** Overexpressing constructs with mcherry tagged Tet1 catalytic domain lead to increased 5mC levels. **(B)** Genomic knock-outs of Dnmt1 (MEF PM) or Dnmt1, Dnmt3a and Dnmt3b (J1 TKO) lead to the loss of all DNA cytosine modifications. **(C)** Transfection of cells with methylated or hydroxymethylated cytosine nucleotides lead to increased genomic 5mC and 5hmC levels, respectively.

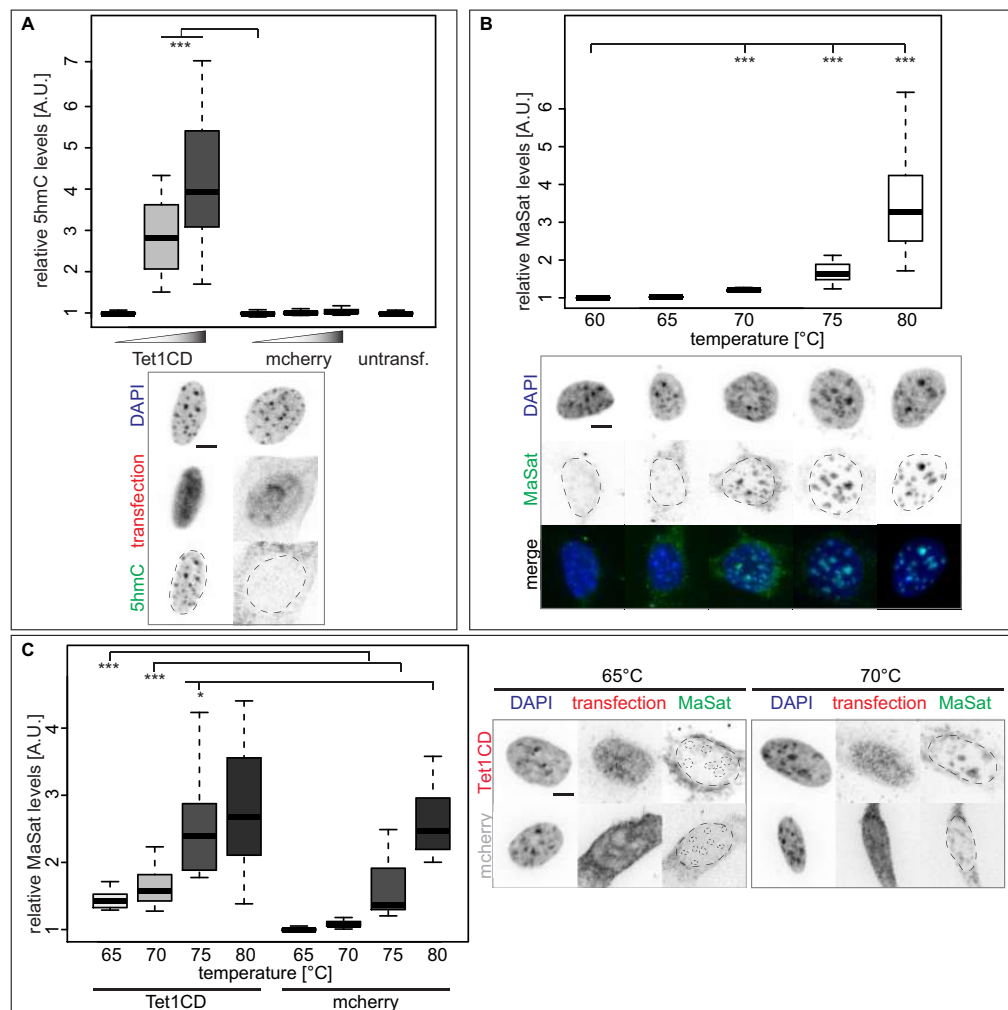
Rausch et al.



**Supplementary Figure 10 – Effect of cytosine modifications on genomic DNA double helix stability *in vivo*.** (A) Experimental setup to determine the optimal denaturing temperatures for major satellite DNA FISH assays. Relative major satellite DNA FISH sum intensities of C2C12 cells denatured at 60°C, 70°C or 80°C and representative spinning disk confocal images for every denaturing temperature are shown. (B) Relative major satellite DNA FISH sum intensities on transfected C2C12 mouse myoblast cells are plotted. DNA denaturation was done at 60°C, 65°C and 70°C, respectively. MaSat signal intensities were measured in the whole nucleus (defined by DAPI counterstaining, up) and in pericentric heterochromatin (chromocenters, down). (C) On a histogram, condensed chromatin is characterized by high DAPI intensities. Thus, cell nuclei containing condensed chromatin show a higher DAPI standard deviation than cells containing decondensed chromatin, which is characterized by low DAPI standard deviations. Normalized DAPI standard deviation values, as a proxy for DNA decondensation levels, 24 hours post transfection of C2C12 cells are plotted. (D) Relative histone H3 acetylation sum intensities, DNA decondensation, 5hmC and major satellite DNA FISH sum intensities in C2C12 cells after 72 hours of TSA treatment. Independent experiments were done in duplicates (A) or in triplicates (B, C and D). All fluorescent signals are plotted as a ratio to the values of the respective control cells. All boxes and whiskers represent 25–75 percentiles and 1.5 times the IQD (interquartile distance), respectively and the center line depicts the median. *p* and *n*-values are summarized in Supplementary Table 8. Scale bar = 5  $\mu$ m. \*\*\* *p* < 0.001, n.s.: non-significant and CC: chromocenters.

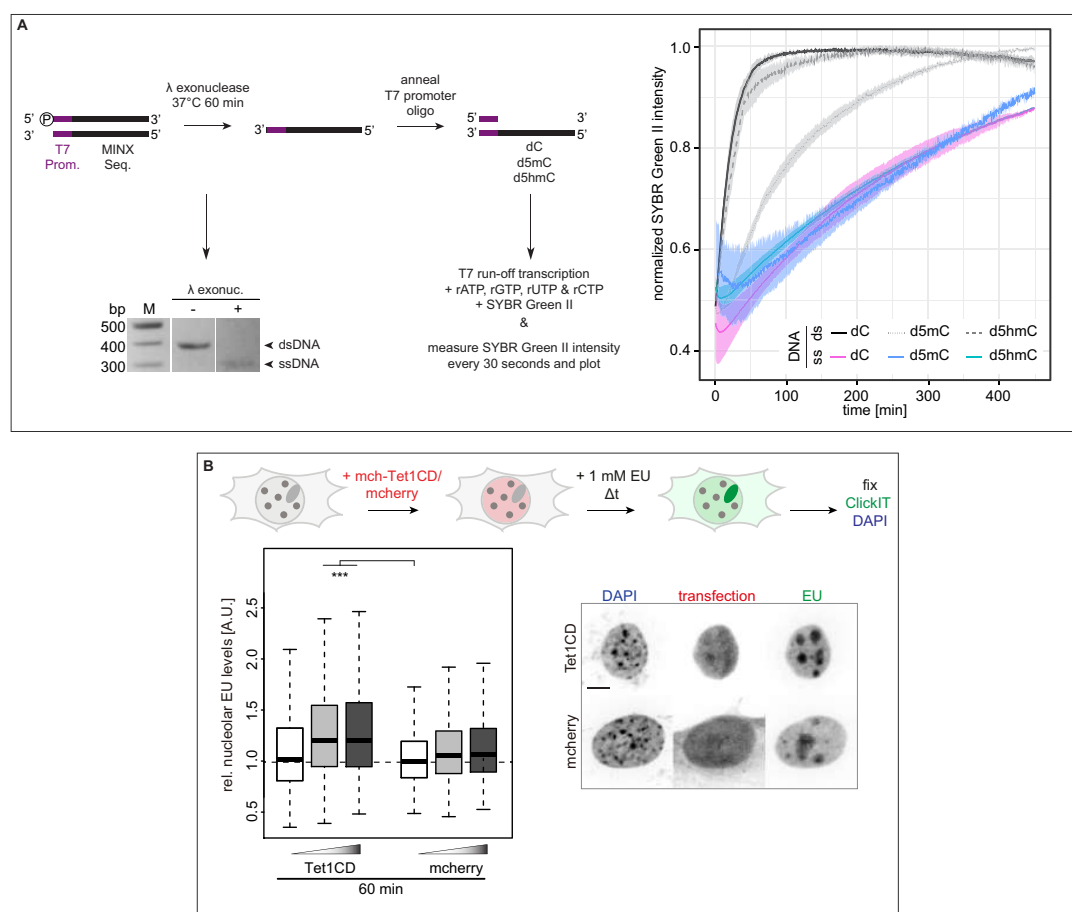
27

Rausch et al.



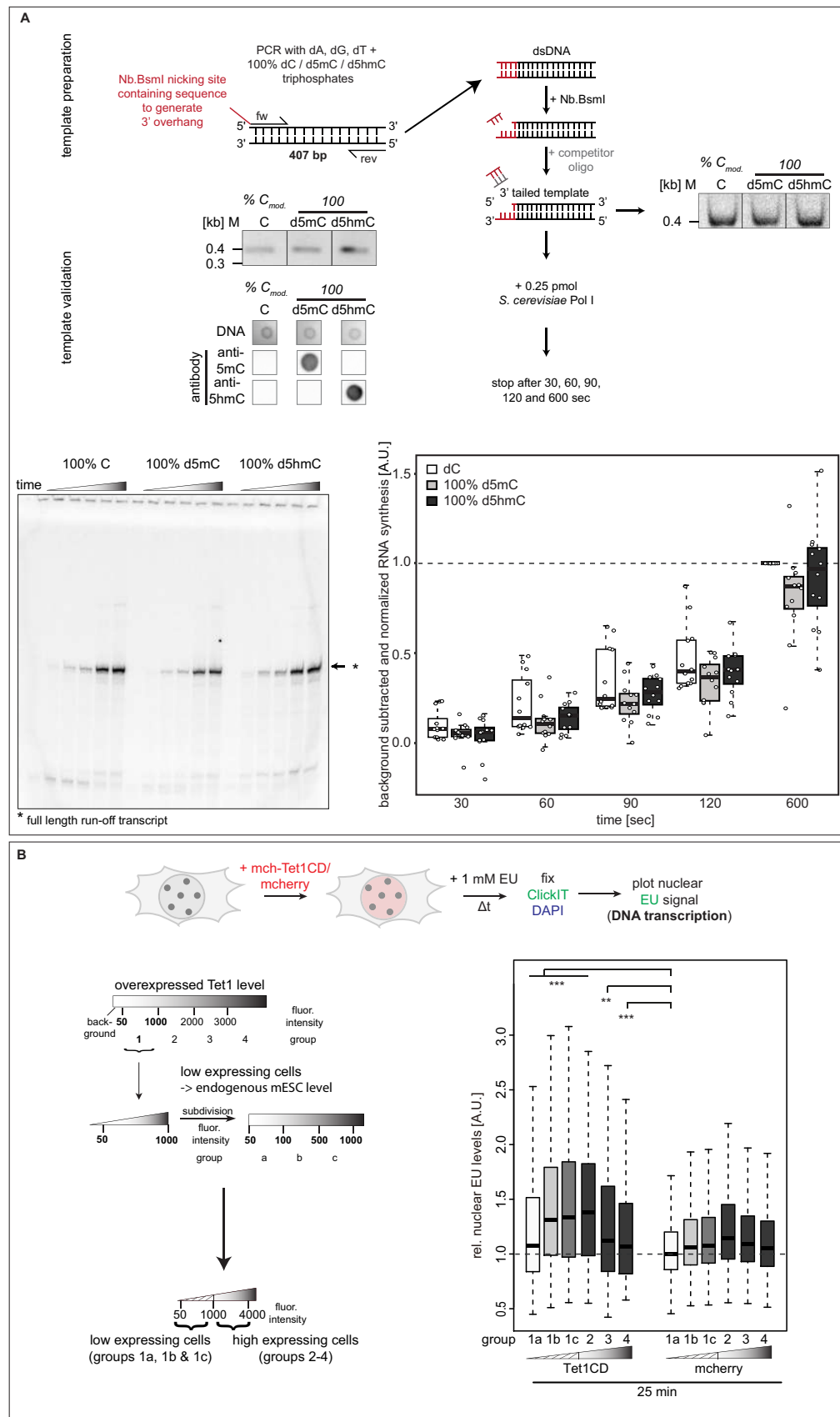
**Supplementary Figure 11 – Effect of cytosine modifications on DNA double helix stability in mouse embryonic fibroblasts *in vitro*.** (A) Relative 5hmC sum intensities in MEF cells, 24 hours after transfection and representative spinning disk confocal images for every transfection condition are shown. (B) Relative major satellite DNA FISH sum intensity signals of MEF cells with denaturing temperatures of 60°C, 65°C, 70°C, 75°C and 80°C and representative spinning disk confocal images for every denaturing temperature are depicted. (C) Relative nuclear major satellite DNA FISH sum intensities transfected mouse embryonic fibroblast cells with DNA denaturation at 65°C, 70°C, 75°C and 80°C, respectively are plotted. Representative spinning disk confocal images at 65°C and 70°C are shown. Independent experiments were done in triplicates. Dotted lines represent nuclear and chromocenter contours. All boxes and whiskers are as in Supplementary Figure 10 and  $p$  and  $n$ -values are summarized in Supplementary Table 8. Scale bar = 5  $\mu$ m. \*  $p < 0.05$  and \*\*\*  $p < 0.001$ .

Rausch et al.



**Supplementary Figure 12 – Effect of cytosine modifications on RNA polymerase activity *in vivo*.** (A) Experimental setup for *in vitro* T7 run-off transcription on double-stranded (dsDNA) and single-stranded (ssDNA). To generate ssDNA templates, dsDNA PCR amplicons with a 5' phosphorylated forward strand were digested with  $\lambda$  exonuclease. To allow polymerase binding, ssDNA templates contained a dsDNA T7 promoter region. Template DNA contained 100% cytosine (dC), methylcytosine (d5mC) or hydroxymethylcytosine (d5hmC) and T7 promoter sequence was unmodified. Normalized mean SYBR green II intensities ( $\pm$  standard deviation), i.e., the amount of RNA transcripts, over time are shown. (B) Schematic representation of the experimental setup for *in vivo* ribonucleotide incorporation analysis. Transfected cells were incubated with 1 mM EU 24 hours post transfection, EU was detected and nucleolar EU sum intensities are plotted. Representative spinning disk confocal images of every transfection condition are depicted. Independent experiments were done in triplicates. All boxes and whiskers are as in Supplementary Figure 10 and  $p$  and  $n$ -values are summarized in Supplementary Table 8. Scale bar = 5  $\mu$ m. \*\*\*  $p < 0.001$ .

Rausch et al.



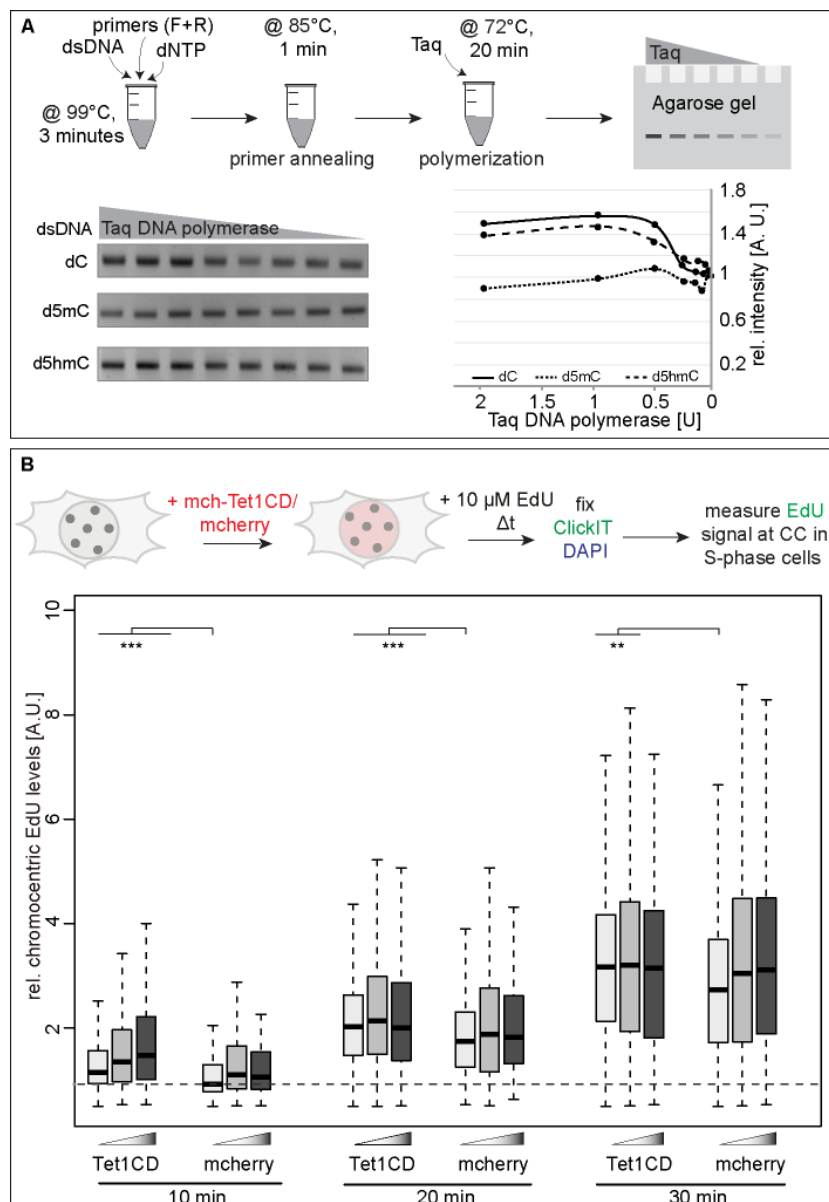
30

Rausch et al.

**Supplementary Figure 13 – Effect of cytosine modifications on eukaryotic RNA polymerase activity.**

**(A)** Experimental setup and biochemical validation of the template generation for *in vitro* polymerization assays. Template DNA contained 100% cytosine (dC) or methylcytosine (d5mC) or hydroxymethylcytosine (d5hmC), respectively (except primer sites). Exemplary agarose gel images and slot blot analysis of dC, d5mC and d5hmC levels are shown. Experimental setup for RNA polymerization assay using purified *S. cerevisiae* Polymerase I (Pol I). DNA templates were digested with the nicking enzyme Nb.BsmI to generate templates with a 3' overhang. Equal DNA amounts after nicking and DNA cleanups are confirmed by native PAGE gel analysis. An exemplary urea-PAGE gel image is shown. Background subtracted and normalized (to dC) full length run-off transcript intensities (400 bp, marked by \*) as a proxy for RNA polymerization are shown. **(B)** Schematic representation of the experimental setup for *in vivo* ribonucleotide incorporation analysis and representation of the cell grouping and binning approach applied. Cells were divided into four main groups, depending on their Tet1CD/mcherry expression levels (group 1: low expressing cells and groups 2-4: high expressing cells). Cells with Tet1 expression levels corresponding to group 1 (fluorescence intensity 50-1000, low Tet1 expressing cells) were further subdivided in three subgroups (fluorescent intensities group 1a: 50-100, group 1b: 100-500 and group 1c: 500-1000) during plotting. Cells were incubated with 1 mM EU 24 hours after transfection, EU was detected and nuclear EU sum intensities are plotted. Data from groups 1a-1a are the same as plotted in Figure 4B. Independent experiments were done with two DNA replicates, two different polymerase batches and in 12 replicates (A) or in triplicates (B) and all boxes and whiskers are as in Supplementary Figure 10. *p* and *n*-values are summarized in Supplementary Table 8. Full gels are shown in Supplementary Figure 7. \*  $p < 0.05$ , \*\*  $p < 0.005$  and \*\*\*  $p < 0.001$ .

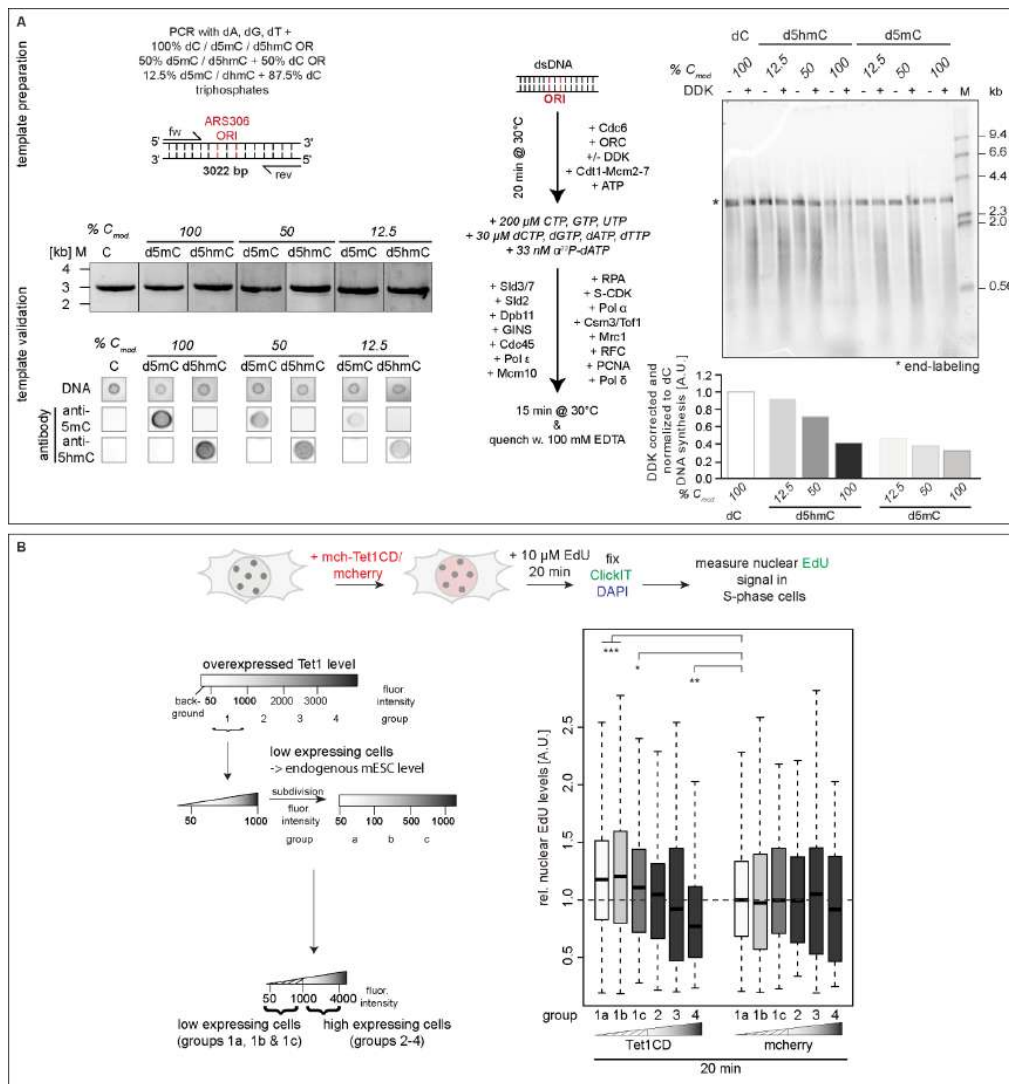
Rausch et al.



**Supplementary Figure 14 – Effect of cytosine modifications on DNA polymerase activity.** (A) Experimental setup for *in vitro* Taq polymerization assay. Template DNA contained 100% cytosine (dC), methylcytosine (d5mC) or hydroxymethylcytosine (d5hmC) (except primer sites). Exemplary gel images for Taq polymerase assay and sum band intensity quantification as a proxy for DNA polymerization, normalized to the DNA input (0 U enzyme) are shown. (B) Schematic representation of the experimental setup for *in vivo* nucleotide incorporation analysis. Cells were incubated with 10  $\mu$ M EdU 24 hours after transfection, EdU was detected and nuclear EdU sum intensities of S-phase cells are plotted. Independent experiments were done twice (A) or in triplicates (B) and all boxes and whiskers are as in Supplementary Figure 10. *p* and *n*-values are summarized in Supplementary Table 8. Full gels are shown in Supplementary Figure 7. \*\*\* *p* < 0.001.

32

Rausch et al.



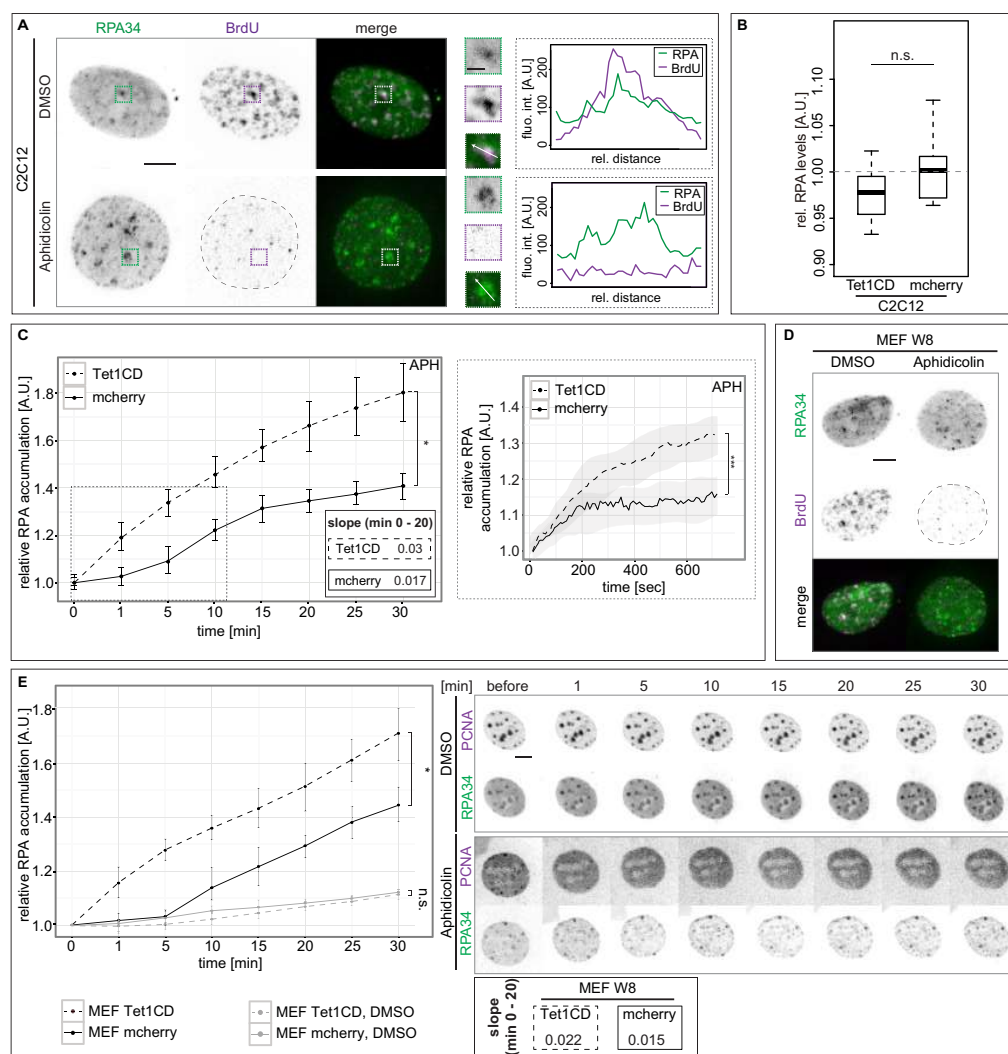
**Supplementary Figure 15 – Effect of various amounts of cytosine modifications on eukaryotic DNA polymerase activity. (A)** Experimental setup and biochemical validation of the template generation for *in vitro* polymerization assays. Template DNA contained 100% cytosine (dC) or 100%, 50% or 12.5% methylcytosine (d5mC) or hydroxymethylcytosine (d5hmC), respectively (except primer sites). The 3 kb templates also contained a centrally arranged yeast ARS306 origin of replication. Exemplary agarose gel images and slot blot analysis of dC, d5mC and d5hmC levels for every nucleotide composition are shown. Experimental setup for soluble replication assay using purified *S. cerevisiae* replisome components. Exemplary alkaline agarose gel images are shown and DDK subtracted and normalized replicated DNA signal intensities (< 3 kb end-labeling band (labeled with \*), normalized to dC) as a proxy for DNA polymerization are shown. **(B)** Schematic representation of the experimental setup for *in vivo* nucleotide incorporation analysis and representation of the cell grouping and binning approach applied. Cells were divided into four main groups, depending on their Tet1CD/mcherry expression levels (group 1: low expressing cells and groups 2-4: high expressing cells). Cells with Tet1 expression levels corresponding to group 1 (fluorescence intensity 50-1000, low Tet1 expressing cells) were further subdivided in three subgroups (fluorescent intensities group 1a: 50-100, group 1b: 100-500 and group 1c: 500-1000) during plotting. Cells were incubated with 10  $\mu$ M EdU 24 hours after transfection, EdU was detected and nuclear EdU sum intensities of S-phase cells are plotted.

