Relevance of heterochromatin topology in genome silencing during (retro) differentiation

TECHNISCHE UNIVERSITÄT DARMSTADT

Vom Fachbereich Biologie der Technischen Universität Darmstadt
zur
Erlangung des akademischen Grades
eines Doctor rerum naturalium
genehmigte Dissertation von

Dipl. Biol. Katharina Laurence Jost
aus Frankfurt am Main

Berichterstatter (1. Referent): Prof. Dr. M. Cristina Cardoso
Mitberichterstatter (2. Referent): Prof. Dr. Gerhard Thiel

Tag der Einreichung: 21.10.2011
Tag der mündlichen Prüfung: 19.12.2011
Darmstadt 2013
D17

Index

Table of figures

Summary

Cellular differentiation depends on the expression of the right genes at the right time. As each cell contains the same DNA, transcriptional and epigenetic factors have to maintain tight control over gene expression. Even a small divergence from the correct differentiation program can lead to severe defects and even death.

The protein–DNA complexes that constitute chromatin provide a global level of gene regulation. Many epigenetic factors directly modify DNA or histones with methylation or acetylation marks that influence transcription by changing chromatin compaction and therefore accessibility of the DNA. Especially heterochromatin which is highly compacted and is known to mostly contain transcriptionally inactive genes could exert silencing functions on neighboring genes even on different chromosomes. In mouse cells, heterochromatin can be found in two distinct forms; as clusters within the nucleus (chromocenter) and as a thin layer at the nuclear periphery. Previous studies analyzing the effect of heterochromatin on gene expression of neighboring genes presented a mixed picture. Some studies underline the repressive effect of heterochromatin proximity on genes whereas others did not observe this effect. The conflicting results might be due to experimental difficulties of obtaining meaningful distance measurements in cells. A point often overlooked is chromatin reorganization. Especially chromocenters undergo large-scale reorganization during differentiation changing from multiple small to few larger chromocenters. In addition, nuclear size and shape changes also influences distance measurements.

In the presented thesis, I developed a method to eliminate morphological bias in distance measurements that improves our ability to study the influence of heterochromatin and its reorganization on gene expression. Using this bias-free method, I compared gene to heterochromatin distances between pluripotency/differentiation states and cell types. To elucidate the role of heterochromatin reorganization in the absence of other regulatory programs, I uncoupled heterochromatin reorganization from the differentiation process. This was achieved by ectopic expression of methyl CpG binding protein 2 (MeCP2), which induces a similar level of chromatin reorganization, as observed during differentiation.

Combining the newly developed tools and known methods, I could establish that gene position within a nucleus is highly influenced by the genetic context of the individual gene. Gene repositioning in relation to chromocenters is not coupled to gene expression changes during differentiation. In the case of distances to the periphery a negative correlation could be observed, suggesting that the periphery may act as an activating rather than a repressing compartment contrary to the previously suggested hypothesis. I was able to show a positive correlation between gene expression and chromocenter proximity when heterochromatin reorganization was induced by expression of MeCP2. This underlines the effect of chromatin reorganization on gene expression which is concealed by other processes during differentiation. Also in these experiments, the periphery did not exert any repressive effect on gene expression but rather showed an activating effect. In conclusion, I could demonstrate that constitutive heterochromatin can have a silencing effect on gene expression whereas facultative heterochromatin acts as an activator of gene expression.

Zusammenfassung

Die Zelldifferenzierung ist unmittelbar von exakter räumlicher und temporärer Genexpression abhängig. Während des Differenzierungsprogrammes muss die Genexpression der mit identischer DNS ausgestatteten Zellen, durch epigenetische und Transkriptionsfaktoren präzise kontrollierten werden, um selbst kleinste Abweichungen, die zu schweren Schäden und Tod führen können, zu vermeiden.