33

Rausch et al.

Data from groups 1a-1a are the same as plotted in Figure 5B. Independent experiments were done twice (A) or in triplicates (B) and all boxes and whiskers are as in Supplementary Figure 10.  $p$  and  $n$ -values are summarized in Supplementary Table 8. Full gels are shown in Supplementary Figure 7. \*  $p < 0.05$ , \*\*  $p < 0.005$  and \*\*\*  $p < 0.001$ .

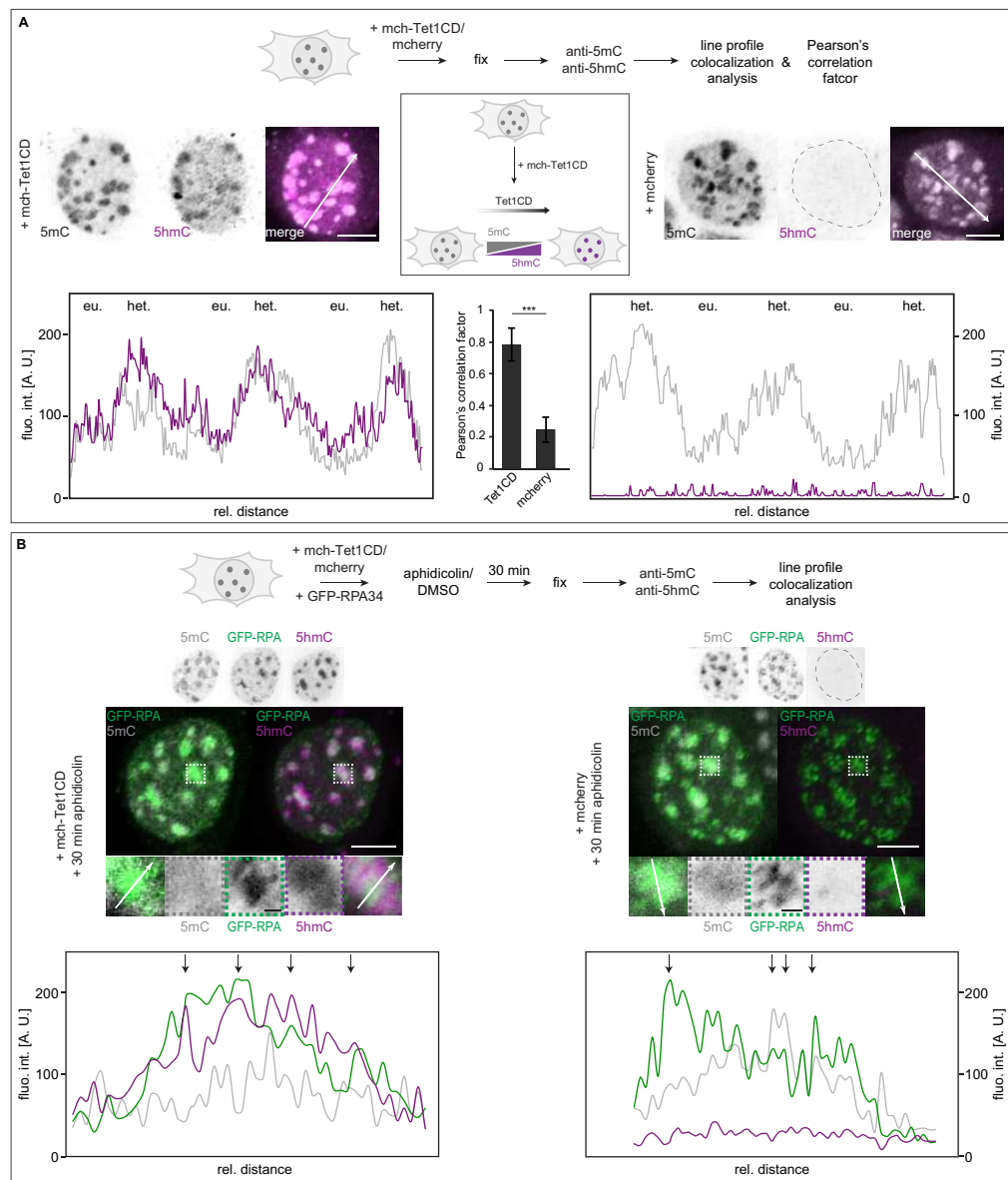
Rausch et al.



**Supplementary Figure 16 – Effect of cytosine modifications on DNA helicase activity *in vivo*.** (A) Representative spinning disk confocal images showing the localization of GFP-RPA and the (lack of) BrdU incorporation in late C2C12 S-phase cells after DMSO or aphidicolin treatment. Dashed boxes represent the selected magnified ROIs for line intensity profile analysis (arrows). Line intensity plots of BrdU (magenta) and RPA (green) through a replicating chromocenter in late C2C12 S-phase cells before and 30 minutes after aphidicolin/DMSO addition are shown. (B) Relative GFP-RPA sum intensities 24 hours after C2C12 triple transfection, showing no significant difference of RPA levels within the different transfection conditions. (C) Graphs showing the normalized average RPA accumulation ( $c_v \pm$  standard deviation) over 30 minutes at 5 minutes interval imaging (left) and over 12 minutes at 10 seconds interval (right) for aphidicolin treated C2C12 S-phase cells, respectively (as in Figure 6B). Slopes of the RPA accumulation were calculated as in Figure 6B. (D) Representative spinning disk confocal images of BrdU incorporation in aphidicolin and DMSO treated MEF W8 cells. (E) Analysis, representation and slopes of RPA accumulation for triple transfected MEF W8 cells are as in (B). Representative spinning disk confocal images of time-lapse microscopy upon aphidicolin treatment of showing RPA and PCNA localization in aphidicolin and DMSO treated MEF cells. Scale bars = 5  $\mu$ m (whole nuclei images) and 2.5  $\mu$ m (magnifications). All boxes and whiskers are as in Supplementary Figure 10 and  $p$  and  $n$ -values are summarized in Supplementary Table 8. \*  $p < 0.05$  and n.s.: non-significant.

35

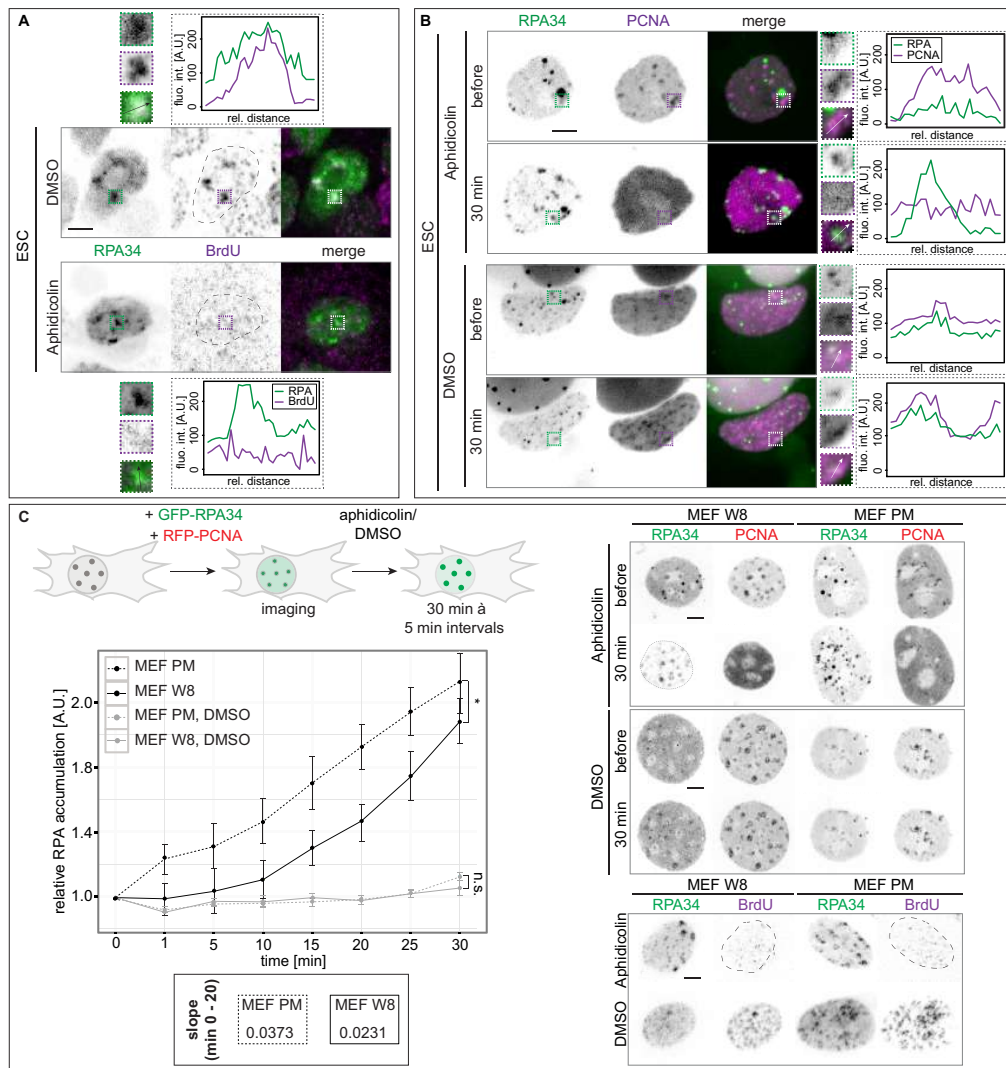
Rausch et al.



**Supplementary Figure 17 – Colocalization analyses of cytosine modifications and DNA replication in mcherry-Tet1CD/mcherry transfected cells. (A)** 24 hours after mcherry-Tet1CD/mcherry transfection, cells were fixed and co-stained for 5mC (grey) and 5hmC (magenta). Line profile analyses and Pearson's correlation factors (mean  $\pm$  StDev) of 5mC and 5hmC are shown. Line profiles are measured along the white arrow depicted in the images. **(B)** 24 hours after mcherry-Tet1CD/mcherry and GFP-RPA double transfection, cells were treated for 30 minutes with aphidicolin or DMSO, fixed and co-stained for 5mC and 5hmC. Line profile analyses of late S-phase for 5mC (grey), 5hmC (magenta) and GFP-RPA (green) are shown for each condition. Line profiles are measured along the white arrow depicted in the images. Black arrows in the plots highlight regions of high GFP-RPA and 5mC and/or 5hmC colocalization. Scale bar = 5  $\mu$ m (whole nuclei) and 2.5  $\mu$ m (magnifications). Dotted lines represent nuclei contours. eu.: euchromatin, het.: heterochromatin and \*\*\*  $p < 0.001$ .

36

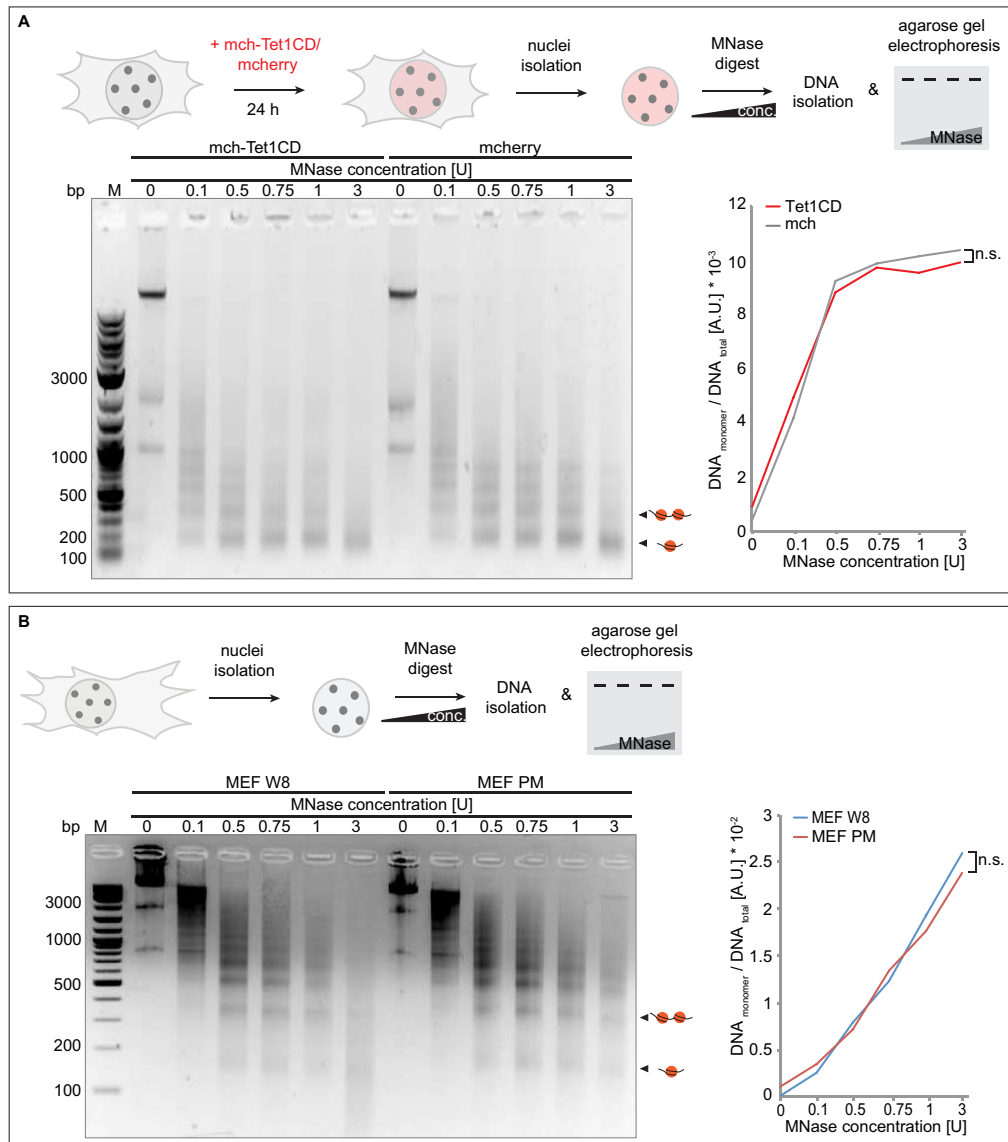
Rausch et al.



**Supplementary Figure 18 – Effect of loss of cytosine methylation on DNA helicase activity *in vivo*.** (A) Representative spinning disk confocal images showing the localization of GFP-RPA and the (lack of) BrdU incorporation in stage II J1 ES cells after aphidicolin/DMSO treatment. Dashed boxes represent the selected magnified ROIs for line intensity profile analysis (white arrows). Line intensity plots of BrdU and RPA are as described in Figure 6B. (B) Representative spinning disk confocal time-lapse microscopy images of aphidicolin and DMSO treated J1 cells as in Figure 6B. Dashed boxes represent the selected magnified ROIs for line intensity profile analysis (arrows) before and 30 minutes after aphidicolin/DMSO addition. (C) MEF W8 and MEF PM cells were co-transfected with GFP-RPA and RFP-PCNA 24 hours before imaging and imaging and analysis of RPA accumulation and slopes of double transfected MEF cells as in Figure 6B. Representative spinning disk confocal images before and after 30 minutes of aphidicolin/DMSO treatment showing RPA localization and BrdU incorporation after 30 minutes of aphidicolin/DMSO treatment are shown. Dashed lines represent nuclear contours. Scale bars = 5  $\mu$ m. All boxes and whiskers are as in Supplementary Figure 10. *p* and *n*-values are summarized in Supplementary Table 8. \* *p* < 0.05 and n.s.: non-significant.

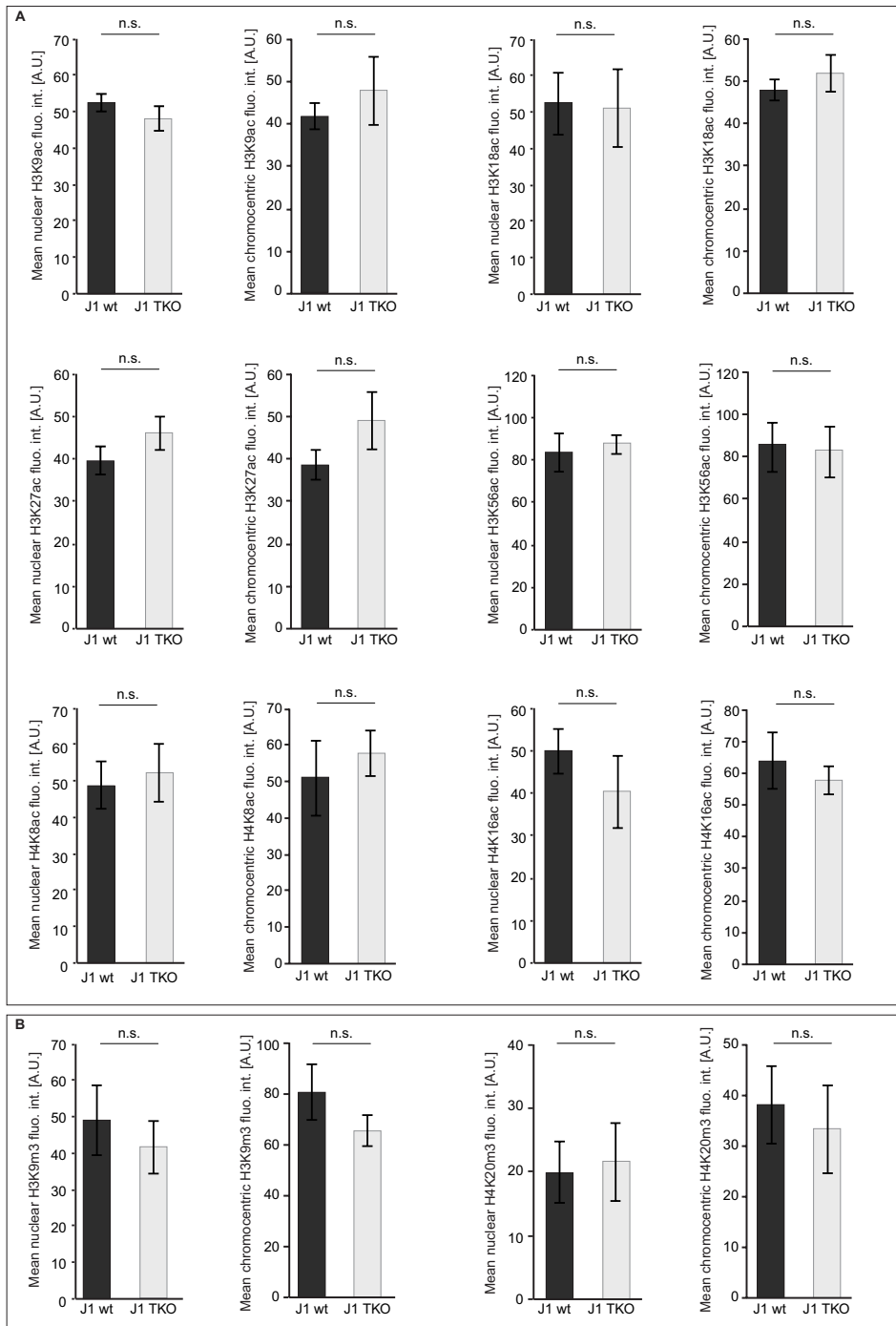
37

Rausch et al.



38

Rausch et al.



**Supplementary Figure 20 – Effect of loss of global DNA methylation on histone modifications in ES cells. (A-B)** Mouse J1 and TKO ES cells were pulse labelled with EdU, fixed and EdU was detected in combination with histone modifications. The nuclear and chromocentric levels of histone acetylation (**A**, H3K9ac, H3K18ac, H3K27ac, H3K56ac, H4K8ac and H4K16ac) and methylation (**B**, H3K9m3 and H4K20m3) were measured in non-replicating (EdU negative) cells. Error bars represent standard deviation. Independent experiments were done in duplicates. *p* and *n*-values are summarized in Supplementary Table 8. n.s.: non-significant.

39

Rausch et al.

## REFERENCES

1. Weber, P., Rausch, C., Scholl, A. and Cardoso, M.C. (2019) Repli-FISH (Fluorescence in Situ Hybridization): Application of 3D-(Immuno)-FISH for the study of DNA replication timing of genetic repeat elements. *OBM Genetics*, **3**.
2. Zhang, P., Hastert, F.D., Ludwig, A.K., Breitwieser, K., Hofstätter, M. and Cardoso, M.C. (2017) DNA base flipping analytical pipeline. *Biology Methods and Protocols*, **2**, 1-12.
3. Zhang, P., Ludwig, A.K., Hastert, F.D., Rausch, C., Lehmkuhl, A., Hellmann, I., Smets, M., Leonhardt, H. and Cardoso, M.C. (2017) L1 retrotransposition is activated by Ten-eleven-translocation protein 1 and repressed by methyl-CpG binding proteins. *Nucleus*, **8**, 548-562.
4. Merkl, P.E., Pils, M., Fremter, T., Schwank, K., Engel, C., Längst, G., Milkereit, P., Griesenbeck, J. and Tschochner, H. (2020) RNA polymerase I (Pol I) passage through nucleosomes depends on Pol I subunits binding its lobe structure. *Journal of Biological Chemistry*, **295**, 4782-4795.
5. Ludwig, A.K., Zhang, P., Hastert, F.D., Meyer, S., Rausch, C., Herce, H.D., Müller, U., Lehmkuhl, A., Hellmann, I., Trummer, C. et al. (2017) Binding of MBD proteins to DNA blocks Tet1 function thereby modulating transcriptional noise. *Nucleic Acids Research*, **45**, 2438-2457.
6. Sporbert, A., Domaing, P., Leonhardt, H. and Cardoso, M.C. (2005) PCNA acts as a stationary loading platform for transiently interacting Okazaki fragment maturation proteins. *Nucleic acids research*, **33**, 3521-3528.
7. Sporbert, A., Gahl, A., Ankerhold, R., Leonhardt, H. and Cardoso, M.C. (2002) DNA polymerase clamp shows little turnover at established replication sites but sequential de novo assembly at adjacent origin clusters. *Mol Cell*, **10**, 1355-1365.
8. Zillmann, M., Zapp, M.L. and Berget, S.M. (1988) Gel electrophoretic isolation of splicing complexes containing U1 small nuclear ribonucleoprotein particles. *Mol Cell Biol*, **8**, 814-821.
9. Yaffe, D. and Saxel, O. (1977) Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature*, **270**, 725-727.
10. Peters, A.H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schofer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A. et al. (2001) Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell*, **107**, 323-337.
11. Lande-Diner, L., Zhang, J., Ben-Porath, I., Amariglio, N., Keshet, I., Hecht, M., Azuara, V., Fisher, A.G., Rechavi, G. and Cedar, H. (2007) Role of DNA methylation in stable gene repression. *J Biol Chem*, **282**, 12194-12200.
12. Graham, F.L., Smiley, J., Russell, W.C. and Nairn, R. (1977) Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *The Journal of general virology*, **36**, 59-74.
13. Li, E., Bestor, T.H. and Jaenisch, R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, **69**, 915-926.
14. Tsumura, A., Hayakawa, T., Kumaki, Y., Takebayashi, S., Sakaue, M., Matsuoka, C., Shimotohno, K., Ishikawa, F., Li, E., Ueda, H.R. et al. (2006) Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells*, **11**, 805-814.
15. Rottach, A., Kremmer, E., Nowak, D., Leonhardt, H. and Cardoso, M.C. (2008) Generation and characterization of a rat monoclonal antibody specific for multiple red fluorescent proteins. *Hybridoma*, **27**, 337-343.



## 6 General Conclusions and Perspectives

---

### 6.1 Replication of the Mammalian Genome – Repeat Elements

---

#### 6.1.1 DNA Repeat Elements Are Well Integrated Into the Mammalian Genome

In chapter I of this thesis, we developed a protocol to analyze replication of DNA repeat elements in time and space in structurally intact 3D cell nuclei (Repli-FISH). While genome-wide sequencing studies of nascent DNA can identify the exact underlying DNA sequence that is replicated, the resolution in time and space is limited. Although the rapidly evolving technical methods enable single-cell profiling by now [269], most studies are based on population-wide analyses, thereby possibly averaging and/or out-diluting single-cell heterogeneity. Additionally, repeat elements are commonly excluded from genome-wide studies, since their repetitive nature makes them difficult to map [270]. Our protocol combines fluorescence *in situ* hybridization (FISH) and visualization of DNA replication (Repli-FISH), either by detecting nascent DNA via nucleotide analogs or by analyzing replisome components via immunofluorescence. Additionally, applying different chase durations between pulse labeling and fixation results in different degrees of replication foci separation and, with increasing chase times, eventually also to transition from one S-phase pattern to the next. Hence, this approach of varying chase times even allows the visualization of a temporal resolution of S-phase progression in fixed cells. The development of sensitive microscopy filters allows the spectral separation of multiple fluorochromes, thereby enabling the detection of several replication signals and repeat elements/DNA sequences (multi-color FISH) per sample. Additionally, the FISH pretreatment procedures preserve the structural integrity of the cells and nuclei without the need of several cycles of freezing and thawing in liquid nitrogen [271]. Hence, the spatial arrangements of different repeat elements and DNA replication signals can be analyzed at different optical resolutions, ranging from wide-field to confocal and even to super-resolution. Similarly, detailed analysis on the spatio-temporal interrelationships can be performed on single nuclei, however, population studies and high-throughput investigations are also feasible.

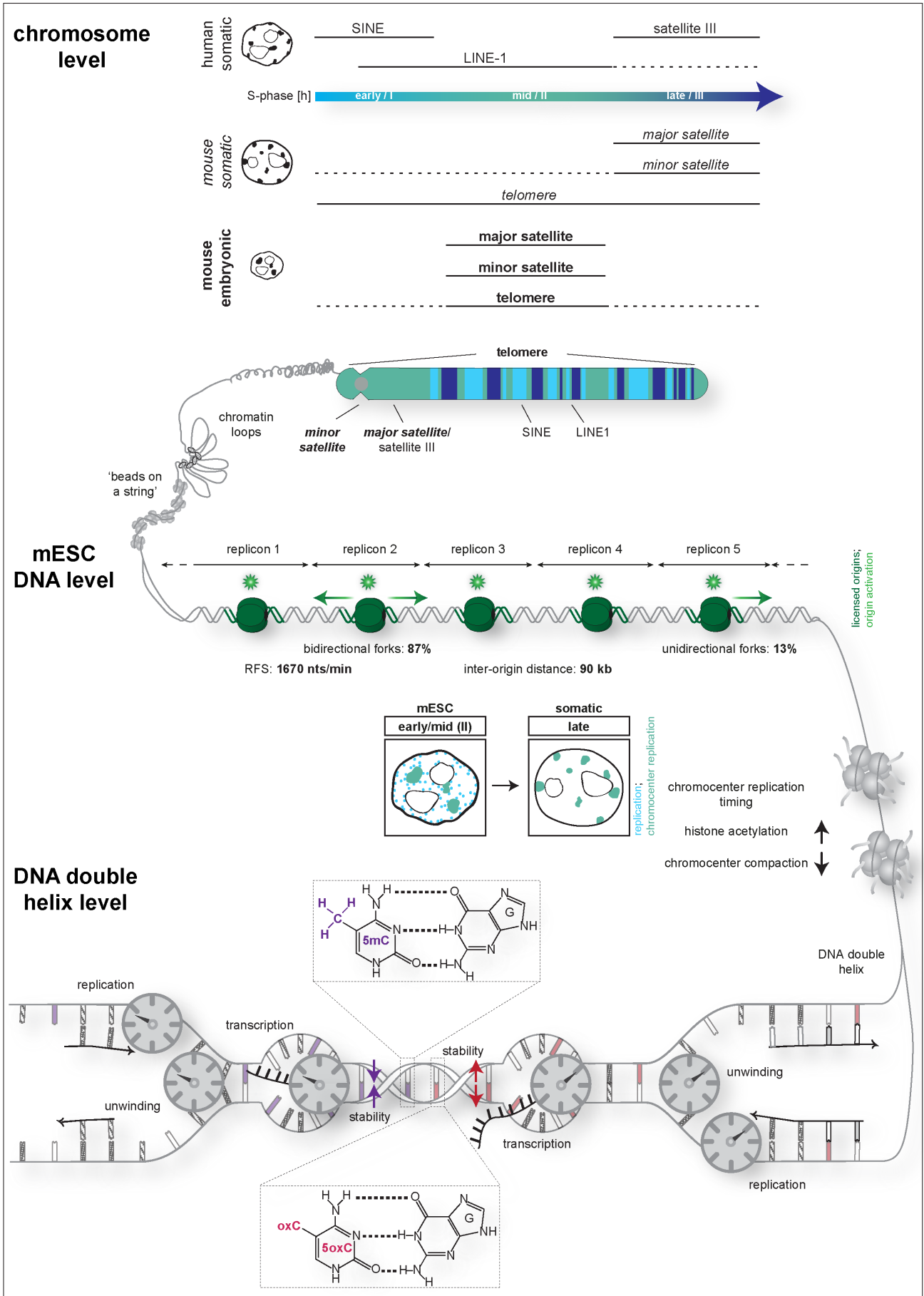
Complete and error-free replication of all (epi)genetic material is of prodigious importance for cells to preserve genome integrity and stability. This holds true not only for the small protein-coding part of the genome but also applies to the non-coding and repetitive fractions of the genome. Opposed to some lower eukaryotes, the majority of the mouse and human genomes consists of non-coding and repetitive DNA. For long, two contradictory hypotheses concerning the *raison d'être* of these elements in the genome persisted. On one side, repetitive DNA was seen as parasitic and selfishly replicating 'junk' DNA with no distinct function in the nuclear metabolism [18, 19]. On the other side, distinct functions in genome organization and maintenance were attributed to repeat elements. The latter are nowadays believed to be involved in the formation of distinctive eu- and heterochromatin compartments [22–26]. In chapters I, II and III, we, therefore, made use of the Repli-FISH protocol to analyze the replication timing of LINE-1, Alu and satellite III in human cervix

carcinoma (HeLa) cells (chapter II) and of major and minor satellites, as well as telomeres in mouse somatic (C2C12 & MEF, chapter I) and pluripotent cells (J1 mES cells, chapter III).

In chapter II, we comprehensively analyzed and connected the replication timing, the double-strand break repair kinetics and the underlying chromatin marks of repetitive DNA elements. In addition, we integrated publicly available genomics data (Chip-Seq and Repli-Seq) and our combination of single-cell FISH and DNA replication/repair correlations. Interestingly, and arguing against the 'junk' DNA theory, repetitive elements appear to be efficiently integrated into the human genome and are duplicated and repaired by the cell's DNA replication and repair machineries. Yet, the accuracy of this maintenance seems to strongly depend on the chromatin type the repeat elements are embedded in. Indeed, mutation frequency analyses revealed that late replicating and generally non-coding DNA displays higher frequencies of point mutations, probably due to a downregulation of DNA mismatch repair proteins during mid S-phase [181, 182, 272]. Nevertheless, we found a dependence on the chromatin state and its structural organization and the timing of repeat element maintenance by DNA replication and repair. While the euchromatic human Alu element was replicated early in S-phase and repaired fast after DNA damage induction, the heterochromatic satellite III repeat elements were duplicated late in S-phase and repaired more slowly (Figure 6.1). Coherently, these observations are also reflected in previous studies, showing earlier replication and faster repair of euchromatin compared to heterochromatic DNA [176, 273, 274]. LINE-1, which was previously shown to correlate with late replicating regions in genome-wide studies [176], was replicated throughout S-phase in our single-cell based study. Integration of genome-wide HeLa chromatin landscape data revealed that LINE-1 elements are generally replicated according to their particular histone marks. This finding underlines the importance of combining genome-wide sequencing and single-cell microscopy analyses (Figure 6.1).

### 6.1.2 mES Cells Are Characterized by a Distinct Replication Program

Similarly to the replication timing analysis of repeat elements in human HeLa cells, we investigated when during S-phase minor and major satellite repeats and telomeres are replicated in mouse embryonic stem (mES) cells (chapter III). We demonstrate that replication of major and minor satellite repeats in mES cells is temporally closely connected, reflecting their close spatial proximity on the linear chromosome. Interestingly, centromeres were proposed to play a role in regulation of the activation time of *cis*-neighboring replication origins in yeast [275]. Our single-cell microscopic analysis showed telomere replication throughout S-phase, with a prevalence in early/mid S-phase. Importantly, telomeres capping the short arm of the mouse acrocentric/telocentric chromosomes were preferentially replicated in conjunction with (peri)centromeric repeats. These observations are in strong favor of a domino-like replication propagation. In this next-in-line model, replication initiation within a replicon cluster at the beginning of S-phase creates replication bubbles subsequently leading to the firing of origins of replication in neighboring replicon clusters, thereby spreading replication onset in *cis* and *trans* along the genome [192]. Although we did not specifically address a correlation between the 3-dimensional nuclear localization and the replication timing of telomeres, such relationships were previously observed [276]. Interestingly, yeast normal-length telomeres were shown to replicate during late S-phase, while DNA replication of shortened telomeres initiated already early in S-phase. The latter correlated with high telomerase activity and, consequently, a subsequent increase in telomere length [277]. The predominant early/mid replication of telomeres in mES cells may, therefore, represent a similar



**Figure 6.1: Scheme summarizing the main conclusions from chapters I-IV.** We show, by a variety of methods, different levels of epigenetic regulation of essential nuclear metabolic processes, ranging from the DNA double helix level up to higher order chromatin structures.

'time buffer' for telomerase activity [278]. Due to the pluripotent nature of mES cells and their ability to generate cells of all three germ layers, intact and long telomeres are essential for successful cell differentiation and organism development. The early onset of chromosome end replication would, consequently, provide enough time to allow telomere length extension/maintenance by the telomerase enzyme at the embryonic stage, thereby guaranteeing sufficient and conscientious end-capping.

Strikingly, (peri)centromeric heterochromatin was replicated during early/mid S-phase in mES cells, which is in sharp contrast to the late replication timing commonly observed in somatic cells. However, a comparable replication timing shift of the compact satellite DNA structures was observed in early *Drosophila* embryos coming along with extremely fast replication. Once satellite replication got delayed, the complete S-phase duration was prolonged. This transition came along with heterochromatin restructuring, resulting in 'mature' chromocenters, and required active DNA transcription [279]. Likewise, we examined chromocenter features of mES cells and compared them to differentiated cells. Although DAPI-intense and compact pericentromeric clusters could clearly be distinguished in pluripotent mouse cells, detailed analysis showed significant structural differences to chromocenters in differentiated cells. While the latter were characterized by a high number of regularly shaped and compacted chromocenters of smaller size, pluripotent stem cells were defined by less but bigger, irregularly shaped and decondensed pericentromeric heterochromatin clusters. Similarly to our observations, previous electron spectroscopic imaging (ESI) studies of mES cells revealed lower heterochromatin packaging densities coming along with poorly delimited chromocenter boundaries. The reduced chromocenter compartmentalization was further defined by a high prevalence of 10-nm chromatin fibers in heterochromatin. Interestingly, the authors also predominantly found this basic 'beads-on-a-string' structure in heterochromatic DNA of mouse embryonic fibroblasts (MEF), spleen lymphocytes and liver cells [280–283]. Although these findings challenge the conventional chromatin packaging model based on the further folding of the 10-nm chromatin fiber to the 30-nm fiber, they support the evidence that the basic structural chromatin unit of the 30-nm fiber was, until now, exclusively observed *in vitro* or in very specialized cells, like the *Patiria miniata* (starfish) sperm [284, 285]. Additional evidence from X-ray scattering, cryo-electron tomography and stochastic optical reconstruction microscopy (STORM) emphasize the lack of the 30-nm fiber in the nucleus and reinforce the fact that, *in vivo*, chromatin adopts irregular shapes that resemble liquid droplets [286]. In line with our findings of a more open mES cell chromatin, fluorescence recovery after photobleaching (FRAP) experiments showed an increased fraction of soluble or loosely bound chromatin architectural proteins such as HP1, the linker histone H1 and core histones H2B and H3 in mES cells [287]. This 'breathing' conformation generates a very active chromatin and allows fast exchange of structural proteins, thereby enabling rapid global reorganization upon differentiation onset. Consequently, the open and mobile chromatin architecture is supposed to be a hallmark of pluripotent cells. It would, therefore, be interesting to further analyze and define the different structure, organization and proteome of mES cell pericentromeric heterochromatin. To analyze potential differences in the mES cell chromocenter protein composition, mass-spectrometry based approaches could be envisaged. We already adapted a protocol to isolate chromocenter fractions from different tissues [288]. This technique could be further modified to compare chromocenter composition of pluripotent and differentiated cells, and possibly also specific knock-out cells with their wildtype counterpart.

Previous reports have shown a clear relationship between histone acetylation and chromatin decompaction [? ]. Accordingly, we also measured chromocentric histone acetylation levels and found significant increases in acetylated H3 and H4 lysines compared to differentiated cells. This different chromatin compo-

sition and organization and the concomitant more open structure may be at the basis for providing better access to a plethora of proteins, including origin activation factors. Similarly, global DNA transcription is highly upregulated in mES cells, and especially normally silenced repeat elements show higher transcript amounts [289]. Akin, histone hypoacetylation and maintenance of late replication timing were tightly linked in previous studies [209, 212, 214–216]. Consistently our histone deacetylase HDAC1 targeting to chromocenters resulted in a replication timing shift of the latter to late S-phase (Figure 6.1). However, how exactly and at which step of DNA replication (origin licensing versus origin firing) histone acetylation regulates early onset of mES cell chromocenter replication needs to be further unveiled. Interestingly, TICRR/Treslin was recently found to interact with acetyl-histone binding proteins, thereby revealing a potential mechanism to regulate DNA replication timing [202]. Additionally, future work will be necessary to clearly define individual chromatin classes [290] in mES cells by distinct epigenetic marks and to correlate these with the 'unusual' DNA replication progression presented in chapter III.

### 6.1.3 Replication of the Y Chromosome Marks the End of S-Phase

**N**ext, we identified the Y chromosome as a specific hallmark for the end of S-phase. Independently of the pluripotency state of the male cell, Y chromosome replication was followed by transition to the G2 phase. Interestingly, the sex chromosome was replicated synchronously and in a relatively short period of time, similarly to the inactivated X (Xi) in female cells [291]. The Y chromosome is marked by a very low gene density (1.7 genes/Mb for mice, [292–294]), of which most show testis-specific expression or have an X encoded paralog [293, 295]. Additionally, the Y chromosome is composed of large heterochromatic DNA blocks and gene expression and transcriptional activity are low [296, 297]. A connection between low gene density and transcriptional activity could indeed explain the late replication timing of the male sex chromosome, since a correlation between gene expression and replication timing was already demonstrated in *Drosophila* [199]. Similarly, a relation between replication timing, GC content, gene density and transcriptional activity was revealed in human cells [176, 298–300]. However, akin to early assumptions that the seemingly 'late' replication timing of the Xi in female cells [301–304] is an important factor to maintain the silenced state of the Xi [304], the late replication of the Y may also be essential to keep it in a transcriptional inert state. Additionally, the synchronous Y chromosome replication, i.e. the simultaneous firing of origins of replication, bears resemblance to observations made for the replication of the female Xi [291]. Moreover, the highly coordinated replication of the Y chromosome is similar to observations in early *Drosophila* and *Xenopus* embryos [252, 254], where replication also takes place in the absence of DNA transcription. Interestingly, in female somatic cells, the pericentromeric heterochromatin region of the Xi is not replicated together with the bulk of the inactive chromosome [291]. If the same phenomenon holds true for the Y chromosome is still open. If so, unveiling the underlying mechanism distinguishing between silenced heterochromatic and bulk inactive chromatin would shed light on different replication organizations in silenced parts of the genome. Thereover, early replicating and indispensable housekeeping genes are more conserved than cell-type specific genes that get replicated later during S-phase [305, 306]. Considering the high testis specificity of most Y chromosome encoded genes [293, 295], late Y replication fits into this general observation. There is evidence that duplication clusters, resulting from gene amplification and recombination events, occur more frequently in the late replicating genome fraction [307, 308]. The late replication timing of the mouse Y chromosome

may, therefore, also be reflected in the massive amounts of amplified DNA sequences present within the male-specific region of the Y chromosome (MSY, [309]). Thus, late replicating DNA tends to have increased genetic variability and instability, in line with theories of a small and degenerated 'junk DNA' chromosome (reviewed in [310]). Future experiments determining specific features of Xi and Y chromosome replication, for example, replication foci number and inter-origin distance, would further underline and highlight a potential universal replication program for inactive chromosomes. In addition, detailed investigations are needed to dismantle the mechanism by which the late replication timing is determined and conserved for a complete chromosome. Promising candidates may be specific epigenetic signatures, as well as the replication timing regulatory factor 1 (RIF1), which is a known genome-wide regulator of DNA replication timing (reviewed in [311]). Importantly, synchronous replication of the transcriptionally silenced Xi was shown to heavily rely on hypoacetylation [291]. Similarly, specific chromatin features responsible for the synchronous late origin firing on the Y chromosomes need to be further unveiled.

---

## 6.2 Replication of the Mammalian Genome – Origin Activation and Fork Speed

---

### 6.2.1 Origin Activation in Pluripotent Mouse Cells Differs From That in Somatic Cells

**D**ramatic differences in replication organization were shown in *Drosophila* and *Xenopus* early embryonic development, where genome duplication is performed in extraordinarily short time frames and the complete absence of transcription [252, 254]. Our data presented in chapter III provides evidence that a similar mechanism exists in early stages of mouse embryo development. However, in contrast to *Xenopus* and *Drosophila*, transcription is not completely absent in the early stages of mouse embryonic development. Instead, transcriptional initiation is already observed in the pronuclei of the zygote and at the 2-cell stage in mice [312, 313]. Nonetheless, we showed by a combination of super-resolution microscopy and DNA combing that mES cell genome replication relies on more activated origins of replication than actually ‘needed’. To further dissect whether this observation is based on a different origin licensing or solely on the subsequent origin activation, future work could compare the levels of essential replisome components and regulators between pluripotent and differentiated cells. For example, the amounts of chromatin-bound MCMs and phosphorylated MCMs, as an indication for licensed and activated origins, respectively, could be further dissected. In line, analysis of checkpoint kinase 1 (CHK1) levels might shed light on the observed increased origin firing in mES cells. Potential low CHK1 levels may, via a network of several kinase-dependent processes, lead to increased ORC activation events and also influence activation timing of origins [314–316]. Similarly, increased levels of the CDK and DDK kinases or essential initiation factors such as CDC45 may result in more activated origins. Since the J1 mES cells used in this study possess 40 chromosomes, such experiments should be performed in diploid somatic cells. Furthermore, the sequences of activated origins from pluripotent cells could be compared to those from differentiated cells. This approach could reveal the potential existence of a set of ‘constitutive’ origins that are used throughout development.

We additionally found mES cell DNA replication to rely on smaller replicons resulting from shorter inter-origin distances (IOD) (Figure 6.1). To further validate these DNA fiber results, we determined IODs and, consequently, replicon sizes by genome-wide sequencing data and observed very similar tendencies. Surprisingly, we could determine a clear dependence of replicon size and inter-origin distance on the developmental stage of the cells. While pluripotent and undifferentiated cells were characterized by small replicons and short inter-origin distances, the latter increased in embryonic fibroblasts, to achieve largest values in terminally differentiated activated B-cells. Since replicons and chromatin loops are by now considered to be tightly connected DNA elements (reviewed in [227]), it would be interesting to measure mES cell chromatin loop sizes, for example, by using the DNA halo technique, and to directly correlate those to the replicon sizes determined in this study. Furthermore, by comparing origin distribution on the male X chromosome to origin organization on autosomes, we found a differential origin activation on the two parental autosomal chromosome homologues. This observation implicates shorter inter-origin distances determined by genome-wide sequencing techniques than observed on 1D stretched single DNA molecules. Lastly, we found that

around 13% of the present replication forks in the mES cell genome are unidirectional/single forks. This is in sharp contrast to amounts of single forks seen in somatic cells (5.5% and [317]). The higher amount of single forks in mES cells, may, on one side, be explained by the transcriptional hyperactivity of stem cells [289], resulting in an increased level of replication-transcription collisions and, thus, fork stalling and disassembly. Given the transcriptional hyperactivity, adapting DNA replication to unidirectional forks traveling in the same direction as RNA synthesis could reduce the risk of fork collision and stalling, thereby ensuring transcription and genome integrity and, hence, stability. Interestingly, highly transcribed ribosomal DNA (rDNA) is replicated by a substantial amount of unidirectional forks [318]. Alike, a high level of single forks was detected in *Drosophila* testes, a tissue also characterized by high transcriptional activity [319]. On the other side, however, this characteristic may represent a distinct feature of pluripotent cells underlying a yet unknown regulation mechanism. Moreover, pluripotent cells were shown to have more and, consequently, smaller replication domains. This domain organization allows a more discordant replication of contiguous chromosome segments. Neighboring domains with different replication timing are separated by timing transition regions (TTR). Interestingly, the latter were found to preferentially be replicated by unidirectional forks [184, 320]. Due to the higher number of replication domains in stem cells, the number of TTRs may increase and, subsequently, also the number of unidirectional forks to replicate these genome segments. As a result of the lack of 'backup' dormant origins, genomic regions whose replication rely on an individual fork may designate regions particularly exposed to DNA damage. Indeed, several TTRs were found to be unstable and assigned to cancer-related genes [321]. Lastly, MTBP, the metazoan Sld7 orthologue, was recently found to represent an important regulation platform for origin firing [133]. In view of a potential role of MTBP in the assembly of a symmetric replisome and/or in keeping the two parts of the replisome together until the origin is initiated [134], it may be interesting to compare MTBP levels in mES cells and differentiated cells. In line, analyzing the amount of unidirectional forks present in MTBP knock-out cells would clarify its role in replisome assembly.

### **6.2.2 Loss of DNA Methylation Enhances Replication Fork Speed**

In view of the dramatic changes of pericentromeric heterochromatin replication timing upon interfering with histone modification levels (chapter III), we also analyzed the effect of the loss of cytosine modifications in mES cells on the spatio-temporal DNA replication progression (chapter IV). Although cytosine methylation is of outstanding importance for embryonic development, regulation of gene expression and cell differentiation, it does not directly influence DNA replication timing in ES cells, since we did not observe spatio-temporal reorganizations of replication patterns during S-phase progression in DNMT triple knock-out (TKO) cells. Interestingly, we and others did not observe global changes in histone acetylation marks and previous studies also confirmed intact HP1 binding to unmethylated chromocenters in TKO cells [322]. Although we did not analyze TKO chromatin organization and structure in detail, Micrococcal nuclease (MNase) digestion of genomic DNA from mouse embryonic fibroblasts deficient for DNMT1 (MEF *p53*<sup>-/-</sup> and *dnmt1*<sup>-/-</sup>) showed similar global nucleosome distributions than the corresponding wildtype cells (MEF W8). Of note, subtle and more local changes may not be resolved by this low resolution method and can, therefore, not completely be excluded. To get deeper insights into the role of DNA methylation and chromatin organization, more detailed analyses of the distribution and amount of chromatin classes [290] and correlating histone modifications

should be performed on DNMT knock-out cells. Indeed, electron spectroscopic imaging (ESI) unveiled that, while DNA methylation is dispensable for chromatin restructuring during differentiation, it is a prerequisite for compacting heterochromatic DNA during differentiation, as J1 TKO epiblast stem cells (EpiSCs) showed even more dispersed chromocenters as wildtype stem cells [283]. While we did not observe a switch in replication patterns in TKO cells, a clear difference in the distribution of cells between the different S-phase stages became visible. Significantly fewer cells were present in S-phase stage II, suggesting that pericentromeric heterochromatin replication is faster in these cells. Faster DNA replication can be achieved in two ways. First, more licensed origins can be activated, and/or, second, replication fork speed can be increased. Hence, we measured the latter by single-molecule DNA fiber analysis. Interestingly, TKO cells have an average replication fork speed that is 10% faster than in wildtype cells. This difference is also approximately reflected in the variation of chromocenter replication time in wildtype and TKO cells. We, therefore, conclude that loss of DNA methylation enhances the replication fork speed. Nonetheless, subsequent studies should perform detailed analyses of the number of replication foci in wildtype and DNMT and TET TKO cells, to further unveil the influence of cytosine base modifications on origin activation.



---

## 6.3 Cytosine Modifications in the Mammalian Genome – DNA Duplex Stability and Metabolism

---

### 6.3.1 Cytosine Modifications Change the DNA Duplex Stability

Since we did not observe any global changes in histone modification levels and chromatin organization, we focussed on the effect of cytosine modifications on the DNA double helix *per se*. As 5hmC and 5fC can also persist as stable epigenetic marks in the genome [244, 245], it is important to know how they influence the basic structure of the DNA duplex itself. Generally, cytosine base modifications are thought not to influence the global canonical B-DNA structure, but local changes around the modification site can arise. They also do not alter the Watson-Crick base pairing in itself, but they, with their C5 modification, introduce a functional group into the double helix. These functional groups on the C5 point towards the major groove of the DNA and change the hydrophobicity, introduce steric hindrance and alter base pair stacking, structure and mechanical properties of the DNA double helix [323]. Since the existing literature is contradictory on the effect of cytosine modifications in DNA double helix stability [323–327], we analyzed the latter with our established high resolution melting temperature protocol [328]. While we observed highest melting temperatures, i.e. highest DNA stability, for methylated DNA, all oxidized variants reverted this stabilization effect and had even lower melting temperatures than unmodified cytosines. Since these were purely *in vitro* experiments, the next step was to establish *in vivo* models to study effects of cytosine modifications on DNA metabolic processes. Therefore, we established several methods based on cytosine modification ‘loss of function’ and ‘gain of function’ systems. On the one hand, by knocking-out the essential cytosine methylation writer enzymes, we generate cell models characterized by the absence of cytosine variants (Dnmt knock-out cells). On the other hand, we increase 5mC and 5hmC levels by transfecting modified deoxy-nucleoside triphosphates (dNTPs), which are phosphorylated and can directly be incorporated into genomic DNA by the cellular DNA polymerases. Lastly, we increase oxidized cytosine variations and, concomitantly decrease 5mC levels, by overexpressing the catalytic domain of Tet1 (Tet1CD). This Tet1 variant lacks the large N-terminal regulatory domain, including the CXXC zinc finger. Unlike full-length Tet1, Tet1CD overexpression can induce significant genome-wide DNA demethylation. Tet1fl is guided towards hypomethylated DNA via its N-terminal CXXC DNA binding motif, thereby mainly preventing 5mC spreading into CpG islands. The absence of this zinc finger domain in the Tet1CD variant allows the binding to unmethylated but also hypermethylated CGIs, thereby inducing significant global 5mC conversion. In line, the genome-wide analyses of Jin *et al.* [329] confirmed that Tet1CD catalytic activity is not biased to specific methylated genomic regions and can act throughout the genome.

To further study the effect of cytosine modifications on DNA helix stability in cells, we established a protocol based on heat denaturation of genomic DNA at different temperatures, similar to fluorescence *in situ* hybridization (FISH). We found similar trends to our *in vitro* observations in all our established cellular systems. All in all, *in vitro* and *in vivo* experiments showed a more stable DNA duplex in the presence of 5mC,

while the oxidized cytosine modifications 5hmC, 5fC and 5caC destabilized the double helix.

How exactly cytosine modifications change the DNA geometry and duplex stability, however, is still under debate (discussed in detail in section 3.3 *DNA Modification Readers and Writers and Their Interplay*). Briefly, cytosine methylation introduces a bulky methyl group in the DNA major groove, thereby reducing DNA bending ability, increasing DNA rigidity and resulting in a wider major and a more narrow minor groove [330, 331]. Oxidation of 5mC to cytosine hydroxymethylation introduces a more polar side group. The higher affinity for water molecules increases fluctuations in the 5hmC:G base pair [332], rendering it less stable and, in turn, less energy and/or time is needed for breaking the hydrogen bonds. Similarly, the formyl and carboxyl groups of 5fC and 5caC, respectively, were found to possess electron-withdrawing abilities [333], weakening the DNA duplex at 5fC:G and 5caC:G base pairs. These changes in DNA structure may influence the working mechanism of essential DNA metabolic enzymes.

### 6.3.2 Cytosine Modifications Influence DNA Transcription, Replication and Unwinding Rates

Since an alteration in DNA stability would be predicted to have major consequences for global genome processes, we investigated the effects of cytosine modifications on DNA replication and DNA transcription. Our *in vitro* phage transcription (IVT) and bacterial polymerization assays showed significantly reduced transcript and amplicon amounts when the template DNA (except promoter regions for IVT) contained 5mC. To rule out that the observed effects in the IVT experiments were due to differences in elongation rate, i.e. incorporation rate of nucleotides opposite of modified cytosines, we additionally compared efficiencies of T7 RNA polymerase on double-stranded (ds) and single-stranded (ss) DNA [334]. To still guarantee T7 polymerase promoter recognition, we used ssDNA fragments containing a dsDNA T7 promoter region and a ssDNA transcription sequence. We observed no significant difference in template elongation between the unmodified, methylated and hydroxymethylated substrates. This is in agreement with previous studies revealing similar rGTP incorporation rates opposite of modified and unmodified cytosines [335]. To extend our *in vitro* enzymatic assay results to a eukaryotic system, we performed RNA polymerization assays and processive soluble replication reactions with purified *Saccharomyces cerevisiae* DNA-dependent RNA polymerase I (Pol I) and replisome components that are required for replication *in vitro*, respectively. Both yeast-based *in vitro* systems are well established [160, 336] and exact template features ensuring RNA and DNA polymerization, respectively, are known. For RNA polymerization by budding yeast Pol I, we used 3' tailed templates that allow nonspecific initiation by RNA polymerases. This implies that no additional transcription factors are needed to initiate RNA polymerization. Moreover, it only recently became possible to purify active DNA-dependent RNA human polymerases [337]. Therefore, reference conditions for human Pol I transcription analysis on dsDNA templates are not yet established. Similarly, and contrary to mammalian DNA replication origins, the origin sequences of *S. cerevisiae* (autonomously replicating sequence, ARS) are known. This is in sharp contrast to mammalian origins, whose prominent characteristic is the lack of a clear sequence and signature defining the initiation sites. In addition, in budding yeast, the proteins essential for origin firing and for effective and processive DNA replication synthesis were defined [137, 160]. Moreover, although most/all homologs of yeast replication factors were identified in mammals, an *in vitro* replication assay based on individually purified components is not yet established. Nonetheless, *in vitro* replication can be performed with mammalian cell extracts using Epstein-Barr or SV40 viral-derived systems [338]. While the yeast-based assay is

no longer a single enzyme experiment, the use of purified proteins allows a very defined composition of the reaction, which is impossible when using cell/nuclear extracts. Interestingly, we observed very similar trends in the yeast RNA and DNA polymerization assays as with the phage and bacterial enzymes. These results are astonishing insofar as budding yeast enzymes are not encountering DNA modification in their natural environment [339]. Hence, this demonstrates a high degree of evolutionary conservation of the enzymes involved in genome metabolism. In line with our *in vitro* results, we see increased 5'-ethynyl uridine (EU) and 5'-ethynyl-2'-deoxyuridine (EdU) levels in d5hmCTP transfected cells. Similarly, we observed increased transcription and polymerization levels in Tet1CD overexpressing cells *in vivo*, where we achieved 5mC loss by Tet1CD mediated oxidation. We, therefore, conclude an at least partial influence of DNA helix stability on DNA dependent RNA and DNA polymerases, however, without completely excluding effects from Tet1 mediated changes in the global transcriptome.

Since both processes rely on opening the DNA double helix by breaking the hydrogen bonds between pairing DNA bases, it is likely that more stable DNA base interactions and, in turn, a more stable DNA helix, would require more energy and/or time for DNA strand separation. Our DNA melting temperature analyses and FISH-based assays described above show exactly this more stable DNA helix upon cytosine methylation (5mC). This effect is reverted by its oxidized forms (5hmC, 5fC and 5caC). To further focus on the helicase enzyme responsible for DNA strand separation during DNA replication, we uncoupled the replisome by a drug-based method [340]. Aphidicolin specifically inhibits DNA polymerases while not interfering with the helicase activity. As a helicase speed readout, we used the accumulation of fluorescently tagged single-stranded DNA binding protein (GFP-RPA34). Tet1CD overexpressing cells, as well as cells deficient for all cytosine modifications (MEF *p53*<sup>-/-</sup> and *dnmt1*<sup>-/-</sup> and J1 TKO) showed increased RPA accumulation levels in comparison to control cells. This effect was particularly strong in normally highly methylated regions, i.e. chromocenters [? ]. This result strongly suggests increased helicase unwinding speeds due to 5mC loss and a concomitant decrease in DNA helix stability, resulting in easier and faster base pair breaking and strand separation. In view of the above results in Dnmt TKO mES cells, we additionally conclude that the reduced DNA stability facilitates DNA unwinding/replication, resulting in a faster completion of S-phase stage II. *In vitro* helicase unwinding assays on modified DNA templates would further enhance our *in vivo* observations.

### 6.3.3 Cytosine Modifications Modulate DNA Metabolic Processes by Changing the DNA Stability

Our study reveals, by *in vitro* and *in vivo* assays, important effects of cytosine modifications at the very basic level of the DNA helix itself, thereby influencing numerous ubiquitous DNA related processes (Figure 6.1). Although the enzymes from different species used in this study may not be compared *per se*, the mere physical event of opening the double helix to allow DNA or RNA polymerization is an essential step in all organisms. Our observations from different experimental setups using proteins from different species are in strong favor of a dependency of the DNA opening process on the modification state of the underlying DNA. We propose that such an elementary mechanism will act in addition to the multiple other ones described before, representing a possible new contribution of modified bases to enable local fine-tuning of nuclear metabolic processes. A compelling idea arising from the changes in modified DNA helix stability is the presence of modified bases reducing duplex stability at origins of replication. In bacteria adenine methylation (m6A) at GATC sites regulates DNA replication initiation [341, 342], possibly via its DNA duplex destabilizing

effect [324]. The presence and function of m6A in mammalian genomes is, however, still debated [343–346]. In line with this idea is a recent study showing 5hmC enrichment at mammalian origins of replication and its implication in replication-licensing factor recruitment [347]. Immunofluorescent detection of base modifications on nucleotide labeled single-molecule DNA fibers may provide further insights into this assumption. Similarly, comparing the replication fork speeds in modified versus unmodified DNA regions would be an additional pivot of our model of base modifications as local regulators of nuclear mechanisms. In this respect, the identification and sequencing of the first mammalian origins (very early origins) to fire by nascent DNA sequencing in single-cells may allow further investigations on the chromatin landscape these sequences are embedded in.

#### **6.3.4 Cytosine Modifications and Their Readers and Writers Regulate Genome Stability**

**B**esides the here-above discussed influences on DNA double helix stability and the ensuing effects on ubiquitous and essential nuclear processes, cytosine modifications are known to greatly influence the genomic distribution of their readers, and, accordingly, gene expression patterns. Recent findings, however, highlight that the opposite point of view is also possible; reader proteins define the DNA modification placement by regulating the accessibility of the modified DNA bases to their respective modifier enzymes. Indeed the MBD protein family members MeCP2, MBD2 and MBD1b were shown to hinder the Tet1 mediated 5mC oxidation to 5hmC by blocking the access to the DNA and the methylated base (guardians of the epigenome, [35, 348]). In line, unrestricted access of TET1 to pericentromeric heterochromatin (MeCP2 knock-out cells) and, in turn, elevated 5hmC levels, increased transcriptional noise from major satellite repeats [35]. Similar observations were made for LINE-1 retrotransposition activation [349]. However, the opposite case was observed for MBD1a, which, via direct interactions with TET1, recruits the dioxygenase to highly methylated pericentromeric heterochromatin [348]. Taken together, these studies demonstrate that complex time and dose-dependent cross-regulations of cytosine modifications and their readers and modifiers exist. Cytosine modifications, therefore, represent an essential epigenetic tool for the cell to translate the modifications into distinct chromatin states, thereby fine-tuning essential nuclear processes. Additionally, this tool can act at distinct levels of the chromatin, ranging from the DNA double helix stability to higher-order chromatin states and is inherently regulated by the tight interplay of DNA modifications and modification readers and writers.

In summary, several levels of epigenetic regulation, ranging from the DNA double helix properties, over the distinct interplay of DNA modification readers and writers, to higher-order chromatin and loop structures, are tightly linked and interconnected to coordinate complex nuclear processes, to maintain genome integrity and stability and, eventually, to ensure organism development and persistence.

## 7 Bibliography

- [1] J. D. Watson and F. H. Crick, "Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid," *Nature*, vol. 171, no. 4356, pp. 737–738, 1953. [Online]. Available: <https://www.nature.com/articles/171737a0>
- [2] K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, and T. J. Richmond, "Crystal structure of the nucleosome core particle at 2.8 Å resolution," *Nature*, vol. 389, no. 6648, pp. 251–260, 1997. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/9305837>
- [3] H. D. Ou, S. Phan, T. J. Deerinck, A. Thor, M. H. Ellisman, and C. C. O'Shea, "ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells," *Science (80-. )*, vol. 357, no. 6349, p. eaag0025, 2017. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/28751582>
- [4] J. Dekker, A. S. Belmont, M. Guttman, V. O. Leshyk, J. T. Lis, S. Lomvardas, L. A. Mirny, C. C. O'Shea, P. J. Park, B. Ren, J. C. Ritland Politz, J. Shendure, and S. Zhong, "The 4D nucleome project," *Nature*, vol. 549, no. 7671, pp. 219–226, 2017. [Online]. Available: [www.4dnucleome.org/](http://www.4dnucleome.org/)
- [5] E. Chargaff, R. Lipshitz, C. Green, and M. E. Hodes, "The composition of the desoxyribonucleic acid of salmon sperm," *J Biol Chem*, vol. 192, pp. 223–230, 1951. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/14917668/>
- [6] J. H. Gommers-Ampt and P. Borst, "Hypermethylated bases in DNA," *FASEB J*, vol. 9, no. 11, pp. 1034–1042, 1995. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/7649402>
- [7] S. S. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, and E. L. Aiden, "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping," *Cell*, vol. 159, no. 7, pp. 1665–1680, 2014. [Online]. Available: <http://dx.doi.org/10.1016/j.cell.2014.11.021>
- [8] M. Merkenschlager and E. P. Nora, "CTCF and cohesin in genome folding and transcriptional gene regulation," *Annu. Rev. Genomics Hum. Genet.*, vol. 17, pp. 17–43, 2016. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/27089971/>
- [9] A. S. Hansen, "CTCF as a boundary factor for cohesin-mediated loop extrusion: evidence for a multi-step mechanism," *Nucleus*, vol. 11, no. 1, pp. 132–148, 2020. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/32631111/>
- [10] L. J. Kaaij, F. Mohn, R. H. van der Weide, E. de Wit, and M. Bühler, "The ChAHP Complex Counteracts Chromatin Looping at CTCF Sites that Emerged from SINE Expansions in Mouse," *Cell*, vol. 178, no. 6, pp. 1437–1451.e14, 2019. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/31491387/>
- [11] E. P. Nora, B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. Van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Blüthgen, J. Dekker, and E. Heard, "Spatial partitioning of the regulatory landscape of the X-inactivation centre," *Nature*, vol. 485, no. 7398, pp. 381–385, 2012. [Online]. Available: <https://www.nature.com/articles/nature11049>
- [12] B. D. Pope, T. Ryba, V. Dileep, F. Yue, W. Wu, O. Denas, D. L. Vera, Y. Wang, R. S. Hansen, T. K. Canfield, R. E. Thurman, Y. Cheng, G. Gülsoy, J. H. Dennis, M. P. Snyder, J. A. Stamatoyannopoulos, J. Taylor, R. C. Hardison, T. Kahveci, B. Ren, and D. M. Gilbert, "Topologically associating domains are stable units of replication-timing regulation," *Nature*, vol. 515, no. 7527, pp. 402–405, 2014. [Online]. Available: <https://www.nature.com/articles/nature13986>
- [13] E. Lieberman-Aiden, N. L. Van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker, "Comprehensive mapping of long-range interactions reveals folding principles of the human genome," *Science (80-. )*, vol. 326, no. 5950, pp. 289–293, 2009. [Online]. Available: [/pmc/articles/PMC2858594/](http://pmc/articles/PMC2858594/)
- [14] S. I. Wright, "Evolution of genome size," *eLS*, pp. 1–6, 2017. [Online]. Available: <https://doi.org/10.1002/9780470015902.a0023983>

- 
- [15] T. A. Elliott and T. R. Gregory, "What's in a genome? The C-value enigma and the evolution of eukaryotic genome content," *Philos. Trans. R. Soc. B*, vol. 370, no. 1678, p. 20140331, 2015. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/26323762/>
- [16] M. B. Clark, P. P. Amaral, F. J. Schlesinger, M. E. Dinger, R. J. Taft, J. L. Rinn, C. P. Ponting, P. F. Stadler, K. V. Morris, A. Morillon, J. S. Rozowsky, M. B. Gerstein, C. Wahlestedt, Y. Hayashizaki, P. Carninci, T. R. Gingeras, and J. S. Mattick, "The reality of pervasive transcription," *PLoS Biol.*, vol. 9, no. 7, p. e1000625, 2011. [Online]. Available: <https://dx.plos.org/10.1371/journal.pbio.1000625>
- [17] A. Woolfe, M. Goodson, D. K. Goode, P. Snell, G. K. McEwen, T. Vavouri, S. F. Smith, P. North, H. Callaway, K. Kelly, K. Walter, I. Abnizova, W. Gilks, Y. J. K. Edwards, J. E. Cooke, and G. Elgar, "Highly conserved non-coding sequences are associated with vertebrate development," *PLoS Biol.*, vol. 3, no. 1, p. e7, 2004. [Online]. Available: <https://dx.plos.org/10.1371/journal.pbio.0030007>
- [18] S. Ohno, "So Much 'Junk DNA' in our Genome," *Brookhaven Symp Biol*, vol. 23, pp. 366–370., 1972. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/5065367/>
- [19] L. E. Orgel and F. H. Crick, "Selfish DNA: The ultimate parasite," *Nature*, vol. 284, no. 5757, pp. 604–607, 1980. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/7366731/>
- [20] F. Jacob and J. Monod, "Genetic regulatory mechanisms in the synthesis of proteins," *J. Mol. Biol.*, vol. 3, no. 3, pp. 318–356, 1961. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/13718526/>
- [21] F. Jacob and S. Brenner, "On the regulation of DNA synthesis in bacteria: the hypothesis of the replicon." *C R Hebd Seances Acad Sci.*, vol. 256, pp. 298–300, 1963. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/13964280/>
- [22] J. A. Shapiro and R. Von Sternberg, "Why repetitive DNA is essential to genome function," *Biol Rev Camb Philos Soc*, vol. 80, no. 2, pp. 227–250, 2005. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/15921050/>
- [23] B. A. Bouwman and W. de Laat, "Getting the genome in shape: the formation of loops, domains and compartments," *Genome Biol*, vol. 16, p. 154, 2015. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/26257189>
- [24] P. H. Krijger and W. de Laat, "Identical cells with different 3D genomes; cause and consequences?" *Curr Opin Genet Dev*, vol. 23, no. 2, pp. 191–196, 2013. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23415810>
- [25] J. C. Politz, D. Scalzo, and M. Groudine, "Something silent this way forms: the functional organization of the repressive nuclear compartment," *Annu Rev Cell Dev Biol*, vol. 29, pp. 241–270, 2013. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23834025>
- [26] I. Solovei, K. Thanisch, and Y. Feodorova, "How to rule the nucleus: divide et impera," *Curr Opin Cell Biol*, vol. 40, pp. 47–59, 2016. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/26938331>
- [27] H. R. Widlund, P. N. Kuduvalli, M. Bengtsson, H. Cao, T. D. Tullius, and M. Kubista, "Nucleosome structural features and intrinsic properties of the TATAACGCC repeat sequence," *J. Biol. Chem.*, vol. 274, no. 45, pp. 31847–31852, 1999. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/10542209/>
- [28] D. Schmidt, P. C. Schwalie, M. D. Wilson, B. Ballester, Â. Goncalves, C. Kutter, G. D. Brown, A. Marshall, P. Flicek, and D. T. Odom, "Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages," *Cell*, vol. 148, no. 1-2, pp. 335–348, 2012. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/22244452/>
- [29] P. A. Jones, "Functions of DNA methylation: islands, start sites, gene bodies and beyond," *Nat Rev Genet*, vol. 13, no. 7, pp. 484–492, 2012. [Online]. Available: <https://www.nature.com/articles/nrg3230>
- [30] Z. D. Smith and A. Meissner, "DNA methylation: Roles in mammalian development," *Nat Rev Genet*, vol. 14, no. 3, pp. 204–220, 2013. [Online]. Available: <https://www.nature.com/articles/nrg3354>
- [31] P. Gerdes, S. R. Richardson, D. L. Mager, and G. J. Faulkner, "Transposable elements in the mammalian embryo: Pioneers surviving through stealth and service," *Genome Biol.*, vol. 17, no. 1, 2016. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/27161170/>

- [32] S. Tabata, A. Oka, K. Sugimoto, M. Takanami, S. Yasuda, and Y. Hirota, "The 245 base-pair oriC sequence of the *E. coli* chromosome directs bidirectional replication at an adjacent region." *Nucleic Acids Res.*, vol. 11, no. 9, pp. 2617–2626, 1983. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/6344019/>
- [33] S. Kearsley, "Structural requirements for the function of a yeast chromosomal replicator," *Cell*, vol. 37, no. 1, pp. 299–307, 1984. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/6327056/>
- [34] G. Schotta, A. Ebert, R. Dorn, and G. Reuter, "Position-effect variegation and the genetic dissection of chromatin regulation in *Drosophila*," *Semin. Cell Dev. Biol.*, vol. 14, no. 1, pp. 67–75, 2003. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/12524009/>
- [35] A. K. Ludwig, P. Zhang, F. D. Hastert, S. Meyer, C. Rausch, H. D. Herce, A. Lehmkuhl, I. Hellmann, C. Trummer, C. Storm, H. Leonhardt, M. Cristina Cardoso, U. Müller, A. Lehmkuhl, I. Hellmann, C. Trummer, C. Storm, H. Leonhardt, and M. C. Cardoso, "Binding of MBD proteins to DNA blocks Tet1 function thereby modulating transcriptional noise," *Nucleic Acids Res.*, vol. 45, no. 5, pp. 2438–2457, 2017. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/27923996/>
- [36] J. R. Korenberg and M. C. Rykowski, "Human genome organization: Alu, LINES, and the molecular structure of metaphase chromosome bands," *Cell*, vol. 53, no. 3, pp. 391–400, 1988. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/3365767/>
- [37] J. A. Croft, J. M. Bridger, S. Boyle, P. Perry, P. Teague, and W. A. Bickmore, "Differences in the localization and morphology of chromosomes in the human nucleus," *J. Cell Biol.*, vol. 145, no. 6, pp. 1119–1131, 1999. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/10366586/>
- [38] G. I. Goldberg, I. Collier, and A. Cassel, "Specific DNA sequences associated with the nuclear matrix in synchronized mouse 3T3 cells," *PNAS*, vol. 80, no. 22 I, pp. 6887–6891, 1983. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/6580619/>
- [39] P. Rollini, S. J. Namciu, M. D. Marsden, and R. E. Fournier, "Identification and characterization of nuclear matrix-attachment regions in the human serpin gene cluster at 14q32.1," *Nucleic Acids Res.*, vol. 27, no. 19, pp. 3779–3791, 1999. [Online]. Available: <http://ftp.genome.gov>
- [40] J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren, "Topological domains in mammalian genomes identified by analysis of chromatin interactions," *Nature*, vol. 485, no. 7398, pp. 376–380, 2012. [Online]. Available: <https://www.nature.com/articles/nature11082>
- [41] K. Kruse, N. Díaz, R. Enriquez-Gasca, X. Gaume, M. E. Torres-Padilla, and J. M. Vaquerizas, "Transposable elements drive reorganisation of 3D chromatin during early embryogenesis," *bioRxiv*, p. 523712, 2019. [Online]. Available: <https://doi.org/10.1101/523712>
- [42] E. B. Chuong, N. C. Elde, and C. Feschotte, "Regulatory activities of transposable elements: From conflicts to benefits," *Nat. Rev. Genet.*, vol. 18, no. 2, pp. 71–86, 2017. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5498291/>
- [43] G. Kunarso, N. Y. Chia, J. Jeyakani, C. Hwang, X. Lu, Y. S. Chan, H. H. Ng, and G. Bourque, "Transposable elements have rewired the core regulatory network of human embryonic stem cells," *Nat. Genet.*, vol. 42, no. 7, pp. 631–634, 2010. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/20526341/>
- [44] J. T. Roberts, S. E. Cardin, and G. M. Borchert, "Burgeoning evidence indicates that microRNAs were initially formed from transposable element sequences," *Mob. Genet. Elements*, vol. 4, no. 3, p. e29255, 2014. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/25054081/>
- [45] A. Kapusta, Z. Kronenberg, V. J. Lynch, X. Zhuo, L. A. Ramsay, G. Bourque, M. Yandell, and C. Feschotte, "Transposable Elements Are Major Contributors to the Origin, Diversification, and Regulation of Vertebrate Long Noncoding RNAs," *PLoS Genet.*, vol. 9, no. 4, 2013. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/23637635/>
- [46] M. Cowley and R. J. Oakey, "Transposable Elements Re-Wire and Fine-Tune the Transcriptome," *PLoS Genet.*, vol. 9,

- 
- no. 1, p. e1003234, 2013. [Online]. Available: [www.plosgenetics.org](http://www.plosgenetics.org)
- [47] C. M. Rubin, R. H. Kimura, and C. W. Schmid, "Selective stimulation of translational expression by Alu RNA," *Nucleic Acids Res.*, vol. 30, no. 14, pp. 3253–3261, 2002. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/12136107/>
- [48] International Human Genome Sequencing Consortium, "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, no. 6822, pp. 860–921, 2001. [Online]. Available: <https://www.nature.com/articles/35057062>
- [49] S. L. Martin and F. D. Bushman, "Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon," *Mol. Cell. Biol.*, vol. 21, no. 2, pp. 467–475, 2001. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/11134335/>
- [50] H. H. Kazazian and J. V. Moran, "The impact of L1 retrotransposons on the human genome," *Nat. Genet.*, vol. 19, no. 1, pp. 19–24, 1998. [Online]. Available: <https://www.nature.com/articles/ng0598-19>
- [51] G. J. Cost, Q. Feng, A. Jacquier, and J. D. Boeke, "Human L1 element target-primed reverse transcription in vitro," *EMBO J.*, vol. 21, no. 21, pp. 5899–5910, 2002. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/12411507/>
- [52] J. O. Kriegs, G. Churakov, J. Jurka, J. Brosius, and J. Schmitz, "Evolutionary history of 7SL RNA-derived SINEs in Suprprimates," *Trends Genet.*, vol. 23, no. 4, pp. 158–161, 2007. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/17307271/>
- [53] K. K. Kojima, "Human transposable elements in Repbase: Genomic footprints from fish to humans," *Mob. DNA*, vol. 9, no. 1, p. 2, 2018. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/29308093/>
- [54] M. Dewannieux, C. Esnault, and T. Heidmann, "LINE-mediated retrotransposition of marked Alu sequences," *Nat. Genet.*, vol. 35, no. 1, pp. 41–48, 2003. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/12897783/>
- [55] M. Dewannieux, T. Heidmann, and M. Yaniv, "L1-mediated retrotransposition of murine B1 and B2 SINEs recapitulated in cultured cells," *J. Mol. Biol.*, vol. 349, no. 2, pp. 241–247, 2005. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/15890192/>
- [56] D. Mager and J. Stoye, "Mammalian endogenous retroviruses," *Microbiol. Spectr.*, vol. 3, no. 1, pp. 1079–1100, 2015. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/26104559/>
- [57] P. O. Brown, B. Bowerman, H. E. Varmus, and J. M. Bishop, "Correct integration of retroviral DNA in vitro," *Cell*, vol. 49, no. 3, pp. 347–356, 1987. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/3032450/>
- [58] T. Sultana, A. Zamborlini, G. Cristofari, and P. Lesage, "Integration site selection by retroviruses and transposable elements in eukaryotes," *Nat. Rev. Genet.*, vol. 18, no. 5, pp. 292–308, 2017. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/28286338/>
- [59] H. M. Rowe, J. Jakobsson, D. Mesnard, J. Rougemont, S. Reynard, T. Aktas, P. V. Maillard, H. Layard-Liesching, S. Verp, J. Marquis, F. Spitz, D. B. Constam, and D. Trono, "KAP1 controls endogenous retroviruses in embryonic stem cells," *Nature*, vol. 463, no. 7278, pp. 237–240, 2010. [Online]. Available: <https://www.nature.com/articles/nature08674>
- [60] H. P. Bogerd, H. L. Wiegand, B. P. Doehle, K. K. Lueders, and B. R. Cullen, "APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells," *Nucleic Acids Res.*, vol. 34, no. 1, pp. 89–95, 2006. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/16407327/>
- [61] M. W. Vandewege, R. N. Platt, D. A. Ray, and F. G. Hoffmann, "Transposable element targeting by piRNAs in laurasiatherians with distinct transposable element histories," *Genome Biol. Evol.*, vol. 8, no. 5, pp. 1327–1337, 2016. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/27060702/>
- [62] R. Rebollo, S. Farivar, and D. L. Mager, "C-GATE - Catalogue of genes affected by transposable elements," *Mob. DNA*, vol. 3, no. 1, p. 9, 2012. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/22621612/>
- [63] J. S. Wayne and H. F. Willard, "Structure, organization, and sequence of alpha satellite DNA from human chromosome 17: evidence for evolution by unequal crossing-over and an ancestral pentamer repeat shared with the human X chromosome." *Mol. Cell. Biol.*, vol. 6, no. 9, pp. 3156–3165, 1986. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/3785225/>

- [64] A. K. Wong and J. B. Rattner, "Sequence organization and cytological localization of the minor satellite of mouse," *Nucleic Acids Res.*, vol. 16, no. 24, pp. 11645–11661, 1988. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC339101/>
- [65] A. Joseph, A. R. Mitchell, and O. J. Miller, "The organization of the mouse satellite DNA at centromeres," *Exp. Cell Res.*, vol. 183, no. 2, pp. 494–500, 1989. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/2767161/>
- [66] Consortium Mouse Genome Sequencing, "Initial sequencing and comparative analysis of the mouse genome," *Nature*, vol. 420, no. 6915, pp. 520–562, 2002. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/12466850>
- [67] M. Guenatri, D. Bailly, C. Maison, and G. Almouzni, "Mouse centric and pericentric satellite repeats form distinct functional heterochromatin," *J Cell Biol*, vol. 166, no. 4, pp. 493–505, 2004. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/15302854>
- [68] M. Jarmuz, C. D. Glotzbach, K. A. Bailey, R. Bandyopadhyay, and L. G. Shaffer, "The evolution of satellite III DNA subfamilies among primates," *Am. J. Hum. Genet.*, vol. 80, no. 3, pp. 495–501, 2007. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1821104/>
- [69] K. F. Sullivan, M. Hechenberger, and K. Masri, "Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere," *J. Cell Biol.*, vol. 127, no. 3, pp. 581–592, 1994. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/7962047/>
- [70] L. Bosch-Presegué, H. Raurell-Vila, J. K. Thackray, J. González, C. Casal, N. Kane-Goldsmith, M. Vizoso, J. P. Brown, A. Gómez, J. Ausió, T. Zimmermann, M. Esteller, G. Schotta, P. B. Singh, L. Serrano, and A. Vaquero, "Mammalian HP1 isoforms have specific roles in heterochromatin structure and organization," *Cell Rep.*, vol. 21, no. 8, pp. 2048–2057, 2017. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/2767161/>
- [71] R. K. Moyzis, J. M. Buckingham, L. S. Cram, M. Dani, L. L. Deaven, M. D. Jones, J. Meyne, R. L. Ratliff, and J. R. Wu, "A highly conserved repetitive DNA sequence, (TTAGGG)(n), present at the telomeres of human chromosomes," *PNAS*, vol. 85, no. 18, pp. 6622–6626, 1988. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/3413114/>
- [72] T. De Lange, "Shelterin: The protein complex that shapes and safeguards human telomeres," *Genes Dev.*, vol. 19, no. 18, pp. 2100–2110, 2005. [Online]. Available: <http://www.genesdev.org/cgi/doi/10.1101/>
- [73] J. R. Williamson, M. K. Raghuraman, and T. R. Cech, "Monovalent cation-induced structure of telomeric DNA: The G-quartet model," *Cell*, vol. 59, no. 5, pp. 871–880, 1989. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/2590943/>
- [74] J. D. Griffith, L. Comeau, S. Rosenfield, R. M. Stansel, A. Bianchi, H. Moss, and T. De Lange, "Mammalian telomeres end in a large duplex loop," *Cell*, vol. 97, no. 4, pp. 503–514, 1999. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/10338214/>
- [75] S. Schatzlein, A. Lucas-Hahn, E. Lemme, W. A. Kues, M. Dorsch, M. P. Manns, H. Niemann, and K. L. Rudolph, "Telomere length is reset during early mammalian embryogenesis," *PNAS*, vol. 101, no. 21, pp. 8034–8038, 2004. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/15148368/>
- [76] N. W. Kim, M. A. Piatyszek, K. R. Prowse, C. B. Harley, M. D. West, P. L. Ho, G. M. Coviello, W. E. Wright, S. L. Weinrich, and J. W. Shay, "Specific association of human telomerase activity with immortal cells and cancer," *Science (80-. )*, vol. 266, no. 5193, pp. 2011–2015, 1994. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/7605428/>
- [77] J. D. Watson and F. H. Crick, "Genetical implications of the structure of deoxyribonucleic acid," *Nature*, vol. 171, no. 4361, pp. 964–967, 1953. [Online]. Available: <https://www.nature.com/articles/171964b0>
- [78] V. O. Chagin, J. H. Stear, and M. C. Cardoso, "Organization of DNA replication," *CSH Persp B*, vol. 2, no. 4, p. a000737, 2010. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/20452942>
- [79] M. W. Parker, M. R. Botchan, and J. M. Berger, "Mechanisms and regulation of DNA replication initiation in eukaryotes," *Crit Rev Biochem Mol Biol*, vol. 52, no. 2, pp. 107–144, 2017. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/28094588>

- 
- [80] M. Fragkos, O. Ganier, P. Coulombe, and M. Méchali, "DNA replication origin activation in space and time," *Nat Rev Mol Cell Biol*, vol. 16, no. 6, pp. 360–374, 2015. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/25999062>
- [81] M. Shimizu, Y. Noguchi, Y. Sakiyama, H. Kawakami, T. Katayama, and S. Takada, "Near-atomic structural model for bacterial DNA replication initiation complex and its functional insights," *PNAS*, vol. 113, no. 50, pp. E8021–E8030, 2016. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5167161/>
- [82] J. M. Kaguni, "Replication initiation at the Escherichia coli chromosomal origin," *Curr. Opin. Chem. Biol.*, vol. 15, no. 5, pp. 606–613, 2011. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3189269/>
- [83] K. K. Swinger and P. A. Rice, "IHF and HU: Flexible architects of bent DNA," *Curr. Opin. Struct. Biol.*, vol. 14, no. 1, pp. 28–35, 2004. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/15102446/>
- [84] K. Keyamura, N. Fujikawa, T. Ishida, S. Ozaki, M. Su'etsugu, K. Fujimitsu, W. Kagawa, S. Yokoyama, H. Kurumizaka, and T. Katayama, "The interaction of DiaA and DnaA regulates the replication cycle in E. coli by directly promoting ATP-DnaA-specific initiation complexes," *Genes Dev.*, vol. 21, no. 16, pp. 2083–2099, 2007. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1948862/>
- [85] K. Keyamura and T. Katayama, "DnaA protein DNA-binding domain binds to Hda protein to promote inter-AAA+ domain interaction involved in regulatory inactivation of DnaA," *J. Biol. Chem.*, vol. 286, no. 33, pp. 29 336–29 346, 2011. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3190739/>
- [86] T. G. Palzkill and C. S. Newlon, "A yeast replication origin consists of multiple copies of a small conserved sequence," *Cell*, vol. 53, no. 3, pp. 441–450, 1988. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/3284655/>
- [87] S. Devbhandari, J. Jiang, C. Kumar, I. Whitehouse, and D. Remus, "Chromatin constrains the initiation and elongation of DNA replication," *Mol. Cell*, vol. 65, pp. 131–141, 2017. [Online]. Available: <http://dx.doi.org/10.1016/j.molcel.2016.10.035>
- [88] C. F. Kurat, J. T. Yeeles, H. Patel, A. Early, and J. F. Diffley, "Chromatin controls DNA replication origin selection, lagging-strand synthesis, and replication fork rates," *Mol. Cell*, vol. 65, no. 1, pp. 117–130, 2017. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/27989438/>
- [89] D. Schaarschmidt, J. Baltin, I. M. Stehle, H. J. Lipps, and R. Knippers, "An episomal mammalian replicon: Sequence-independent binding of the origin recognition complex," *EMBO J.*, vol. 23, no. 1, pp. 191–201, 2004. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/14685267/>
- [90] S. Paixão, I. N. Colaluca, M. Cubells, F. A. Peverali, A. Destro, S. Giadrossi, M. Giacca, A. Falaschi, S. Riva, and G. Biamonti, "Modular structure of the human Lamin B2 replicator," *Mol. Cell. Biol.*, vol. 24, no. 7, pp. 2958–2967, 2004. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/15024083/>
- [91] A. L. Altman and E. Fanning, "Defined sequence modules and an architectural element cooperate to promote initiation at an ectopic mammalian chromosomal replication origin," *Mol. Cell. Biol.*, vol. 24, no. 10, pp. 4138–4150, 2004. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC400449/>
- [92] L. Wang, C.-M. Lin, S. Brooks, D. Cimbora, M. Groudine, and M. I. Aladjem, "The human  $\beta$ -Globin replication initiation region consists of two modular independent replicators," *Mol. Cell. Biol.*, vol. 24, no. 8, pp. 3373–3386, 2004. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/15060158/>
- [93] E. Besnard, A. Babled, L. Lapasset, O. Milhavet, H. Parrinello, C. Dantec, J. M. Marin, and J. M. Lemaitre, "Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs," *Nat. Struct. Mol. Biol.*, vol. 19, no. 8, pp. 837–844, 2012. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/22751019/>
- [94] A. R. Langley, S. Gräf, J. C. Smith, and T. Krude, "Genome-wide identification and characterisation of human DNA replication origins by initiation site sequencing (ini-seq)," *Nucleic Acids Res.*, vol. 44, no. 21, pp. 10 230–10 247, 2016. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/27587586/>
- [95] B. Bartholdy, R. Mukhopadhyay, J. Lajugie, M. I. Aladjem, and E. E. Bouhassira, "Allele-specific analysis of

- DNA replication origins in mammalian cells," *Nat. Commun.*, vol. 6, no. 1, pp. 1–12, 2015. [Online]. Available: <https://www.nature.com/articles/ncomms8051>
- [96] P. Prorok, M. Artufel, A. Aze, P. Coulombe, I. Peiffer, L. Lacroix, A. Guédin, J. L. Mergny, J. Damaschke, A. Schepers, C. Cayrou, M.-P. Teulade-Fichou, B. Ballester, and M. Méchali, "Involvement of G-quadruplex regions in mammalian replication origin activity," *Nat Commun*, vol. 10, no. 1, p. 3274, 2019. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/31332171/>
- [97] M. Giacca, L. Zentilin, P. Norio, S. Diviacco, D. Dimitrova, G. Contreas, G. Biamonti, G. Perini, F. Weighardt, S. Riva, and A. Falaschi, "Fine mapping of a replication origin of human DNA," *PNAS*, vol. 91, no. 15, pp. 7119–7123, 1994. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC44350/>
- [98] G. Abdurashidova, M. Deganuto, R. Klima, S. Riva, G. Biamonti, M. Giacca, and A. Falaschi, "Start sites of bidirectional DNA synthesis at the human lamin B2 origin," *Science (80-. )*, vol. 287, no. 5460, pp. 2023–2026, 2000. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/10720330/>
- [99] C. Mcwhinney and M. Leffak, "Autonomous replication of a DNA fragment containing the chromosomal replication origin of the human c-myc gene," *Nucleic Acids Res.*, vol. 18, no. 5, pp. 1233–1242, 1990. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC330439/>
- [100] M. Ghosh, M. Kemp, G. Liu, M. Ritzi, A. Schepers, and M. Leffak, "Differential binding of replication proteins across the human c-myc replicator," *Mol. Cell. Biol.*, vol. 26, no. 14, pp. 5270–5283, 2006. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1592723/>
- [101] M. Malott and M. Leffak, "Activity of the c-myc replicator at an ectopic chromosomal location," *Mol. Cell. Biol.*, vol. 19, no. 8, pp. 5685–5695, 1999. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC84420/>
- [102] P. A. Dijkwel and J. L. Hamlin, "The Chinese hamster dihydrofolate reductase origin consists of multiple potential nascent-strand start sites." *Mol. Cell. Biol.*, vol. 15, no. 6, pp. 3023–3031, 1995. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC230533/>
- [103] P. A. Dijkwel, S. Wang, and J. L. Hamlin, "Initiation sites are distributed at frequent intervals in the chinese hamster dihydrofolate reductase origin of replication but are used with very different efficiencies," *Mol. Cell. Biol.*, vol. 22, no. 9, pp. 3053–3065, 2002. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/11940663/>
- [104] L. D. Mesner, X. Li, P. A. Dijkwel, and J. L. Hamlin, "The dihydrofolate reductase origin of replication does not contain any nonredundant genetic elements required for origin activity," *Mol. Cell. Biol.*, vol. 23, no. 3, pp. 804–814, 2003. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC140713/>
- [105] M. Gossen, D. T. Pak, S. K. Hansen, J. K. Acharya, and M. R. Botchan, "A Drosophila homolog of the yeast origin recognition complex," *Science (80-. )*, vol. 270, no. 5242, pp. 1674–1677, 1995. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/7502079/>
- [106] A. Rowles, J. P. Chong, L. Brown, M. Howell, G. I. Evan, and J. J. Blow, "Interaction between the origin recognition complex and the replication licensing system in *Xenopus*," *Cell*, vol. 87, no. 2, pp. 287–296, 1996. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/8861912/>
- [107] S. K. Dhar and A. Dutta, "Identification and characterization of the human ORC6 homolog," *J. Biol. Chem.*, vol. 275, no. 45, pp. 34 983–34 988, 2000. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/10945994/>
- [108] S. Donovan, J. Harwood, L. S. Drury, and J. F. Diffley, "Cdc6p-dependent loading of Mcm proteins onto pre-replicative chromatin in budding yeast," *PNAS*, vol. 94, no. 11, pp. 5611–5616, 1997. [Online]. Available: <https://www.pnas.org/content/94/11/5611>
- [109] M. Weinreich, C. Liang, and B. Stillman, "The Cdc6p nucleotide-binding motif is required for loading Mcm proteins onto chromatin," *PNAS*, vol. 96, no. 2, pp. 441–446, 1999. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC15155/>
- [110] A. Rowles and J. J. Blow, "Chromatin proteins involved in the initiation of DNA replication," *Curr. Opin. Genet. Dev.*,

- 
- vol. 7, no. 2, pp. 152–157, 1997. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/9115430/>
- [111] C. Speck and B. Stillman, “Cdc6 ATPase activity regulates ORC-Cdc6 stability and the selection of specific DNA sequences as origins of DNA replication,” *J. Biol. Chem.*, vol. 282, no. 16, pp. 11705–11714, 2007. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/17314092/>
- [112] T. Mizushima, N. Takahashi, and B. Stillman, “Cdc6p modulates the structure and DNA binding activity of the origin recognition complex in vitro,” *Genes Dev.*, vol. 14, no. 13, pp. 1631–1641, 2000. [Online]. Available: <http://genesdev.cshlp.org/content/14/13/1631.full.html>
- [113] D. Y. Takeda, Y. Shibata, J. D. Parvin, and A. Dutta, “Recruitment of ORC or CDC6 to DNA is sufficient to create an artificial origin of replication in mammalian cells,” *Genes Dev.*, vol. 19, no. 23, pp. 2827–2836, 2005. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1315390/>
- [114] J. H. Shin, B. Grabowski, R. Kasiviswanathan, S. D. Bell, and Z. Kelman, “Regulation of minichromosome maintenance helicase activity by Cdc6,” *J. Biol. Chem.*, vol. 278, no. 39, pp. 38059–38067, 2003. [Online]. Available: <http://www.jbc.org/>
- [115] S. A. Samel, A. Fernández-Cid, J. Sun, A. Riera, S. Tognetti, M. Carmen Herrera, H. Li, and C. Speck, “A unique DNA entry gate serves for regulated loading of the eukaryotic replicative helicase MCM2-7 onto DNA,” *Genes Dev.*, vol. 28, no. 15, pp. 1653–1666, 2014. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4117941/>
- [116] J. Sun, C. Evrin, S. A. Samel, A. Fernández-Cid, A. Riera, H. Kawakami, B. Stillman, C. Speck, and H. Li, “Cryo-EM structure of a helicase loading intermediate containing ORC-Cdc6-Cdt1-MCM2-7 bound to DNA,” *Nat. Struct. Mol. Biol.*, vol. 20, no. 8, pp. 944–951, 2013. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/23851460/>
- [117] Y. Zhai, E. Cheng, H. Wu, N. Li, P. Y. K. Yung, N. Gao, and B. K. Tye, “Open-ringed structure of the Cdt1-Mcm2-7 complex as a precursor of the MCM double hexamer,” *Nat. Struct. Mol. Biol.*, vol. 24, no. 3, pp. 300–308, 2017. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/28191894/>
- [118] S. Ticau, L. J. Friedman, K. Champasa, I. R. Corrêa, J. Gelles, and S. P. Bell, “Mechanism and timing of Mcm2-7 ring closure during DNA replication origin licensing,” *Nat. Struct. Mol. Biol.*, vol. 24, no. 3, pp. 309–315, 2017. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/28191892/>
- [119] S. Ticau, L. J. Friedman, N. A. Ivica, J. Gelles, and S. P. Bell, “Single-molecule studies of origin licensing reveal mechanisms ensuring bidirectional helicase loading,” *Cell*, vol. 161, no. 3, pp. 513–525, 2015. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4445235/>
- [120] G. Coster, J. Frigola, F. Beuron, E. P. Morris, and J. F. Diffley, “Origin licensing requires ATP binding and hydrolysis by the MCM replicative helicase,” *Mol. Cell*, vol. 55, no. 5, pp. 666–677, 2014. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/25087873/>
- [121] A. Fernández-Cid, A. Riera, S. Tognetti, M. C. Herrera, S. Samel, C. Evrin, C. Winkler, E. Gardenal, S. Uhle, and C. Speck, “An ORC/Cdc6/MCM2-7 complex is formed in a multistep reaction to serve as a platform for MCM double-hexamer assembly,” *Mol. Cell*, vol. 50, no. 4, pp. 577–588, 2013. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/23603117/>
- [122] S. Kang, M. D. Warner, and S. P. Bell, “Multiple functions for Mcm2-7 ATPase motifs during replication initiation,” *Mol. Cell*, vol. 55, no. 5, pp. 655–665, 2014. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/25087876/>
- [123] J. Sun, A. Fernández-Cid, A. Riera, S. Tognetti, Z. Yuan, B. Stillman, C. Speck, and H. Li, “Structural and mechanistic insights into Mcm2-7 double-hexamer assembly and function,” *Genes Dev.*, vol. 28, no. 20, pp. 2291–2303, 2014. [Online]. Available: <http://www.genesdev.org/cgi/doi/10.1101/gad.242313.114>
- [124] C. Evrin, A. Fernández-Cid, J. Zech, M. C. Herrera, A. Riera, P. Clarke, S. Brill, R. Lurz, and C. Speck, “In the absence of ATPase activity, pre-RC formation is blocked prior to MCM2-7 hexamer dimerization,” *Nucleic Acids Res.*, vol. 41, no. 5, pp. 3162–3172, 2013. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/23376927/>
- [125] C. Evrin, A. Fernández-Cid, A. Riera, J. Zech, P. Clarke, M. C. Herrera, S. Tognetti, R. Lurz, and

- C. Speck, "The ORC/Cdc6/MCM2-7 complex facilitates MCM2-7 dimerization during prereplicative complex formation," *Nucleic Acids Res.*, vol. 42, no. 4, pp. 2257–2269, 2014. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3936773/>
- [126] G. Coster and J. F. Diffley, "Bidirectional eukaryotic DNA replication is established by quasi-symmetrical helicase loading," *Science (80-. )*, vol. 357, no. 6348, pp. 314–318, 2017. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5608077/>
- [127] T. C. Miller, J. Locke, J. F. Greiwe, J. F. Diffley, and A. Costa, "Mechanism of head-to-head MCM double-hexamers formation revealed by cryo-EM," *Nature*, vol. 575, no. 7784, pp. 704–710, 2019. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6887548/>
- [128] O. Hyrien, K. Marheineke, and A. Goldar, "Paradoxes of eukaryotic DNA replication: MCM proteins and the random completion problem," *Bioessays*, vol. 25, no. 2, pp. 116–125, 2003. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/12539237/>
- [129] A. M. Woodward, T. Göhler, M. G. Luciani, M. Oehlmann, X. Ge, A. Gartner, D. A. Jackson, and J. J. Blow, "Excess MCM2-7 license dormant origins of replication that can be used under conditions of replicative stress," *J. Cell Biol.*, vol. 173, no. 5, pp. 673–683, 2006. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2063885/>
- [130] X. Q. Ge, D. A. Jackson, and J. J. Blow, "Dormant origins licensed by excess MCM2-7 are required for human cells to survive replicative stress," *Genes Dev.*, vol. 21, no. 24, pp. 3331–3341, 2007. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2113033/>
- [131] A. Ibarra, E. Schwob, and J. Méndez, "Excess MCM proteins protect human cells from replicative stress by licensing backup origins of replication," *PNAS*, vol. 105, no. 26, pp. 8956–8961, 2008. [Online]. Available: <https://www.pnas.org/content/105/26/8956.abstract>
- [132] Y. J. Sheu and B. Stillman, "Cdc7-Dbf4 Phosphorylates MCM Proteins via a Docking Site-Mediated Mechanism to Promote S-Phase Progression," *Mol. Cell*, vol. 24, no. 1, pp. 101–113, 2006. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/17018296/>
- [133] K. Köhler, L. Sanchez-Pulido, V. Höfer, A. Marko, C. P. Ponting, A. P. Snijders, R. Feederle, A. Schepers, and D. Boos, "The Cdk8/19-cyclin C transcription regulator functions in genome replication through metazoan Sld7," *PLoS Biol.*, vol. 17, no. 1, p. e2006767, 2019. [Online]. Available: <https://doi.org/10.1371/journal.pbio.2006767>
- [134] P. Ferreira, V. Höfer, N. Kronshage, A. Marko, K. U. Reusswig, B. Tetik, C. Dießel, K. Köhler, N. Tschernoster, J. Altmüller, N. Schulze, B. Pfander, and D. Boos, "MTBP phosphorylation controls DNA replication origin firing," *Sci. Rep.*, vol. 11, no. 1, 2021. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7895959/>
- [135] J. Walter and J. Newport, "Initiation of eukaryotic DNA replication: Origin unwinding and sequential chromatin association of Cdc45, RPA, and DNA polymerase  $\alpha$ ," *Mol. Cell*, vol. 5, no. 4, pp. 617–627, 2000. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/10882098/>
- [136] R. C. Heller, S. Kang, W. M. Lam, S. Chen, C. S. Chan, and S. P. Bell, "Eukaryotic origin-dependent DNA replication in vitro reveals sequential action of DDK and S-CDK kinases," *Cell*, vol. 146, no. 1, pp. 80–91, 2011. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/21729781/>
- [137] J. T. Yeeles, T. D. Deegan, A. Janska, A. Early, and J. F. Diffley, "Regulated eukaryotic DNA replication origin firing with purified proteins," *Nature*, vol. 519, no. 7544, pp. 431–435, 2015. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/25739503/>
- [138] S. E. Moyer, P. W. Lewis, and M. R. Botchan, "Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase," *PNAS*, vol. 103, no. 27, pp. 10236–10241, 2006. [Online]. Available: [www.pnas.org/cgi/doi/10.1073/pnas.0602400103](https://www.pnas.org/cgi/doi/10.1073/pnas.0602400103)
- [139] M. Pacek, A. V. Tutter, Y. Kubota, H. Takisawa, and J. C. Walter, "Localization of MCM2-7, Cdc45, and GINS to the site of DNA unwinding during eukaryotic DNA replication," *Mol. Cell*, vol. 21, no. 4, pp. 581–587, 2006. [Online].

---

Available: <https://pubmed.ncbi.nlm.nih.gov/16483939/>

- [140] Z. Yuan, R. Georgescu, L. Bai, D. Zhang, H. Li, and M. E. O'Donnell, "DNA unwinding mechanism of a eukaryotic replicative CMG helicase," *Nat. Commun.*, vol. 11, no. 1, pp. 1–10, 2020. [Online]. Available: <https://doi.org/10.1038/s41467-020-14577-6>
- [141] P. Perez-Arnaiz, I. Bruck, and D. L. Kaplan, "Mcm10 coordinates the timely assembly and activation of the replication fork helicase," *Nucleic Acids Res.*, vol. 44, no. 1, pp. 315–329, 2016. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/26582917/>
- [142] F. Van Deursen, S. Sengupta, G. De Piccoli, A. Sanchez-Diaz, and K. Labib, "Mcm 10 associates with the loaded DNA helicase at replication origins and defines a novel step in its activation," *EMBO J.*, vol. 31, no. 9, pp. 2195–2206, 2012. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/22433841/>
- [143] M. Lööke, M. F. Maloney, and S. P. Bell, "Mcm10 regulates DNA replication elongation by stimulating the CMG replicative helicase," *Genes Dev.*, vol. 31, no. 3, pp. 291–305, 2017. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5358725/>
- [144] Y. Quan, Y. Xia, L. Liu, J. Cui, Z. Li, Q. Cao, X. S. Chen, J. L. Campbell, and H. Lou, "Cell-cycle-regulated interaction between Mcm10 and double hexameric Mcm2-7 is required for helicase splitting and activation during S-phase," *Cell Rep.*, vol. 13, no. 11, pp. 2576–2586, 2015. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5536962/>
- [145] A. Costa, L. Renault, P. Swuec, T. Petojevic, J. J. Pesavento, I. Ilves, K. MacLellan-Gibson, R. A. Fleck, M. R. Botchan, and J. M. Berger, "DNA binding polarity, dimerization, and ATPase ring remodeling in the CMG helicase of the eukaryotic replisome," *Elife*, vol. 3, pp. 1–17, 2014. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/25117490/>
- [146] R. Georgescu, Z. Yuan, L. Bai, R. De Luna Almeida Santos, J. Sun, D. Zhang, O. Yurieva, H. Li, and M. E. O'Donnell, "Structure of eukaryotic CMG helicase at a replication fork and implications to replisome architecture and origin initiation," *PNAS*, vol. 114, no. 5, pp. E697–E706, 2017. [Online]. Available: [www.pnas.org/cgi/doi/10.1073/pnas.1620500114](http://www.pnas.org/cgi/doi/10.1073/pnas.1620500114)
- [147] M. E. Douglas, F. A. Ali, A. Costa, and J. F. Diffley, "The mechanism of eukaryotic CMG helicase activation," *Nature*, vol. 555, no. 7695, pp. 265–268, 2018. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/29489749/>
- [148] I. Ilves, T. Petojevic, J. J. Pesavento, and M. R. Botchan, "Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins," *Mol. Cell*, vol. 37, no. 2, pp. 247–258, 2010. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6396293/>
- [149] M. L. DePamphilis, "Eukaryotic DNA replication: Anatomy of an origin," *Annu. Rev. Biochem.*, vol. 62, pp. 29–63, 1993. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/8352592/>
- [150] I. Akerman, B. Kasaai, A. Bazarova, P. B. Sang, I. Peiffer, M. Artufel, R. Derelle, G. Smith, M. Rodriguez-Martinez, M. Romano, S. Kinet, P. Tino, C. Theillet, N. Taylor, B. Ballester, and M. Méchali, "A predictable conserved DNA base composition signature defines human core DNA replication origins," *Nat. Commun.*, vol. 11, no. 1, pp. 1–15, 2020. [Online]. Available: <https://doi.org/10.1038/s41467-020-18527-0>
- [151] R. C. Conaway and I. R. Lehman, "A DNA primase activity associated with DNA polymerase alpha from *Drosophila melanogaster* embryos." *PNAS*, vol. 79, no. 8, pp. 2523–2527, 1982. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC346231/>
- [152] T. Kozu, T. Yagura, and T. Seno, "De novo DNA synthesis by a novel mouse DNA polymerase associated with primase activity," *Nature*, vol. 298, no. 5870, pp. 180–182, 1982. [Online]. Available: <https://www.nature.com/articles/298180a0>
- [153] M. L. Kilkenny, M. A. Longo, R. L. Perera, and L. Pellegrini, "Structures of human primase reveal design of nucleotide elongation site and mode of Pol  $\alpha$  tethering," *PNAS*, vol. 110, no. 40, pp. 15 961–15 966, 2013. [Online].

Available: [www.pnas.org/cgi/doi/10.1073/pnas.1311185110](http://www.pnas.org/cgi/doi/10.1073/pnas.1311185110)

- [154] R. L. Perera, R. Torella, S. Klinge, M. L. Kilkenny, J. D. Maman, and L. Pellegrini, "Mechanism for priming DNA synthesis by yeast DNA Polymerase  $\alpha$ ," *Elife*, vol. 2013, no. 2, p. e00482, 2013. [Online]. Available: <https://elifesciences.org/articles/00482>
- [155] Z. F. Pursell, I. Isoz, E. B. Lundström, E. Johansson, and T. A. Kunkel, "Yeast DNA polymerase  $\epsilon$  participates in leading-strand DNA replication," *Science (80-. )*, vol. 317, no. 5834, pp. 127–130, 2007. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2233713/>
- [156] S. A. Nick McElhinny, D. A. Gordenin, C. M. Stith, P. M. Burgers, and T. A. Kunkel, "Division of labor at the eukaryotic replication fork," *Mol. Cell*, vol. 30, no. 2, pp. 137–144, 2008. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2654179/>
- [157] I. Miyabe, T. A. Kunkel, and A. M. Carr, "The major roles of DNA polymerases epsilon and delta at the eukaryotic replication fork are evolutionarily conserved," *PLoS Genet.*, vol. 7, no. 12, p. 1002407, 2011. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3228825/>
- [158] A. R. Clausen, S. A. Lujan, A. B. Burkholder, C. D. Orebaugh, J. S. Williams, M. F. Clausen, E. P. Malc, P. A. Mieczkowski, D. C. Fargo, D. J. Smith, and T. A. Kunkel, "Tracking replication enzymology in vivo by genome-wide mapping of ribonucleotide incorporation," *Nat. Struct. Mol. Biol.*, vol. 22, no. 3, pp. 185–191, 2015. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/25622295/>
- [159] R. E. Johnson, R. Klassen, L. Prakash, and S. Prakash, "A major role of DNA polymerase  $\delta$  in replication of both the leading and lagging DNA strands," *Mol. Cell*, vol. 59, no. 2, pp. 163–175, 2015. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/26145172/>
- [160] J. T. Yeeles, A. Janska, A. Early, and J. F. Diffley, "How the Eukaryotic Replisome Achieves Rapid and Efficient DNA Replication," *Mol. Cell*, vol. 65, no. 1, pp. 105–116, 2017. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/27989442/>
- [161] G. L. Moldovan, B. Pfander, and S. Jentsch, "PCNA, the maestro of the replication fork," *Cell*, vol. 129, no. 4, pp. 665–679, 2007. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/17512402/>
- [162] O. Chilkova, P. Stenlund, I. Isoz, C. M. Stith, P. Grabowski, E. B. Lundström, P. M. Burgers, and E. Johansson, "The eukaryotic leading and lagging strand DNA polymerases are loaded onto primer-ends via separate mechanisms but have comparable processivity in the presence of PCNA," *Nucleic Acids Res.*, vol. 35, no. 19, pp. 6588–6597, 2007. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2095795/>
- [163] A. Sporbert, A. Gahl, R. Ankerhold, H. Leonhardt, and M. C. Cardoso, "DNA polymerase clamp shows little turnover at established replication sites but sequential de novo assembly at adjacent origin clusters," *Mol. Cell*, vol. 10, no. 6, pp. 1355–1365, 2002. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/12504011/>
- [164] A. Sporbert, P. Domaing, H. Leonhardt, and M. C. Cardoso, "PCNA acts as a stationary loading platform for transiently interacting Okazaki fragment maturation proteins." *Nucleic Acids Res.*, vol. 33, no. 11, pp. 3521–3528, 2005. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/15972794/>
- [165] D. Dovrat, J. L. Stodola, P. M. J. Burgers, and A. Aharoni, "Sequential switching of binding partners on PCNA during in vitro Okazaki fragment maturation," *PNAS*, vol. 111, no. 39, pp. 14 118–14 123, 2014. [Online]. Available: [www.pnas.org/cgi/doi/10.1073/pnas.1321349111](http://www.pnas.org/cgi/doi/10.1073/pnas.1321349111)
- [166] M. S. Kang, E. Ryu, S. W. Lee, J. Park, N. Y. Ha, J. S. Ra, Y. J. Kim, J. Kim, M. Abdel-Rahman, S. H. Park, K.-y. Lee, H. Kim, S. Kang, and K. Myung, "Regulation of PCNA cycling on replicating DNA by RFC and RFC-like complexes," *Nat. Commun.*, vol. 10, no. 1, pp. 1–16, 2019. [Online]. Available: <https://doi.org/10.1038/s41467-019-10376-w>
- [167] O. Mortusewicz and H. Leonhardt, "XRCC1 and PCNA are loading platforms with distinct kinetic properties and different capacities to respond to multiple DNA lesions," *BMC Mol. Biol.*, vol. 8, no. 1, p. 81, 2007. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/17880707/>

- 
- [168] R. Okazaki, T. Okazaki, K. Sakabe, K. Sugimoto, R. Kainuma, A. Sugino, and N. Iwatsuki, "In vivo mechanism of DNA chain growth," *CSH Symp. Quant.*, vol. 33, pp. 129–143, 1968. [Online]. Available: <http://symposium.cshlp.org/content/33/129>
- [169] J. L. Stodola and P. M. Burgers, "Resolving individual steps of Okazaki-fragment maturation at a millisecond timescale," *Nat. Struct. Mol. Biol.*, vol. 23, no. 5, pp. 402–408, 2016. [Online]. Available: <https://www.nature.com/articles/nsmb.3207>
- [170] N. K. Sinha, C. F. Morris, and B. M. Alberts, "Efficient in vitro replication of double-stranded DNA templates by a purified T4 bacteriophage replication system." *J. Biol. Chem.*, vol. 255, pp. 4290–93, 1980. [Online]. Available: <https://www.jbc.org/content/255/9/4290.abstract>
- [171] A. C. Simon, J. C. Zhou, R. L. Perera, F. Van Deursen, C. Evrin, M. E. Ivanova, M. L. Kilkenny, L. Renault, S. Kjaer, D. Matak-Vinkovi, K. Labib, A. Costa, and L. Pellegrini, "A Ctf4 trimer couples the CMG helicase to DNA polymerase  $\alpha$  in the eukaryotic replisome," *Nature*, vol. 510, no. 7504, pp. 293–297, 2014. [Online]. Available: <https://www.nature.com/articles/nature13234>
- [172] F. Villa, A. C. Simon, M. A. Ortiz Bazan, M. L. Kilkenny, D. Wirthensohn, M. Wightman, D. Matak-Vinković, L. Pellegrini, and K. Labib, "Ctf4 Is a hub in the eukaryotic replisome that links multiple CIP-box proteins to the CMG helicase," *Mol. Cell*, vol. 63, no. 3, pp. 385–396, 2016. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/27397685/>
- [173] J. S. Hanna, E. S. Kroll, V. Lundblad, and F. A. Spencer, "Saccharomyces cerevisiae CTF18 and CTF4 are required for sister chromatid cohesion," *Mol. Cell. Biol.*, vol. 21, no. 9, pp. 3144–3158, 2001. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC86942/>
- [174] M. Fumasoni, K. Zwicky, F. Vanoli, M. Lopes, and D. Branzei, "Error-free DNA damage tolerance and sister chromatid proximity during DNA replication rely on the Pol $\alpha$ /Primase/Ctf4 complex," *Mol. Cell*, vol. 57, no. 5, pp. 812–823, 2015. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/25661486/>
- [175] J. H. Taylor, "Asynchronous duplication of chromosomes in cultured cells of Chinese hamster." *J. Biophys. Biochem. Cytol.*, vol. 7, no. 3, pp. 455–464, 1960. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/13837165/>
- [176] K. Woodfine, H. Fiegler, D. M. Beare, J. E. Collins, O. T. McCann, B. D. Young, S. Debernardi, R. Mott, I. Dunham, and N. P. Carter, "Replication timing of the human genome," *Hum Mol Genet*, vol. 13, no. 2, pp. 191–202, 2004. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/14645202>
- [177] D. Mantiero, A. MacKenzie, A. Donaldson, and P. Zegerman, "Limiting replication initiation factors execute the temporal programme of origin firing in budding yeast," *EMBO J.*, vol. 30, no. 23, pp. 4805–4814, 2011. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3243606/>
- [178] C. A. Müller and C. A. Nieduszynski, "DNA replication timing influences gene expression level," *J. Cell Biol.*, vol. 216, no. 7, pp. 1907–1914, 2017. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5496624/>
- [179] O. Padovan-Merhar, G. P. Nair, A. G. Biaesch, A. Mayer, S. Scarfone, S. W. Foley, A. R. Wu, L. S. Churchman, A. Singh, and A. Raj, "Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms," *Mol. Cell*, vol. 58, no. 2, pp. 339–352, 2015. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/25866248/>
- [180] Y. Voichek, R. Bar-Ziv, and N. Barkai, "Expression homeostasis during DNA replication," *Science (80-. )*, vol. 351, no. 6277, pp. 1087–1090, 2016. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/26941319/>
- [181] A. Koren, P. Polak, J. Nemesh, J. J. Michaelson, J. Sebat, S. R. Sunyaev, and S. A. McCarroll, "Differential relationship of DNA replication timing to different forms of human mutation and variation," *Am. J. Hum. Genet.*, vol. 91, no. 6, pp. 1033–1040, 2012. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3516607/>
- [182] F. Supek and B. Lehner, "Differential DNA mismatch repair underlies mutation rate variation across the human genome," *Nature*, vol. 521, no. 7550, pp. 81–84, 2015. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/25707793/>

- [183] R. S. Hansen, S. Thomas, R. Sandstrom, T. K. Canfield, R. E. Thurman, M. Weaver, M. O. Dorschner, S. M. Gartler, and J. A. Stamatoyannopoulos, "Sequencing newly replicated DNA reveals widespread plasticity in human replication timing," *PNAS*, vol. 107, no. 1, pp. 139–144, 2010. [Online]. Available: [www.pnas.org/cgi/doi/10.1073/pnas.0912402107](http://www.pnas.org/cgi/doi/10.1073/pnas.0912402107)
- [184] I. Hiratani, T. Ryba, M. Itoh, T. Yokochi, M. Schwaiger, C.-W. W. Chang, Y. Lyou, T. M. Townes, D. Schübeler, and D. M. Gilbert, "Global reorganization of replication domains during embryonic stem cell differentiation," *PLoS Biol*, vol. 6, no. 10, p. e245, 2008. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/18842067>
- [185] D. A. Jackson and A. Pombo, "Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S-phase in human cells," *J Cell Biol*, vol. 140, no. 6, pp. 1285–1295, 1998. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2132671/>
- [186] H. Leonhardt, H. P. Rahn, P. Weinzierl, A. Sporberr, T. Cremer, D. Zink, and M. C. Cardoso, "Dynamics of DNA replication factories in living cells." *J. Cell Biol.*, vol. 149, no. 2, pp. 271–280, 2000. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/10769021/>
- [187] E. M. Manders, J. Stap, J. Strackee, R. Van Driel, and J. A. Aten, "Dynamic behavior of DNA replication domains," *Exp Cell Res*, vol. 226, no. 2, pp. 328–335, 1996. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/8806436>
- [188] J. H. van Dierendonck, R. Keyzer, C. J. van de Velde, and C. J. Cornelisse, "Subdivision of S-phase by analysis of nuclear 5-bromodeoxyuridine staining patterns," *Cytometry*, vol. 10, no. 2, pp. 143–150, 1989. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/2469556>
- [189] T. Ryba, I. Hiratani, J. Lu, M. Itoh, M. Kulik, J. Zhang, T. C. Schulz, A. J. Robins, S. Dalton, and D. M. Gilbert, "Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types," *Genome Res.*, vol. 20, no. 6, pp. 761–770, 2010. [Online]. Available: <http://www.genome.org/cgi/doi/10.1101/gr.099655.109>.
- [190] P. Hozak, D. A. Jackson, and P. R. Cook, "Replication factories and nuclear bodies: the ultrastructural characterization of replication sites during the cell cycle," *J Cell Sci*, vol. 107, no. 8, pp. 2191–2202, 1994. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/7983177>
- [191] D. Löb, N. Lengert, V. O. Chagin, M. Reinhart, C. S. Casas-Delucchi, M. C. Cardoso, and B. Drossel, "3D replicon distributions arise from stochastic initiation and domino-like DNA replication progression," *Nat Commun*, vol. 7, no. 1, p. 11207, 2016. [Online]. Available: [www.nature.com/naturecommunications](http://www.nature.com/naturecommunications)
- [192] K. S. Heinz, C. S. Casas-Delucchi, T. Török, D. Cmarko, A. Rapp, I. Raska, and M. C. Cardoso, "Peripheral re-localization of constitutive heterochromatin advances its replication timing and impairs maintenance of silencing marks," *Nucleic Acids Res*, vol. 46, no. 12, pp. 6112–6128, 2018. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/29750270>
- [193] V. O. Chagin, B. Reinhart, A. Becker, O. Mortusewicz, K. L. Jost, A. Rapp, H. Leonhardt, and M. C. Cardoso, "Processive DNA synthesis is associated with localized decompaction of constitutive heterochromatin at the sites of DNA replication and repair," *Nucleus*, vol. 10, no. 1, pp. 231–253, 2019. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/31744372>
- [194] P. K. Patel, B. Arcangioli, S. P. Baker, A. Bensimon, and N. Rhind, "DNA replication origins fire stochastically in fission yeast," *Mol. Biol. Cell*, vol. 17, no. 1, pp. 308–316, 2006. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/16251353/>
- [195] M. Eshaghi, R. K. M. Karuturi, J. Li, Z. Chu, E. T. Liu, and J. Liu, "Global profiling of DNA replication timing and efficiency reveals that efficient replication/firing occurs late during S-Phase in *S. pombe*," *PLoS One*, vol. 2, no. 8, p. e722, 2007. [Online]. Available: <https://dx.plos.org/10.1371/journal.pone.0000722>
- [196] O. Hyrien and A. Goldar, "Mathematical modelling of eukaryotic DNA replication," *Chromosom. Res.*, vol. 18, no. 1, pp. 147–161, 2010. [Online]. Available: <https://link.springer.com/article/10.1007/s10577-009-9092-4>

- 
- [197] J. D. Braunstein, D. Schulze, T. Delgiudice, A. Furst, and C. L. Schildkraut, "The temporal order of replication of murine immunoglobulin heavy chain constant region sequences corresponds to their linear order in the genome," *Nucleic Acids Res.*, vol. 10, no. 21, pp. 6887–6902, 1982. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC326972/>
- [198] A. Maya-Mendoza, P. Olivares-Chauvet, A. Shaw, and D. A. Jackson, "S-phase progression in human cells is dictated by the genetic continuity of DNA foci," *PLoS Genet.*, vol. 6, no. 4, p. 1000900, 2010. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2851568/>
- [199] D. Schübeler, D. Scalzo, C. Kooperberg, B. van Steensel, J. Delrow, and M. Groudine, "Genome-wide DNA replication profile for *Drosophila melanogaster*: A link between transcription and replication timing," *Nat. Genet.*, vol. 32, no. 3, pp. 438–442, 2002. [Online]. Available: <http://www.nature.com/naturegenetics>
- [200] S. Gnan, I. M. Flyamer, K. N. Klein, E. Castelli, A. Rapp, A. Maiser, N. Chen, P. Weber, E. Enervald, M. C. Cardoso, W. A. Bickmore, D. M. Gilbert, and S. C. B. Buonomo, "Nuclear organisation and replication timing are coupled through RIF1-PP1 interaction," *Nat. Commun.*, vol. 12, no. 1, p. 2910, 2021. [Online]. Available: <http://www.nature.com/articles/s41467-021-22899-2>
- [201] T. Yokochi, K. Poduch, T. Ryba, J. Lu, I. Hiratani, M. Tachibana, Y. Shinkai, and D. M. Gilbert, "G9a selectively represses a class of late-replicating genes at the nuclear periphery," *PNAS*, vol. 106, no. 46, pp. 19363–19368, 2009. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/19889976/>
- [202] C. G. Sansam, K. Pietrzak, B. Majchrzycka, M. A. Kerlin, J. Chen, S. Rankin, and C. L. Sansam, "A mechanism for epigenetic control of DNA replication," *Genes Dev.*, vol. 32, no. 3-4, pp. 224–229, 2018. [Online]. Available: </pmc/articles/PMC5859964/>
- [203] B. Miotto, Z. Ji, and K. Struhl, "Selectivity of ORC binding sites and the relation to replication timing, fragile sites, and deletions in cancers," *PNAS*, vol. 113, no. 33, pp. E4810–E4819, 2016. [Online]. Available: <https://www.pnas.org/content/113/33/E4810.abstract>
- [204] S. Jørgensen, G. Schotta, and C. S. Sørensen, "Histone H4 Lysine 20 methylation: Key player in epigenetic regulation of genomic integrity," *Nucleic Acids Res.*, vol. 41, no. 5, pp. 2797–2806, 2013. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/23345616/>
- [205] M. Tardat, R. Murr, Z. Herceg, C. Sardet, and E. Julien, "PR-Set7-dependent lysine methylation ensures genome replication and stability through S-phase," *J. Cell Biol.*, vol. 179, no. 7, pp. 1413–1426, 2007. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/18158331/>
- [206] M. Tardat, J. Brustel, O. Kirsh, C. Lefevbre, M. Callanan, C. Sardet, and E. Julien, "The histone H4 Lys 20 methyltransferase PR-Set7 regulates replication origins in mammalian cells," *Nat. Cell Biol.*, vol. 12, no. 11, pp. 1086–1093, 2010. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/20953199/>
- [207] D. B. Beck, A. Burton, H. Oda, C. Ziegler-Birling, M. E. Torres-Padilla, and D. Reinberg, "The role of PR-Set7 in replication licensing depends on Suv4-20h," *Genes Dev.*, vol. 26, no. 23, pp. 2580–2589, 2012. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3521623/>
- [208] J. G. Aparicio, C. J. Viggiani, D. G. Gibson, and O. M. Aparicio, "The Rpd3-Sin3 histone deacetylase regulates replication timing and enables intra-S origin control in *Saccharomyces cerevisiae*," *Mol. Cell. Biol.*, vol. 24, no. 11, pp. 4769–4780, 2004. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC416400/>
- [209] W. A. Bickmore and A. D. Carothers, "Factors affecting the timing and imprinting of replication on a mammalian chromosome," *J. Cell Sci.*, vol. 108, no. 8, pp. 2801–2809, 1995. [Online]. Available: <https://europepmc.org/article/med/7593321>
- [210] H. F. Jørgensen, V. Azuara, S. Amoils, M. Spivakov, A. Terry, T. Nesterova, B. S. Cobb, B. Ramsahoye, M. Merckenschlager, and A. G. Fisher, "The impact of chromatin modifiers on the timing of locus replication in mouse embryonic stem cells," *Genome Biol.*, vol. 8, no. 8, p. R169, 2007. [Online]. Available:

<http://www.ncbi.nlm.nih.gov/pubmed/17705870>

- [211] A. M. Keohane, L. P. O'Neill, N. D. Belyaev, J. S. Lavender, and B. M. Turner, "X-inactivation and histone H4 acetylation in embryonic stem cells," *Dev. Biol.*, vol. 180, no. 2, pp. 618–630, 1996. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/8954732/>
- [212] C. M. Lin, H. Fu, M. Martinovsky, E. Bouhassira, and M. I. Aladjem, "Dynamic Alterations of Replication Timing in Mammalian Cells," *Curr. Biol.*, vol. 13, no. 12, pp. 1019–1028, 2003. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0960982203003828>
- [213] P. Perry, S. Sauer, N. Billon, W. D. Richardson, M. Spivakov, G. Warnes, F. J. Livesey, M. Merckenschlager, A. G. Fisher, and V. Azuara, "A dynamic switch in the replication timing of key regulator genes in embryonic stem cells upon neural induction," *Cell Cycle*, vol. 3, no. 12, pp. 1619–1624, 2004. [Online]. Available: <https://europepmc.org/article/med/15611653>
- [214] M. Vogelauer, L. Rubbi, I. Lucas, B. J. Brewer, and M. Grunstein, "Histone acetylation regulates the time of replication origin firing," *Mol Cell*, vol. 10, no. 5, pp. 1223–1233, 2002. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/12453428>
- [215] D. Schübeler, C. Francastel, D. M. Cimbara, A. Reik, D. I. Martin, and M. Groudine, "Nuclear localization and histone acetylation: A pathway for chromatin opening and transcriptional activation of the human  $\beta$ -globin locus," *Genes Dev.*, vol. 14, no. 8, pp. 940–950, 2000. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/10783166/>
- [216] M. G. Kemp, M. Ghosh, G. Liu, and M. Leffak, "The histone deacetylase inhibitor trichostatin A alters the pattern of DNA replication origin activity in human cells," *Nucleic Acids Res.*, vol. 33, no. 1, pp. 325–336, 2005. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/15653633/>
- [217] A. Goren, A. Tabib, M. Hecht, and H. Cedar, "DNA replication timing of the human  $\beta$ -globin domain is controlled by histone modification at the origin," *Genes Dev.*, vol. 22, no. 10, pp. 1319–1324, 2008. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2377185/>
- [218] S. Leidescher, J. Nübler, Y. Feodorova, E. Hildebrand, S. Ullrich, S. Bultmann, S. Link, K. Thanisch, J. Dekker, H. Leonhardt, L. Mirny, and I. Solovei, "Spatial organization of transcribed eukaryotic genes," *bioRxiv*, 2020. [Online]. Available: <https://doi.org/10.1101/2020.05.20.106591>
- [219] L. A. Mirny and I. Solovei, "Keeping chromatin in the loop(s)," *Nat. Rev. Mol. Cell Biol.*, vol. 22, pp. 439–440, 2021. [Online]. Available: [www.nature.com/nrm](http://www.nature.com/nrm)
- [220] R. Berezney, D. D. Dubey, and J. A. Huberman, "Heterogeneity of eukaryotic replicons, replicon clusters, and replication foci," *Chromosoma*, vol. 108, no. 8, pp. 471–484, 2000. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/10794569>
- [221] M. Buongiorno-Nardelli, G. Micheli, M. T. Carri, and M. Marilley, "A relationship between replicon size and supercoiled loop domains in the eukaryotic genome," *Nature*, vol. 298, no. 5869, pp. 100–102, 1982. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/7088157>
- [222] M. Marilley and M. Buongiorno-Nardelli, "Relationship between the organization of DNA loop domains and of replicons in the eukaryotic genome," *Adv Exp Med Biol*, vol. 179, pp. 163–168, 1984. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/6524497>
- [223] H. Nakamura, T. Morita, and C. Sato, "Structural organizations of replicon domains during DNA synthetic phase in the mammalian nucleus," *Exp Cell Res*, vol. 165, no. 2, pp. 291–297, 1986. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/3720850>
- [224] M. Reinhart and M. C. Cardoso, "A journey through the microscopic ages of DNA replication," *Protoplasma*, vol. 254, no. 3, pp. 1151–1162, 2017. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/27943022>
- [225] V. O. Chagin, C. S. Casas-Delucchi, M. Reinhart, L. Schermelleh, Y. Markaki, A. Maiser, J. J. Bolius, A. Bensimon, M. Fillies, P. Domaing, Y. M. Rozanov, H. Leonhardt, and M. C. Cardoso, "4D Visualization of replication foci in

- 
- mammalian cells corresponding to individual replicons," *Nat Commun*, vol. 7, p. 11231, 2016. [Online]. Available: <https://www.nature.com/articles/ncomms11231>
- [226] H. Ma, J. Samarabandu, R. S. Devdhar, R. Acharya, P. C. Cheng, C. Meng, and R. Berezney, "Spatial and temporal dynamics of DNA replication sites in mammalian cells," *J Cell Biol*, vol. 143, no. 6, pp. 1415–1425, 1998. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/9852140>
- [227] S. Mamberti and M. C. Cardoso, "Are the processes of DNA replication and DNA repair reading a common structural chromatin unit?" *Nucleus*, vol. 11, no. 1, pp. 66–82, 2020. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/32275847>
- [228] M. Wijgerde, F. Grosveld, and P. Fraser, "Transcription complex stability and chromatin dynamics in vivo," *Nature*, vol. 377, no. 6546, pp. 209–213, 1995. [Online]. Available: <https://www.nature.com/articles/377209a0>
- [229] E. Guillou, A. Ibarra, V. Coulon, J. Casado-Vela, D. Rico, I. Casal, E. Schwob, A. Losada, and J. Mendez, "Cohesin organizes chromatin loops at DNA replication factories," *Genes Dev*, vol. 24, no. 24, pp. 2812–2822, 2010. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21159821>
- [230] M. Cremer, D. M. Gilbert, K. Brandstetter, A. Maiser, S. S. P. Rao, M. Guirao-ortiz, N. Mitra, S. Mamberti, K. N. Klein, H. Leonhardt, M. C. Cardoso, E. L. Aiden, and T. Cremer, "Cohesin depleted cells rebuild functional nuclear compartments after endomitosis," *Nat. Commun.*, vol. 11, p. 6146, 2020. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/33262376/>
- [231] J. Song, O. Rechkoblit, T. H. Bestor, and D. J. Patel, "Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation," *Science (80-. )*, vol. 331, no. 6020, pp. 1036–1040, 2011. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21163962>
- [232] J. Song, M. Teplova, S. Ishibe-Murakami, and D. J. Patel, "Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation," *Science (80-. )*, vol. 335, no. 6069, pp. 709–712, 2012. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/22323818/>
- [233] X. Liu, Q. Gao, P. Li, Q. Zhao, J. Zhang, J. Li, H. Koseki, and J. Wong, "UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9," *Nat. Commun.*, vol. 4, p. 1563, 2013. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/23463006/>
- [234] J. Fang, J. Cheng, J. Wang, Q. Zhang, M. Liu, R. Gong, P. Wang, X. Zhang, Y. Feng, W. Lan, Z. Gong, C. Tang, J. Wong, H. Yang, C. Cao, and Y. Xu, "Hemi-methylated DNA opens a closed conformation of UHRF1 to facilitate its histone recognition," *Nat. Commun.*, vol. 7, p. 11197, 2016. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/27045799/>
- [235] W. Qin, P. Wolf, N. Liu, S. Link, M. Smets, F. L. Mastra, I. Forné, G. Pichler, D. Hörl, K. Fellingner, F. Spada, I. M. Bonapace, A. Imhof, H. Harz, and H. Leonhardt, "DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination," *Cell Res.*, vol. 25, pp. 911–929, 2015. [Online]. Available: <https://www.nature.com/articles/cr201572>
- [236] J. S. Harrison, E. M. Cornett, D. Goldfarb, P. A. Darosa, Z. M. Li, F. Yan, B. M. Dickson, A. H. Guo, D. V. Cantu, L. Kaustov, P. J. Brown, C. H. Arrowsmith, D. A. Erie, M. B. Major, R. E. Klevit, K. Krajewski, B. Kuhlman, B. D. Strahl, and S. B. Rothbart, "Hemi-methylated DNA regulates DNA methylation inheritance through allosteric activation of H3 ubiquitylation by UHRF1," *Elife*, vol. 5, 2016. [Online]. Available: <https://elifesciences.org/articles/17101>
- [237] B. M. Foster, P. Stolz, C. B. Mulholland, A. Montoya, H. Kramer, S. Bultmann, and T. Bartke, "Critical role of the UBL domain in stimulating the E3 ubiquitin ligase activity of UHRF1 toward chromatin," *Mol. Cell*, vol. 72, no. 4, pp. 739–752.e9, 2018. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/30392929/>
- [238] H. Leonhardt, A. W. Page, H.-U. Weier, and T. H. Bestor, "A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei," *Cell*, vol. 71, pp. 865–873, 1992. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/1423634/>
-

- [239] L. Schermelleh, A. Haemmer, F. Spada, N. Rösing, D. Meilinger, U. Rothbauer, C. M. Cardoso, and H. Leonhardt, "Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation." *Nucleic Acids Res.*, vol. 35, no. 13, pp. 4301–4312, 2007. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1934996/>
- [240] K. Schneider, C. Fuchs, A. Dobay, A. Rottach, W. Qin, P. Wolf, J. M. Álvarez-Castro, M. M. Nalaskowski, E. Kremmer, V. Schmid, H. Leonhardt, and L. Schermelleh, "Dissection of cell cycle-dependent dynamics of Dnmt1 by FRAP and diffusion-coupled modeling," *Nucleic Acids Res.*, vol. 41, no. 9, pp. 4860–4876, 2013. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/23535145/>
- [241] S. Ishiyama, A. Nishiyama, Y. Saeki, K. Moritsugu, D. Morimoto, L. Yamaguchi, N. Arai, R. Matsumura, T. Kawakami, Y. Mishima, H. Hojo, S. Shimamura, F. Ishikawa, S. Tajima, K. Tanaka, M. Ariyoshi, M. Shirakawa, M. Ikeguchi, A. Kidera, I. Suetake, K. Arita, and M. Nakanishi, "Structure of the Dnmt1 Reader Module Complexed with a Unique Two-Mono-Ubiquitin Mark on Histone H3 Reveals the Basis for DNA Methylation Maintenance," *Mol. Cell*, vol. 68, no. 2, pp. 350–360.e7, 2017. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/29053958/>
- [242] A. Nishiyama, C. B. Mulholland, S. Bultmann, S. Kori, A. Endo, Y. Saeki, W. Qin, C. Trummer, Y. Chiba, H. Yokoyama, S. Kumamoto, T. Kawakami, H. Hojo, G. Nagae, H. Aburatani, K. Tanaka, K. Arita, H. Leonhardt, and M. Nakanishi, "Two distinct modes of DNMT1 recruitment ensure stable maintenance DNA methylation," *Nat. Commun.*, vol. 11, p. 1222, 2020. [Online]. Available: <https://doi.org/10.1038/s41467-020-15006-4>
- [243] H. P. Easwaran, L. Schermelleh, H. Leonhardt, and M. C. Cardoso, "Replication-independent chromatin loading of Dnmt1 during G2 and M phases." *EMBO Rep*, vol. 5, no. 12, pp. 1181–1186, 2004. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1299190/>
- [244] M. Bachman, S. Uribe-Lewis, X. Yang, M. Williams, A. Murrell, and S. Balasubramanian, "5-Hydroxymethylcytosine is a predominantly stable DNA modification," *Nat Chem*, vol. 6, no. 12, pp. 1049–1055, 2014. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/25411882>
- [245] M. Bachman, S. Uribe-Lewis, X. Yang, H. E. Burgess, M. Iurlaro, W. Reik, A. Murrell, and S. Balasubramanian, "5-Formylcytosine can be a stable DNA modification in mammals," *Nat Chem Biol*, vol. 11, pp. 555–557, 2015. [Online]. Available: <https://www.nature.com/articles/nchembio.1848>
- [246] C. Alabert, T. K. Barth, N. Reverón-Gómez, S. Sidoli, A. Schmidt, O. Jensen, A. Imhof, and A. Groth, "Two distinct modes for propagation of histone PTMs across the cell cycle," *Genes Dev.*, vol. 29, no. 6, pp. 585–590, 2015. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4378191/>
- [247] A. T. Annunziato, "The fork in the road: Histone partitioning during DNA replication," *Genes (Basel)*, vol. 6, no. 2, pp. 353–371, 2015. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4488668/>
- [248] H. Tagami, D. Ray-Gallet, G. Almouzni, and Y. Nakatani, "Histone H3.1 and H3.3 Complexes Mediate Nucleosome Assembly Pathways Dependent or Independent of DNA Synthesis," *Cell*, vol. 116, no. 1, pp. 51–61, 2004. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/14718166/>
- [249] M. Xu, C. Long, X. Chen, C. Huang, S. Chen, and B. Zhu, "Partitioning of histone H3-H4 tetramers during DNA replication-dependent chromatin assembly," *Science (80-. )*, vol. 328, no. 5974, pp. 94–98, 2010. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/20360108/>
- [250] V. Jackson, "In vivo studies on the dynamics of histone-DNA interaction: evidence for nucleosome dissolution during replication and transcription and a low level of dissolution independent of both," *Biochemistry*, vol. 29, no. 3, pp. 719–731, 1990. [Online]. Available: <https://pubs.acs.org/doi/abs/10.1021/bi00455a019>
- [251] H. Tachiwana, M. Dacher, K. Maehara, A. Harada, Y. Seto, R. Katayama, Y. Ohkawa, H. Kimura, H. Kurumizaka, and N. Saitoh, "Chromatin structure-dependent histone incorporation revealed by a genome-wide deposition assay," *Elife*, vol. 10, 2021. [Online]. Available: <https://elifesciences.org/articles/66290>
- [252] J. J. Blow, P. J. Gillespie, D. Francis, and D. A. Jackson, "Replication origins in *Xenopus* egg extract Are 5-15

- 
- kilobases apart and are activated in clusters that fire at different times," *J Cell Biol*, vol. 152, no. 1, pp. 15–25, 2001. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/11149917>
- [253] J. Newport and M. Kirschner, "A major developmental transition in early *Xenopus* embryos: I. characterization and timing of cellular changes at the midblastula stage," *Cell*, vol. 30, no. 3, pp. 675–686, 1982. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/6183003/>
- [254] A. B. Blumenthal, H. J. Kriegstein, and D. S. Hogness, "The units of DNA replication in *Drosophila melanogaster* chromosomes," *CSH Symp. Quant.*, vol. 38, pp. 205–223, 1974. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/4208784>
- [255] J. Newport and M. Kirschner, "A major developmental transition in early *Xenopus* embryos: II. control of the onset of transcription," *Cell*, vol. 30, no. 3, pp. 687–696, 1982. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/7139712/>
- [256] O. Hyrien and M. Méchali, "Chromosomal replication initiates and terminates at random sequences but at regular intervals in the ribosomal DNA of *Xenopus* early embryos," *EMBO J*, vol. 12, no. 12, pp. 4511–4520, 1993. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/8223461>
- [257] O. Hyrien, C. Maric, and M. Méchali, "Transition in specification of embryonic metazoan DNA replication origins," *Science (80-. )*, vol. 270, no. 5238, pp. 994–997, 1995. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/7481806>
- [258] J. Walter and J. W. Newport, "Regulation of replicon size in *Xenopus* egg extracts," *Science (80-. )*, vol. 275, no. 5302, pp. 993–995, 1997. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/9020085>
- [259] B. A. Edgar and G. Schubiger, "Parameters controlling transcriptional activation during early *drosophila* development," *Cell*, vol. 44, no. 6, pp. 871–877, 1986. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/2420468/>
- [260] J. M. Lemaitre, E. Danis, P. Pasero, Y. Vassetzky, and M. Méchali, "Mitotic remodeling of the replicon and chromosome structure," *Cell*, vol. 123, no. 5, pp. 787–801, 2005. [Online]. Available: [https://www.cell.com/cell/abstract/S0092-8674\(05\)00973-6](https://www.cell.com/cell/abstract/S0092-8674(05)00973-6)
- [261] S. Courbet, S. Gay, N. Arnoult, G. Wronka, M. Anglana, O. Brison, and M. Debatisse, "Replication fork movement sets chromatin loop size and origin choice in mammalian cells," *Nature*, vol. 455, no. 7212, pp. 557–560, 2008. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/18716622>
- [262] M. L. Eaton, J. A. Prinz, H. K. MacAlpine, G. Tretyakov, P. V. Kharchenko, and D. M. MacAlpine, "Chromatin signatures of the *Drosophila* replication program," *Genome Res.*, vol. 21, no. 2, pp. 164–174, 2011. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3032920/>
- [263] N. Sher, G. W. Bell, S. Li, J. Nordman, T. Eng, M. L. Eaton, D. M. MacAlpine, and T. L. Orr-Weaver, "Developmental control of gene copy number by repression of replication initiation and fork progression," *Genome Res.*, vol. 22, no. 1, pp. 64–75, 2012. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3246207/>
- [264] P. Norio, S. Kosiyatrakul, Q. Yang, Z. Guan, N. M. Brown, S. Thomas, R. Riblet, and C. L. Schildkraut, "Progressive activation of DNA replication initiation in large domains of the immunoglobulin heavy chain locus during B cell development," *Mol. Cell*, vol. 20, no. 4, pp. 575–587, 2005. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/16307921/>
- [265] L. D. Mesner, V. Valsakumar, N. Karnani, A. Dutta, J. L. Hamlin, and S. Bekiranov, "Bubble-chip analysis of human origin distributions demonstrates a genomic scale significant clustering into zones and significant association with transcription," *Genome Res.*, vol. 21, no. 3, pp. 377–389, 2011. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/21173031/>
- [266] D. Grégoire, K. Brodolin, and M. Méchali, "HoxB domain induction silences DNA replication origins in the locus and specifies a single origin at its boundary," *EMBO Rep.*, vol. 7, no. 8, pp. 812–816, 2006. [Online]. Available:
-

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1525151/>

- [267] Y. H. Chen, S. Keegan, M. Kahli, P. Tonzi, D. Fenyő, T. T. Huang, and D. J. Smith, "Transcription shapes DNA replication initiation and termination in human cells," *Nat. Struct. Mol. Biol.*, vol. 26, no. 1, pp. 67–77, 2019. [Online]. Available: [/pmc/articles/PMC6320713/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6320713/)
- [268] J. Ferreira and M. Carmo-Fonseca, "Genome replication in early mouse embryos follows a defined temporal and spatial order," *J Cell Sci*, vol. 110, pp. 889–897, 1997. [Online]. Available: <https://jcs.biologists.org/content/110/7/889.long>
- [269] V. Dileep and D. M. Gilbert, "Single-cell replication profiling to measure stochastic variation in mammalian replication timing," *Nat. Commun.*, vol. 9, no. 1, pp. 1–8, 2018. [Online]. Available: [www.nature.com/naturecommunications](http://www.nature.com/naturecommunications)
- [270] R. K. Slotkin, "The case for not masking away repetitive DNA," *Mob. DNA*, vol. 9, no. 1, p. 15, 2018. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/29743957/>
- [271] M. Cremer, F. Grasser, C. Lanctôt, S. Müller, M. Neusser, R. Zinner, I. Solovei, and T. Cremer, "Multicolor 3D Fluorescence In Situ Hybridization for Imaging Interphase Chromosomes," *Methods Mol. Biol.*, vol. 463, pp. 205–239, 2012. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/18951171/>
- [272] J. A. Stamatoyannopoulos, I. Adzhubei, R. E. Thurman, G. V. Kryukov, S. M. Mirkin, and S. R. Sunyaev, "Human mutation rate associated with DNA replication timing," *Nat. Genet.*, vol. 41, no. 4, pp. 393–395, 2009. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/19287383/>
- [273] R. T. O'Keefe, S. C. Henderson, and D. L. Spector, "Dynamic organization of DNA replication in mammalian cell nuclei: spatially and temporally defined replication of chromosome-specific alpha-satellite DNA sequences," *J Cell Biol*, vol. 116, no. 5, pp. 1095–1110, 1992. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/1740468>
- [274] A. A. Goodarzi, A. T. Noon, D. Deckbar, Y. Ziv, Y. Shiloh, M. Löbrich, and P. A. Jeggo, "ATM Signaling Facilitates Repair of DNA Double-Strand Breaks Associated with Heterochromatin," *Mol. Cell*, vol. 31, no. 2, pp. 167–177, 2008. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/18657500/>
- [275] T. J. Pohl, B. J. Brewer, and M. K. Raghuraman, "Functional Centromeres Determine the Activation Time of Pericentric Origins of DNA Replication in *Saccharomyces cerevisiae*," *PLoS Genet.*, vol. 8, no. 5, p. e1002677, 2012. [Online]. Available: <https://dx.plos.org/10.1371/journal.pgen.1002677>
- [276] N. Arnoult, C. Schluth-Bolard, A. Letessier, I. Drascovic, R. Bouarich-Bourimi, J. Campisi, S. H. Kim, A. Boussouar, A. Ottaviani, F. Magdinier, E. Gilson, and A. Londono-Vallejo, "Replication timing of human telomeres is chromosome arm-specific, influenced by subtelomeric structures and connected to nuclear localization," *PLoS Genet*, vol. 6, no. 4, p. e1000920, 2010. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/20421929>
- [277] A. Bianchi and D. Shore, "Early Replication of Short Telomeres in Budding Yeast," *Cell*, vol. 128, no. 6, pp. 1051–1062, 2007. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/17382879/>
- [278] J. Huang, F. Wang, M. Okuka, N. Liu, G. Ji, X. Ye, B. Zuo, M. Li, P. Liang, W. W. Ge, J. C. Tsibris, D. L. Keefe, and L. Liu, "Association of telomere length with authentic pluripotency of ES/iPS cells," *Cell Res.*, vol. 21, no. 5, pp. 779–792, 2011. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/21283131/>
- [279] A. W. Shermoen, M. L. McClelland, and P. H. O'Farrell, "Developmental control of late replication and S-phase length," *Curr Biol*, vol. 20, no. 23, pp. 2067–2077, 2010. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21074439>
- [280] K. Ahmed, H. Dehghani, P. Rugg-Gunn, E. Fussner, J. Rossant, and D. P. Bazett-Jones, "Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo," *PLoS One*, vol. 5, no. 5, 2010. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/20479880/>
- [281] E. Fussner, U. Djuric, M. Strauss, A. Hotta, C. Perez-Iratxeta, F. Lanner, F. J. Dilworth, J. Ellis, and D. P. Bazett-Jones, "Constitutive heterochromatin reorganization during somatic cell reprogramming," *EMBO J.*, vol. 30, no. 9, pp. 1778–1789, 2011. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/21468033/>

- 
- [282] E. Fussner, M. Strauss, U. Djuric, R. Li, K. Ahmed, M. Hart, J. Ellis, and D. P. Bazett-Jones, "Open and closed domains in the mouse genome are configured as 10-nm chromatin fibres," *EMBO Rep.*, vol. 13, no. 11, pp. 992–996, 2012. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/22986547/>
- [283] V. Hassan-Zadeh, P. Rugg-Gunn, and D. P. Bazett-Jones, "DNA methylation is dispensable for changes in global chromatin architecture but required for chromocentre formation in early stem cell differentiation," *Chromosoma*, vol. 126, no. 5, pp. 605–614, 2017. [Online]. Available: <https://link.springer.com/article/10.1007/s00412-017-0625-x>
- [284] C. L. F. Woodcock, L. L. . Y. Frado, and J. . B. . Rattner, "The Higher-order Structure of Chromatin : Evidence for a Helical Ribbon Arrangement," *J. Cell Biol.*, vol. 99, pp. 42–52, 1984. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/6736132/>
- [285] D. P. Bazett-Jones, "Electron spectroscopic imaging of chromatic and other nucleoprotein complexes," *Electron Microsc. Rev.*, vol. 5, no. 1, pp. 37–58, 1992. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/1730076/>
- [286] B. A. Gibson, L. K. Doolittle, M. W. Schneider, L. E. Jensen, N. Gamarra, L. Henry, D. W. Gerlich, S. Redding, and M. K. Rosen, "Organization of Chromatin by Intrinsic and Regulated Phase Separation," *Cell*, vol. 179, no. 2, pp. 470–484.e21, 2019. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/31543265/>
- [287] E. Meshorer, D. Yellajoshula, E. George, P. J. Scambler, D. T. Brown, and T. Misteli, "Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells," *Dev Cell*, vol. 10, no. 1, pp. 105–116, 2006. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/16399082>
- [288] A. N. Prusov and O. V. Zatsepina, "Isolation of the chromocenter fraction from mouse liver nuclei," *Biochem.*, vol. 67, no. 4, pp. 423–431, 2002. [Online]. Available: <https://link.springer.com/article/10.1023/A:1015229922183>
- [289] S. Efroni, R. Duttagupta, J. Cheng, H. Dehghani, D. J. Hoepfner, C. Dash, D. P. Bazett-Jones, S. Le Grice, R. D. McKay, K. H. Buetow, T. R. Gingeras, T. Misteli, and E. Meshorer, "Global transcription in pluripotent embryonic stem cells," *Cell Stem Cell*, vol. 2, no. 5, pp. 437–447, 2008. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/18462694>
- [290] V. J. Schmid, M. Cremer, and T. Cremer, "Quantitative analyses of the 3D nuclear landscape recorded with super-resolved fluorescence microscopy," *Methods*, vol. 123, pp. 33–46, 2017. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/28323041/>
- [291] C. S. Casas-Delucchi, A. Brero, H. P. Rahn, I. Solovei, A. Wutz, T. Cremer, H. Leonhardt, and M. C. Cardoso, "Histone acetylation controls the inactive X chromosome replication dynamics," *Nat Commun*, vol. 2, p. 222, 2011. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21364561>
- [292] B. L. Aken, S. Ayling, D. Barrell, L. Clarke, V. Curwen, S. Fairley, J. Fernandez Banet, K. Billis, C. Garcia Giron, T. Hourlier, K. Howe, A. Kahari, F. Kokocinski, F. J. Martin, D. N. Murphy, R. Nag, M. Ruffier, M. Schuster, Y. A. Tang, J. H. Vogel, S. White, A. Zadissa, P. Flicek, and S. M. Searle, "The Ensembl gene annotation system," *Database (Oxford)*, vol. 2016, 2016. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/27337980>
- [293] H. Skaletsky, T. Kuroda-Kawaguchi, P. J. Minx, H. S. Cordum, L. D. Hillier, L. G. Brown, S. Reppng, T. Pyntikova, J. All, T. Blerl, A. Chinwalla, A. Delehaunty, H. Du, G. Fewell, L. Fulton, R. Fulton, T. Graves, S. F. Hou, P. Latrielle, S. Leonard, E. Mardis, R. Maupin, J. McPherson, T. Miner, W. Nash, C. Nguyen, P. Ozersky, K. Pepin, S. Rock, T. Rohlfing, K. Scott, B. Schultz, C. Strong, A. Tin-Wollam, S. P. Yang, R. H. Waterston, R. K. Willson, S. Rozen, and D. C. Page, "The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes," *Nature*, vol. 423, no. 6942, pp. 825–837, 2003. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/12815422/>
- [294] R. Mayer, A. Brero, J. von Hase, T. Schroeder, T. Cremer, and S. Dietzel, "Common themes and cell type specific variations of higher order chromatin arrangements in the mouse," *BMC Cell Biol*, vol. 6, p. 44, 2005. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/16336643>
- [295] A. A. Maan, J. Eales, A. Akbarov, J. Rowland, X. Xu, M. A. Jobling, F. J. Charchar, and M. Tomaszewski, "The Y

- chromosome: A blueprint for men's health?" *Eur. J. Hum. Genet.*, vol. 25, no. 11, pp. 1181–1188, 2017. [Online]. Available: [www.nature.com/ejhg](http://www.nature.com/ejhg)
- [296] B. Charlesworth, "The organization and evolution of the human Y chromosome," *Genome Biol.*, vol. 4, no. 9, p. 226, 2003. [Online]. Available: <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2003-4-9-226>
- [297] L. Quintana-Murci and M. Fellous, "The human Y chromosome: The biological role of a "functional wasteland"," *J. Biomed. Biotechnol.*, vol. 2001, no. 1, pp. 18–24, 2001. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC79676/>
- [298] J. C. Cadoret, F. Meisch, V. Hassan-Zadeh, I. Luyten, C. Guillet, L. Duret, H. Quesneville, and M. N. Prioleau, "Genome-wide studies highlight indirect links between human replication origins and gene regulation," *PNAS*, vol. 105, no. 41, pp. 15 837–15 842, 2008. [Online]. Available: <https://www.pnas.org/content/105/41/15837>
- [299] M. Costantini and G. Bernardi, "Replication timing, chromosomal bands, and isochores," *PNAS*, vol. 105, no. 9, pp. 3433–3437, 2008. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2265141/>
- [300] F. Picard, J. C. Cadoret, B. Audit, A. Arneodo, A. Alberti, C. Battail, L. Duret, and M. N. Prioleau, "The Spatiotemporal Program of DNA Replication Is Associated with Specific Combinations of Chromatin Marks in Human Cells," *PLoS Genet.*, vol. 10, no. 5, 2014. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/24785686/>
- [301] C. W. Gilbert, S. Muldal, L. G. Lajtha, and J. Rowley, "Time-sequence of human chromosome duplication," *Nature*, vol. 195, no. 4844, pp. 869–873, 1962. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/13898534/>
- [302] A. Morishima, M. M. Grumbach, and J. H. Taylor, "Asynchronous duplication of human chromosomes and the origin of sex chromatin," *PNAS*, vol. 48, no. 2, pp. 756–763, may 1962. [Online]. Available: <https://www.pnas.org/content/48/5/756>
- [303] N. Takagi, "Differentiation of X chromosomes in early female mouse embryos," *Exp. Cell Res.*, vol. 86, no. 1, pp. 127–135, 1974. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/4133904/>
- [304] B. P. Chadwick and H. F. Willard, "Chromatin of the Barr body: Histone and non-histone proteins associated with or excluded from the inactive X chromosome," *Hum. Mol. Genet.*, vol. 12, no. 17, pp. 2167–2178, 2003. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/12915472/>
- [305] J. H. Chuang and H. Li, "Functional bias and spatial organization of genes in mutational hot and cold regions in the human genome," *PLoS Biol.*, vol. 2, no. 2, p. e29, 2004. [Online]. Available: <http://biology.plosjournals.org>
- [306] L. Zhang and W. H. Li, "Mammalian Housekeeping Genes Evolve More Slowly than Tissue-Specific Genes," *Mol. Biol. Evol.*, vol. 21, no. 2, pp. 236–239, 2004. [Online]. Available: <https://europepmc.org/article/med/14595094>
- [307] M. Cardoso-Moreira, J. J. Emerson, A. G. Clark, and M. Long, "Drosophila Duplication Hotspots Are Associated with Late-Replicating Regions of the Genome," *PLoS Genet.*, vol. 7, no. 11, p. e1002340, 2011. [Online]. Available: <https://dx.plos.org/10.1371/journal.pgen.1002340>
- [308] D. Juan, D. Rico, T. Marques-Bonet, Ó. Fernández-Capetillo, and A. Valencia, "Late-replicating CNVs as a source of new genes," *Biol. Open*, vol. 2, no. 12, pp. 1402–1411, 2013. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/24285712/>
- [309] Y. Q. Soh, J. Alföldi, T. Pyntikova, L. G. Brown, T. Graves, P. J. Minx, R. S. Fulton, C. Kremitzki, N. Koutseva, J. L. Mueller, S. Rozen, J. F. Hughes, E. Owens, J. E. Womack, W. J. Murphy, Q. Cao, P. De Jong, W. C. Warren, R. K. Wilson, H. Skaletsky, and D. C. Page, "Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes," *Cell*, vol. 159, no. 4, pp. 800–813, 2014. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/25417157/>
- [310] D. Bachrog, "Y-chromosome evolution: Emerging insights into processes of Y-chromosome degeneration," pp. 113–124, 2013. [Online]. Available: <https://www.nature.com/articles/nrg3366>
- [311] S. B. Buonomo, "Rif1-dependent regulation of genome replication in mammals," *Adv Exp Med Biol*, vol. 1042, pp. 259–272, 2017. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/29357062/>

- 
- [312] T. Hamatani, M. G. Carter, A. A. Sharov, and M. S. Ko, "Dynamics of global gene expression changes during mouse preimplantation development," *Dev Cell*, vol. 6, no. 1, pp. 117–131, 2004. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/14723852/>
- [313] M. E. Torres-Padilla and M. Zernicka-Goetz, "Role of TIF1 $\alpha$  as a modulator of embryonic transcription in the mouse zygote," *J. Cell Biol.*, vol. 174, no. 3, pp. 329–338, 2006. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2064229/>
- [314] C. S. Sørensen, R. G. Syljuåsen, J. Falck, T. Schroeder, L. Rønnstrand, K. K. Khanna, B. B. Zhou, J. Bartek, and J. Lukas, "Chk1 regulates the S phase checkpoint by coupling the physiological turnover and ionizing radiation-induced accelerated proteolysis of Cdc25A," *Cancer Cell*, vol. 3, no. 3, pp. 247–258, 2003. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/12676583/>
- [315] D. Shechter, V. Costanzo, and J. Gautier, "ATR and ATM regulate the timing of DNA replication origin firing," *Nat. Cell Biol.*, vol. 6, no. 7, pp. 648–655, 2004. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/15220931/>
- [316] A. Maya-Mendoza, E. Petermann, D. A. Gillespie, K. W. Caldecott, and D. A. Jackson, "Chk1 regulates the density of active replication origins during the vertebrate S-phase," *EMBO J.*, vol. 26, no. 11, pp. 2719–2731, 2007. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/17491592/>
- [317] N. Petryk, M. Kahli, Y. D'Aubenton-Carafa, Y. Jaszczyszyn, Y. Shen, M. Silvain, C. Thermes, C. L. Chen, and O. Hyrien, "Replication landscape of the human genome," *Nat Commun*, vol. 7, p. 10208, 2016. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/26751768>
- [318] R. Lebofsky and A. Bensimon, "DNA Replication Origin Plasticity and Perturbed Fork Progression in Human Inverted Repeats," *Mol. Cell. Biol.*, vol. 25, no. 15, pp. 6789–6797, 2005. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1190330/>
- [319] M. Wooten, J. Snedeker, Z. F. Nizami, X. Yang, R. Ranjan, E. Urban, J. M. Kim, J. Gall, J. Xiao, and X. Chen, "Asymmetric histone inheritance via strand-specific incorporation and biased replication fork movement," *Nat. Struct. Mol. Biol.*, vol. 26, no. 8, pp. 732–743, 2019. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6684448/>
- [320] G. Guilbaud, A. Rappailles, A. Baker, C.-L. Chen, A. Arneodo, A. Goldar, Y. D'Aubenton-Carafa, C. Thermes, B. Audit, and O. Hyrien, "Evidence for Sequential and Increasing Activation of Replication Origins along Replication Timing Gradients in the Human Genome," *PLoS Comput. Biol.*, vol. 7, no. 12, p. e1002322, 2011. [Online]. Available: <https://dx.plos.org/10.1371/journal.pcbi.1002322>
- [321] Y. Watanabe, A. Fujiyama, Y. Ichiba, M. Hattori, T. Yada, Y. Sakaki, and T. Ikemura, "Chromosome-wide assessment of replication timing for human chromosomes 11q and 21q: Disease-related genes in timing-switch regions," *Hum. Mol. Genet.*, vol. 11, no. 1, pp. 13–21, 2002. [Online]. Available: <http://www.ncbi.nlm.nih.gov/Omim/>
- [322] A. Tsumura, T. Hayakawa, Y. Kumaki, S. Takebayashi, M. Sakaue, C. Matsuoka, K. Shimotohno, F. Ishikawa, E. Li, H. R. Ueda, J. Nakayama, and M. Okano, "Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b," *Genes Cells*, vol. 11, no. 7, pp. 805–814, 2006. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/16824199>
- [323] E. A. Raiber, P. Murat, D. Y. Chirgadze, D. Beraldi, B. F. Luisi, and S. Balasubramanian, "5-formylcytosine alters the structure of the DNA double helix," *Nat. Struct. Mol. Biol.*, vol. 22, no. 1, pp. 44–49, 2015. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/25504322><http://dx.doi.org/10.1038/nsmb.2936>
- [324] C. M. Lopez, A. J. Lloyd, K. Leonard, and M. J. Wilkinson, "Differential effect of three base modifications on DNA thermostability revealed by high resolution melting," *Anal Chem*, vol. 84, no. 17, pp. 7336–7342, 2012. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/22882125>
- [325] J. S. Hardwick, D. Ptchelkine, A. H. El-Sagheer, I. Tear, D. Singleton, S. E. Phillips, A. N. Lane, and T. Brown, "5-Formylcytosine does not change the global structure of DNA," *Nat. Struct. Mol. Biol.*, vol. 24, no. 6, pp. 544–552, 2017. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/28504696/>

- [326] D. Renciuik, O. Blacque, M. Vorlickova, and B. Spingler, "Crystal structures of B-DNA dodecamer containing the epigenetic modifications 5-hydroxymethylcytosine or 5-methylcytosine," *Nucleic Acids Res.*, vol. 41, no. 21, pp. 9891–9900, 2013. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23963698>
- [327] A. Thalhammer, A. S. Hansen, A. H. El-Sagheer, T. Brown, and C. J. Schofield, "Hydroxylation of methylated CpG dinucleotides reverses stabilisation of DNA duplexes by cytosine 5-methylation," *Chem Commun*, vol. 47, no. 18, pp. 5325–5327, 2011. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21451870>
- [328] P. Zhang, F. D. Hastert, A. K. Ludwig, K. Breitwieser, M. Hofstätter, and M. C. Cardoso, "DNA base flipping analytical pipeline," *Biol. Methods Protoc.*, vol. 2, no. 1, pp. 1–12, 2017. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/32161792/>
- [329] C. Jin, Y. Lu, J. Jelinek, S. Liang, M. R. H. Estecio, M. C. Barton, and J.-P. J. P. J. Issa, "TET1 is a maintenance DNA demethylase that prevents methylation spreading in differentiated cells," *Nucleic Acids Res.*, vol. 42, no. 11, pp. 6956–6971, 2014. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/24875481/>
- [330] A. Pérez, C. L. Castellazzi, F. Battistini, K. Collinet, O. Flores, O. Deniz, M. L. Ruiz, D. Torrents, R. Eritja, M. Soler-López, M. Orozco, A. Perez, C. L. Castellazzi, F. Battistini, K. Collinet, O. Flores, O. Deniz, M. L. Ruiz, D. Torrents, R. Eritja, M. Soler-Lopez, M. Orozco, A. Pérez, C. L. Castellazzi, F. Battistini, K. Collinet, O. Flores, O. Deniz, M. L. Ruiz, D. Torrents, R. Eritja, M. Soler-López, and M. Orozco, "Impact of methylation on the physical properties of DNA," *Biophys. J.*, vol. 102, no. 9, pp. 2140–2148, 2012. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/22824278>
- [331] A. C. D. Machado, T. Zhou, S. Rao, P. Goel, C. Rastogi, A. Lazarovici, H. J. Bussemaker, R. Rohs, A. C. Dantas Machado, T. Zhou, S. Rao, P. Goel, C. Rastogi, A. Lazarovici, H. J. Bussemaker, and R. Rohs, "Evolving insights on how cytosine methylation affects protein-DNA binding," *Brief. Funct. Genomics*, vol. 14, no. 1, pp. 61–73, 2015. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/25319759>
- [332] M. Wanunu, D. Cohen-Karni, R. R. Johnson, L. Fields, J. Benner, N. Peterman, Y. Zheng, M. L. Klein, M. Drndic, R. R. Johnson, D. Cohen-Karni, M. Drndic, N. Peterman, M. L. Klein, J. Benner, Y. Zheng, and M. Wanunu, "Discrimination of Methylcytosine from Hydroxymethylcytosine in DNA Molecules," *J. Am. Chem. Soc.*, vol. 133, no. 3, pp. 486–492, 2010. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21155562>
- [333] Q. Dai, P. J. Sanstead, C. S. Peng, D. Han, C. He, and A. Tokmakoff, "Weakened N3 hydrogen bonding by 5-formylcytosine and 5-carboxylcytosine reduces their base-pairing stability," *ACS Chem Biol*, vol. 11, no. 2, pp. 470–477, 2016. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/26641274>
- [334] N. Arnaud-Barbe, V. Cheynet-Sauvion, G. Oriol, B. Mandrand, and F. Mallet, "Transcription of RNA templates by T7 RNA polymerase," *Nucleic Acids Res.*, vol. 26, no. 15, pp. 3550–3554, 1998. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/9671817/>
- [335] M. W. Kellinger, C. X. Song, J. Chong, X. Y. Lu, C. He, and D. Wang, "5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription," *Nat. Struct. Mol. Biol.*, vol. 19, no. 8, pp. 831–833, 2012. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/22820989/>
- [336] P. E. Merkl, M. Pils, T. Fremter, K. Schwank, C. Engel, G. Längst, P. Milkereit, J. Griesenbeck, and H. Tschochner, "RNA polymerase I (Pol I) passage through nucleosomes depends on Pol I subunits binding its lobe structure," *J. Biol. Chem.*, vol. 295, no. 15, pp. 4782–4795, 2020. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/32060094/>
- [337] E. P. Ramsay, G. Abascal-Palacios, J. L. Daiß, H. King, J. Gouge, M. Pils, F. Beuron, E. Morris, P. Gunkel, C. Engel, and A. Vannini, "Structure of human RNA polymerase III," *Nat. Commun.*, vol. 11, no. 1, pp. 1–12, 2020. [Online]. Available: <https://doi.org/10.1038/s41467-020-20262-5>
- [338] J. J. Li and T. J. Kelly, "Simian virus 40 DNA replication in vitro," *PNAS*, vol. 81, no. 22, pp. 6973–6977, 1984. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/6095264/>
- [339] J. H. Proffitt, J. R. Davie, D. Swinton, and S. Hattman, "5-Methylcytosine is not detectable in *Saccharomyces*

- 
- cerevisiae DNA." *Mol. Cell. Biol.*, vol. 4, no. 5, pp. 985–988, 1984. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/6374428/>
- [340] S. M. Gorisch, A. Sporbert, J. H. Stear, I. Grunewald, D. Nowak, E. Warbrick, H. Leonhardt, and M. C. Cardoso, "Uncoupling the replication machinery: replication fork progression in the absence of processive DNA synthesis," *Cell Cycle*, vol. 7, no. 13, pp. 1983–1990, 2008. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/18604168>
- [341] A. Abeles, T. Brendler, and S. Austin, "Evidence of two levels of control of P1 oriR and host oriC replication origins by DNA adenine methylation," *J Bacteriol*, vol. 175, no. 24, pp. 7801–7807, 1993. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/8253669>
- [342] H. Yamaki, O. E. N. K, and M. Y, "The oriC unwinding by dam methylation in Escherichia coli," *Nucleic Acids Res.*, vol. 44, no. 1, pp. 5067–5073, 2016. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC336717/>
- [343] S. Schiffers, C. Ebert, R. Rahimoff, O. Kosmatchev, J. Steinbacher, A. V. Bohne, F. Spada, S. Michalakis, J. Nickelsen, M. Muller, and T. Carell, "Quantitative LC-MS Provides No Evidence for m(6) dA or m(4) dC in the Genome of Mouse Embryonic Stem Cells and Tissues," *Angew Chem Int Ed Engl*, vol. 56, no. 37, pp. 11268–11271, 2017. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/28371147>
- [344] T. P. Wu, T. Wang, M. G. Seetin, Y. Lai, S. Zhu, K. Lin, Y. Liu, S. D. Byrum, S. G. Mackintosh, M. Zhong, A. Tackett, G. Wang, L. S. Hon, G. Fang, J. A. Swenberg, and A. Z. Xiao, "DNA methylation on N(6)-adenine in mammalian embryonic stem cells," *Nature*, vol. 532, no. 7599, pp. 329–333, 2016. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/27027282>
- [345] C. L. Xiao, S. Zhu, M. He, Chen, Q. Zhang, Y. Chen, G. Yu, J. Liu, S. Q. Xie, F. Luo, Z. Liang, D. P. Wang, X. C. Bo, X. F. Gu, K. Wang, and G. R. Yan, "N(6)-Methyladenine DNA Modification in the Human Genome," *Mol. Cell*, vol. 71, no. 2, pp. 306–318 e7, 2018. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/30017583>
- [346] X. Liu, W. Lai, Y. Li, S. Chen, B. Liu, N. Zhang, J. Mo, C. Lyu, J. Zheng, Y.-R. Du, G. Jiang, G.-L. Xu, and H. Wang, "N6-methyladenine is incorporated into mammalian genome by DNA polymerase," *Cell Res*, vol. 31, no. 1, pp. 94–97, 2020. [Online]. Available: <https://www.nature.com/articles/s41422-020-0317-6>
- [347] T. Prikrylova, J. Robertson, F. Ferrucci, D. Konorska, H. Aanes, A. Manaf, B. Zhang, C. B. Vågbø, A. Kuśnierczyk, K. M. Gilljam, C. Løvkvam-Køster, M. Otterlei, J. A. Dahl, J. Enserink, A. Klungland, A. B. Robertson, C. B. Vagbo, A. Kusnierczyk, K. M. Gilljam, C. Lovkvam-Koster, M. Otterlei, J. A. Dahl, J. Enserink, A. Klungland, and A. B. Robertson, "5-hydroxymethylcytosine Marks Mammalian Origins Acting as a Barrier to Replication," *Sci Rep*, vol. 9, no. 1, p. 11065, 2019. [Online]. Available: <http://www.nature.com/articles/s41598-019-47528-3>
- [348] P. Zhang, C. Rausch, F. D. Hastert, B. Boneva, A. Filatova, S. J. Patil, U. A. Nuber, Y. Gao, X. Zhao, and M. C. Cardoso, "Methyl-CpG binding domain protein 1 regulates localization and activity of Tet1 in a CXXC3 domain-dependent manner," *Nucleic Acids Res.*, vol. 45, no. 12, pp. 7118–7136, 2017. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/28449087>
- [349] P. Zhang, A. K. Ludwig, F. D. Hastert, C. Rausch, A. Lehmkuhl, I. Hellmann, M. Smets, H. Leonhardt, and M. C. Cardoso, "L1 retrotransposition is activated by Ten-eleven-translocation protein 1 and repressed by methyl-CpG binding proteins," *Nucleus*, vol. 8, no. 5, pp. 548–562, 2017. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/28524723>

## 8 Annex

### 8.1 Abbreviations

A	Adenine
aa	Amino acid
Aph	Aphidicolin
AML	Acute myeloid leukemia
ARS	Autonomously replicating sequence
ATP	Adenosine triphosphate
BAZ2A/B	Bromodomain Adjacent To Zinc Finger Domain 2A/B
BER	Base excision repair
bp	Base pair(s)
BrdU	5-Bromo-2'-deoxyuridine
C	Cytosine
CC	Chromocenter
CDK	Cyclin-dependent kinase
cDNA	Complementary DNA
CTCF	CCCTC-Binding Factor
CGI	CpG island
ChIP-Seq	Chromatin immunoprecipitation followed by DNA sequencing
CI	Confidence interval
CldU	5-Chloro-2'-deoxyuridine
c-Myc	Cellular myelocytomatosis
CTD	C-terminal domain
CXXC	Cysteine/X/X/cysteine zinc finger motif
Cy3	Cyanine 3
Cy5	Cyanine 5
DAPI	4',6-diamidino-2-phenylindole
d5caC	2'-Deoxy-5-carboxyl-cytosine
d5fC	2'-Deoxy-5-formyl-cytosine
d5hmC	2'-Deoxy-5-hydroxymethyl-cytosine
d5mC	2'-Deoxy-5-methyl-cytosine
d5caCTP	2'-Deoxy-5-carboxyl-cytosine 5'-triphosphate
d5fCTP	2'-Deoxy-5-formyl-cytosine 5'-triphosphate
d5hmCTP	2'-Deoxy-5-hydroxymethyl-cytosine 5'-triphosphate
d5mCTP	2'-Deoxy-5-methyl-cytosine 5'-triphosphate
dATP	2'-Deoxyadenosine 5'-triphosphate
dCTP	2'-Deoxycytosine 5'-triphosphate
dGTP	2'-Deoxyguanosine 5'-triphosphate
DDK	Dbf4-dependent kinase

DDR	DNA damage response
DMEM	Dulbecco's modified eagle medium
DMSO	Dimethylsulfoxid
DNA	Deoxyribonucleic acid
Dnmt	DNA methyltransferase
dNTP	Deoxyribonucleoside 5'-triphosphate
DSB	Double-strand break
DSBH	Double-stranded beta helix
DTT	Dithiothreitol
dTTP	2'-Deoxythymidine 5'-triphosphate
dUTP	2'-Deoxyuridine 5'-triphosphate
EDT	Euclidian distance transform
EDTA	Ethylenediaminetetraacetic acid
EdU	5-Ethynyl-2'-deoxyuridine
ES cell	Embryonic stem cell
ESI	Electron spectroscopic Imaging
EU	5-Ethynyl-uridine
exo <sup>-</sup>	Exonuclease deficient
FACS	Fluorescence activated cell sorting
FBS	Fetal bovine serum
Fe	Iron
FISH	Fluorescence in situ hybridization
FRAP	Fluorescence recovery after photobleaching
γH2AX	Phosphorylation of histone H2AX
G-phase	Gap phase
G	Guanine
Gb	Gigabase pair(s)
GBP	GFP binding protein
gDNA	Genomic DNA
GFP	Green fluorescent protein
Gy	Gray
H3K9ac	Acetylation of histone 3 at lysine 9
H3K9me3	Tri-methylation of histone 3 at lysine 9
H4K5ac	Acetylation of histone 4 at lysine 5
H4K8ac	Acetylation of histone 4 at lysine 8
H4K12ac	Acetylation of histone 4 at lysine 12
HDAC	Histone deacetylase
HEK	Human embryonic kidney
HRM	High-resolution melting temperature analysis
HRP	Horseradish peroxidase
ID	Intervening domain

IdU	5-iodo-2'-deoxyuridine
Idax	Inhibition of the Dvl and Axin complex
IOD	Inter-origin distance
IR	Ionizing irradiation
IS	Initiation site
IVT	<i>In vitro</i> transcription
IZ	Initiation zone
kDa	Kilo dalton
kb	Kilobase pair(s)
Klf4	Kruppel-like factor 4
KO	Knock-out
L1 & LINE1	Long interspersed nuclear element 1
LIF	Leukemia inhibitory factor
LTR	Long terminal repeat
M	Marker
m6A	N6-methyladenine
MaSat	Major Satellite
Mb	Megabase pair(s)
MBD	Methyl-CpG binding domain
Mbd1-6	Methyl-CpG binding domain proteins 1-6
MBP	Methyl-CpG binding protein
mch	mcherry
MCM helicase	Minichromosome maintenance helicase
mCpG	Methylated CpG dinucleotide
Mecp2	Methyl-CpG binding protein 2
mES cells	Mouse embryonic stem cells
MEF	Mouse embryonic fibroblasts
MiSat	Minor Satellite
miRFP	Monomeric near infrared fluorescent protein
mNSC	Mouse neuronal stem cells
mRFP	Monomeric red fluorescent protein
MLL	Mixed lineage leukemia
MTase	Methyltransferase
nanoRFi	Nano replication foci
NCoR	Nuclear receptor corepressor
NLS	Nuclear localization signal
nt	Nucleotides
NTD	N-terminal domain
NuRD	Nucleosome Remodeling Deacetylase
Oct4	Octamer-binding transcription factor 4
Ok-seq	Okazaki fragment sequencing

ORF	Open reading frame
ORI	Origin of replication
p53	Tumor suppressor protein of 53 kDa
PBS	Phosphate buffered saline
PBST	PBS containing Tween
Pen/Strep	Penicilin/Streptavidin
Ph	Phase contrast
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PEI	Polyethylenimine
PTM	Post-translational modification
pWF	Pseudo widefield
PZF	Polydactyl zinc finger
RA	Retinoic Acid
rDNA	Ribosomal DNA
Repli-FISH	Combination of DNA replication detection (Repli) and fluorecence in situ hybridization (FISH)
RF	Replication focus
RFi	Replication foci
RFS	Replication fork speed
Repli-Seq	Replicated DNA sequencing
RNA	Ribonucleic acid
ROI	Region of interest
RPA	Replication protein A
RTT	Rett syndrome
S <sub>e</sub>	Early S-phase
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SEM	Standard error of the mean
S <sub>l</sub>	Late S-phase
S <sub>m</sub>	Mid S-phase
SNS-seq	Short-nascent-strand sequencing
S-phase	Synthesis phase
SINE	Short interspersed nuclear element
Sox2	SRY (sex determining region Y)-box 2
SSB	Single stranded binding protein
SSC	Saline-sodium citrate
StDev	Standard deviation
S <sub>ve</sub>	Very early S-phase
S <sub>Y</sub>	Male specific S-phase substage of Y chromosome replication
T	Thymine
TAD	Topologically associated domain
TDG	Thymine DNA glycosylase

---

TET	Ten eleven translocation
TET1CD	Tet1 catalytic domain
TKO	Triple knock-out
T <sub>m</sub>	Melting temperature
TPRT	Targeted primed reverse transcription
TRD	Transcriptional repression domain
TS	Targeting sequence
TSA	Trichostatin A
Uhrf1	Ubiquitin-like with plant homeodomain and ring finger domains 1
UTR	Untranslated region
wt	Wild type
2i	2 Inhibitors (Gsk3 $\beta$ and MEK inhibitors)
3D-SIM	3-Dimensional structured illumination microscopy
5caC	5-Carboxylcytosine
5fC	5-Formylcytosine
5hmC	5-Hydroxymethylcytosine
5hmU	5-Hydroxymethyluracil
5mC	5-Methylcytosine
6-FAM	6-Carboxyfluorescein



## 8.2 List of Contributions

### 8.2.1 DNA Modification Readers and Writers and Their Interplay

C.R. prepared figure 3 and wrote the text for sections 1 (Introduction), 2 (Effect of cytosine modifications on DNA double helix structure and recognition by proteins), 3 (DNMT writers, MBD readers, and TET modifiers), 3.1 (DNMT writer enzymes and the 5mC mark), 3.2 (MBD reader proteins), 3.2.1 (MecP2), 3.2.2 (Mbd1), 3.2.3 (Mbd2), 3.2.4 (Mbd3), 3.2.5 (Mbd4), 4 (DNA methylation and binding of MBD readers), 4.1 (TET modifiers and oxidized cytosine variants), 4.1.1 (Ten-eleven-translocation proteins, oxidative cytosine variants, and DNA demethylation), 4.1.2 (Genomic distribution of oxidative cytosine variants), 5 (Interplay of DNA methylation readers and modifiers) and 6 (Concluding remarks/Perspectives). F.H. prepared figures 1 and 2 and text sections 3.1 (DNMT writer enzymes and the 5mC mark) and 4.1.3 (TET isoforms). M.C.C. conceived and supervised the project and wrote the manuscript.

### 8.2.2 Repli-FISH: Application of 3D-(Immuno)-FISH for the Study of DNA Replication Timing of Genetic Repeat Elements

C.R. performed experiments for figures 4C, 5, 7, S2 and S3, analyzed data in figures 4A-C, 5, 6, 7, S2 and S3, prepared figures and wrote the manuscript. P.W. performed experiments for figures 4C, 5, 8, S2 and S4, analyzed data in figures 4C, 5, 6, 8, S2 and S4, prepared figures and wrote the manuscript. A.S. performed experiments for figures 4A-B and 6. M.C.C. conceived and supervised the project and wrote the manuscript.

### 8.2.3 DNA Replication and Repair Kinetics of Alu, LINE-1 and Satellite III Genomic Repetitive Elements

C.R. contributed to data analysis in figure S3a. F.N. performed the ChIP-Seq experiment, genomic data analysis and wrote the manuscript. A.R. performed the ChIP-Seq experiment, genomic data analysis, image analysis and wrote the manuscript. A.S. performed and analyzed FISH experiments and wrote the manuscript. W.Y. performed genomic data analysis. M.C.C. conceived and supervised the project and wrote the manuscript.

### 8.2.4 Developmental Differences in Genome Replication Program and Origin Activation

C.R. performed experiments for figures 1A-D, 2, 4B, 4D, 5, S8, S10, S14, S15, S18 and S19 and analyzed data from figures 1, 2, 3, 4, 5, 6, 7B-C, 7F-H, S8, S10, S14, S15, S17, S18 and S19. C.R. prepared figures and wrote the manuscript. P.W. performed experiments for figures 1, 2B-D, 3, 4C, 5, 6, 7B-C, 7F-H, S6, S8, S11, S15, S16, S17 and S19 and analyzed data from figures 1, 2B-C, 4, 5, 6, 7B-C, 7F-H, S8, S11, S15, S16 and S19. P.P. performed genome wide origin mapping analysis shown in figures 7D, 7E, S20, S21 and S22. D.H. performed RFI feature analysis from figure 1E, S12 and S13. A.M. performed 3D SIM imaging in figure 6. A.L. performed growth curve measurement in figure 5B. V.C. contributed to image analysis in figures 6B and 6D-F. C.C-D. conceived and supervised the project. H.L. contributed to 3D SIM imaging in figure 6. M.C.C. conceived and supervised the project and wrote the manuscript.

**8.2.5 Cytosine Base Modifications Regulate DNA Duplex Stability and Metabolism**

C.R. performed experiments for figures 1, 2, 3, 4, 5, 6, 7, S8, S10, S11, S12, S13B, S14, S15B, S16, S17, S18, S19 and S20 and generated and characterized templates used in figures S13A and S15A. C.R. analyzed data for figures 1, 2, 3, 4, 5, 6, 7, S8, S10, S11, S12, S13B, S14, S15B, S16, S17, S18, S19 and S20. C.R. prepared final figures and wrote the manuscript. P.Z. conceived the project, performed and analyzed experiments for figures 1 and S8A, S8C and S8E-G. C.C.-D. performed and analyzed assays in figure S15A. J.D. performed and analyzed experiments in figure S13A. C.E. analyzed experiments in figure S13A. G.C. contributed to experiments in figure S15A. F.H. performed and analyzed *in vitro* run-off transcription experiments in figure 4A. P.W. performed DNA fiber experiments and pulse-labeling assays in figures 7 and S3. M.C.C. conceived and supervised the project and wrote the manuscript.

### 8.3 Acknowledgment

First, I would like to thank Prof. Dr. M. Cristina Cardoso for the supervision of my PhD thesis during the last five and a half years. I am grateful for your constant support and your contagious (🤗🤗) and passionate dedication to science. Thanks that your door was always open for discussions and for your helpful advice. Thanks for taking care of me and always pushing me forward. I very much appreciated that you let me work independently and were, yet, always there to guide me. Being part of your scientific family was and still is an amazing experience and I enjoy it very much!

I would also like to thank Prof. Dr. Ulrike A. Nuber for agreeing to be the second supervisor of this thesis.

Many thanks to Florian, Maruthi, Patrick and Paulina for their fast and conscientious proof-reading of this thesis. Your comments and suggestions improved it a lot!

Alex, Anne, Anne Lu, Annika, Annina, Beata, Bianca, Diana, Eduardo, Florian, Francesco, Hector, Henry, Hui, Katalina, Kathrin, Katrin, Ksenia, Manu, Maria, Marina, Marius, Maruthi, Nicola, Patrick, Paulina, Peng, Stefania, Steffi, Stephan, Sunil and Vadim, thanks for creating a great working atmosphere over all those years; it was a pleasure. Special thanks to Anne Lu, Florian and Peng, the initial Met group, for introducing me to ongoing projects and helping me to integrate to the lab.

I am deeply grateful to Manu who became a close friend over the last years. Thanks for testing so many antibodies, discussing scientific stuff and, of course, for your famous cooking!! Thanks for listening to daily problems while struggling with your own ones and for building me up in difficult times, especially during the last year 🤔. And, of course, many, many thanks for always finding the time to cook, thereby preventing me from starving! 😊 Thanks for all the lovely (lunch) breaks! The Prosecco and cheese industries owe us a lot 🤗. You kept me motivated and always believed in a happy-end.

Liebe Anne, where should I start? Thanks for keeping the lab running, for always dealing with unexpected cell orders and for finding solutions to the strangest requests in the lab. But most of all, thanks for your constant support and your open ear. Thanks for always having time for a small chat and all your maternal advice - you are the best! 🤗

An den alten, weisen Mann: Alex, thanks for the countless times you have solved microscope, computer or other problems. Thanks for all your help during the last years and for sharing your infinite knowledge.

Diana and Katrin, thanks for last minute plasmid preps and orders and for listening to small and bigger problems.

Thanks to the lunchclub for many scientific discussions and a lot of very funny non-scientific conversations. Many thanks also to Annina for being such a great bench partner and to Beata and Marina for lunch breaks and a lot more.

Particular thanks go to the Lieblingsbiologen: Anne Lu, Kathrin, Florian and Paddy, thanks for scientific as well as non-scientific discussions, funny lunch sessions, botanical garden walks, a great office working atmosphere (👉) and, most importantly, great friendship. In the beginning you were colleagues, but you became good friends very quickly. Thanks for all your input and support also from the distance.

Finally I want to thank my parents and my sister for making this dissertation possible and great support during the last years. Gina, Rom, Chiara, dir hutt warscheinlech (ok, warscheinlech zimmlech sécher) méi oft drun gegleeft, dass ech hei alles géing hin kréien wei ech selwer (obwuel der näischt dovunner verstanenen hutt)! Dofir an fir nach vill méi woufir hei net méi genuch Platz ass, e groussen MERCI!!!! 🤗❤️



#### **8.4 Declaration - Ehrenwörtliche Erklärung**

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit entsprechend den Regeln guter wissenschaftlicher Praxis selbstständig und ohne unzulässige Hilfe Dritter angefertigt habe.

Sämtliche aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie sämtliche von Anderen direkt oder indirekt übernommenen Daten, Techniken und Materialien sind als solche kenntlich gemacht. Die Arbeit wurde bisher bei keiner anderen Hochschule zu Prüfungszwecken eingereicht.

Darmstadt, den 5. Juli 2021

Cathia Pascale Fabienne Rausch



## 8.5 Curriculum Vitæ

Personal Information	
Name:	Cathia Pascale Fabienne Rausch
Adress:	Schnittspahnstraße 10, 64287 Darmstadt, Germany
Date of birth:	27.11.1990
Place of birth:	Luxembourg
Doctoral studies	
01/2016 - 07/2021	<p><b>PhD thesis</b></p> <p>Technische Universität Darmstadt</p> <p>Department of Biology – Cell Biology and Epigenetics</p> <p>Prof. Dr. M. Cristina Cardoso</p> <p>"Cytosine base modifications, DNA metabolism and development"</p>
Academic education	
10/2015 - 12/2015	<p><b>Research assistant</b></p> <p>Ludwig-Maximilians-Universität Munich</p> <p>Faculty of Biology – Human Biology and Bioluminescence</p> <p>Prof. Dr. Heinrich Leonhardt</p>
03/2015 - 09/2015	<p><b>Master thesis</b></p> <p>Ludwig-Maximilians-Universität Munich</p> <p>Faculty of Biology – Human Biology and Bioluminescence</p> <p>Prof. Dr. Heinrich Leonhardt</p> <p>"Characterization of different TET1 deletion constructs"</p>
12/2014 - 02/2015	<p><b>Research internship</b></p> <p>Ludwig-Maximilians-Universität Munich</p> <p>Faculty of Biology – Human Biology and Bioluminescence</p> <p>Prof. Dr. Heinrich Leonhardt</p> <p>"Characterization of the monoclonal anti-MIN antibody"</p>
08/2014 - 11/2014	<p><b>Research internship</b></p> <p>Ludwig-Maximilians-Universität Munich</p> <p>Faculty of Biology – Human Biology and Bioluminescence</p> <p>Prof. Dr. Heinrich Leonhardt</p> <p>"Characterization of monoclonal anti-TauP antibodies"</p>

- 10/2013 - 09/2015      **Master studies**  
Ludwig-Maximilians-Universität Munich  
Faculty of Biology
- 09/2013 - 10/2013      **Research internship**  
Ludwig-Maximilians-Universität Munich  
Faculty of Biology – Microbiology  
Prof. Dr. Kirsten Jung  
"Role of the *cadC* localization and characterization of the unstructured cytoplasmic domain of the transcriptional activator CadC of *E. coli*"
- 04/2013 - 06/2013      **Bachelor thesis**  
Ludwig-Maximilians-Universität Munich  
Faculty of Biology – Microbiology  
Prof. Dr. Kirsten Jung  
"Mutagenesis and characterization of the cytoplasmic domain of the membrane integrated transcriptional activator CadC of *E. coli*"
- 10/2010 - 09/2013      **Bachelor studies**  
Ludwig-Maximilians-Universität Munich  
Faculty of Biology

---

**Publications**

---

- 2021      **Rausch, C.**, Zhang, P., Casas-Delucchi, C. S.\*, Daiß, J. L.\*, Engel, C., Coster, G., Hastert, F. D., Weber, P. and Cardoso, M. C.  
**Cytosine base modifications regulate DNA duplex stability and metabolism**  
\* equally contributing authors  
NAR, doi: 10.1093/nar/gkab509
- 2020      **Rausch, C.\***, Weber, P.\*, Prorok, P., Hörl, D., Maiser, A., Lehmkühl, A., Chagin, V. O., Casas-Delucchi, C. S., Leonhardt, H. and Cardoso, M. C.  
**Developmental differences in genome replication program and origin activation**  
\* equally contributing authors  
NAR 48(22), doi: 10.1093/nar/gkaa1124
- 2019      **Rausch C.**, Hastert, F. D. and Cardoso, M. C.  
**DNA modification readers, writers and their interplay**  
J. Mol. Biol. 432, doi: 10.1016/j.jmb.2019.12.018

Weber, P.\*, **Rausch, C.\***, Scholl, A. and Cardoso, M. C.

**Repli-FISH (Fluorescence in Situ Hybridization): Application of 3D-(Immuno)-FISH for the study of DNA replication timing of genetic repeat elements**

\* equally contributing authors

OBM Genetics 3(31), doi: 10.21926/obm.genet.1901062

2018

Natale, F.\*, Scholl, A.\*, Rapp, A.\*, Yu, W., **Rausch, C.** and Cardoso, M. C.

**DNA replication and repair kinetics of Alu, LINE-1 and satellite III genomic repetitive elements**

\* equally contributing authors

Epigenetics & Chromatin 11(61), doi: 10.1186/s13072-018-0226-9

2017

Ludwig, A. K.\*, Zhang, P.\*, Hastert, F. D., Meyer, S., **Rausch, C.**, Herce, H. D., Müller, U., Lehmkuhl, A., Hellmann, I., Trummer, C., Storm, C., Leonhardt, H. and Cardoso, M. C.

**Binding of MBD proteins to DNA blocks Tet1 function thereby modulating transcriptional noise**

\* equally contributing authors

NAR 45(5), doi: 10.1093/nar/gkw1197

Zhang, P.\*, Ludwig, A. K.\*, Hastert, F. D., **Rausch, C.**, Lehmkuhl, A., Hellmann, I., Smets, M., Leonhardt, H. and Cardoso, M. C.

**L1 retrotransposition is activated by Ten-eleven-translocation protein 1 and repressed by methyl-CpG binding proteins**

\* equally contributing authors

Nucleus 8(5), doi: 10.1080/19491034.2017.1330238

Zhang, P., **Rausch, C.**, Hastert, F. D., Boneva, B., Filatova, A., Patil, S. J., Nuber, U. A., Gao, Y., Zhao, X. and Cardoso, M. C.

**Methyl-CpG binding domain protein 1 regulates localization and activity of Tet1 in a CXXC3 domain-dependent manner**

NAR (45) 12, doi: 10.1093/nar/gkx281