Chromatin, ein Protein-DNA-Komplex, ist für eine globale Form der Genregulation verantwortlich. DNA und Histone werden durch viele epigenetische Faktoren direkt modifiziert, unter anderem durch das Anbringen von Methylierungs- und Azetylierungsmarkierungen, welche unmittelbar Einfluss auf die Chromatin-Kompaktierung und damit dessen Zugänglichkeit nehmen. Ein negativer Effekt durch stark kompaktiertes Heterochromatin, welches hauptsächlich inaktive Gene enthält, könnte somit auch auf die Expression benachbarter Gene ausgeübt werden, selbst wenn diese sich auf unterschiedlichen Chromosomen befinden. Heterochromatin in Mauszellen existiert in zwei charakteristischen Variationen, ein Teil des Heterochromatins formt Cluster innerhalb des Nucleus (Chromocenter) während der übrige Teil an die nukleäre Peripherie angelagert ist. Bisherige Studien zum Einfluss von Heterochromatin auf die Genexpression, kommen zu widersprüchlichen Ergebnissen: während einige Studien den repressiven Effekt herausstellen, können andere diesen Effekt nicht nachweisen. Zurückzuführen sind diese widersprüchlichen Ergebnisse wahrscheinlich auf methodische Unzulänglichkeiten der Distanzmessung, nicht zuletzt verursacht durch die Vernachlässigung ausgeprägter Chromatin Umstrukturierungen, von mehreren kleinen zu wenigen großen Clustern, während der Differenzierung, sowie Änderungen in Form und Größe des Zellkerns.

Die im Rahmen der hier vorgestellten Arbeit von uns entwickelte Methode zur Distanzmessung kompensiert den Einfluss der Zellmorphologie und ermöglicht einen objektiven Vergleich verschiedener Zelltypen und Stadien. Diese neu etablierte Methode konnte ich erfolgreich zum Vergleich der Gen/Heterochromatin Distanzen zwischen verschiedenen Pluripotenz- und Differenzierungsstadien verwenden. Zur Erforschung der differenzierungsunabhängigen Effekte einer heterochomatischen Umorganisation auf die Genexpression, muss diese unabhängig vom Differenzierungsprozess evaluiert werden. Eine induzierte ektopische Expression des Methyl CpG binding protein (MeCP2) wird genutzt, um eine heterochromatische Umorganisation zu initiieren, welche die Reorganisation im Verlauf der Differenzierung imitiert.

Die Kombination neuentwickelter und bereits etablierter Methoden ermöglichte mir den Nachweis einer starken Korrelation zwischen der Geneposition relativ zum Heterochromatin und seinem genetischen Umfeld. Genrepositionierung in Relation zu Chromocentern ist hingegen nicht an die Änderungen der Genexpression während der Differenzierung gekoppelt. Bereits veröffentlichte Hypothesen konnte ich im Falle einer Repositionierung zur Peripherie durch die beobachtete negative Korrelation zur Genexpression wiederlegen. Die Peripherie würde in diesem Fall vor allem einen aktivierenden Einfluss ausüben. MeCP2 Expression induzierte Chromatin-Reorganisation hingegen zeigt eine positive Korrelation zwischen Genexpression und Distanzänderung zu Chromocentern. Dies unterstreicht den Effekt der Chromatin-Reorganisation auf die Genexpression, welcher ansonsten von anderen Prozessen während der Differenzierung überdeckt wird. Der repressive Einfluss der Peripherie auf die Genexpression konnte auch in diesem Versuch negiert werden. Zusammenfassend ist in einigen Fällen ein repressiver Effekt durch konstitutives Heterochromatin (Chromocenter) auf die Genexpression nachweisbar, während fakultatives Heterochromatin eine aktivierende Funktion besitzt.

General introduction

Obtaining the genome sequence of different organisms was one of the main goals to understand genome organization in living organisms. Unfortunately the sequence itself proved not to answer all questions. It soon became clear that there is more information on DNA then its sequence. This field of research was dubbed epigenetics (Waddington, 1942); the information above the DNA. Epigenetics represents heritable changes in gene function in the absence of changes in DNA sequence.

In the following Introduction I would like to give a brief overview on chromatin —the combination of DNA and proteins— and its influence on gene expression during cellular differentiation including the role of MeCP2 as a transcriptional regulator associated with chromatin. As well as higher-order levels of chromatin organization which can be brought about by proteins and influence genome activity and stability.

Historical definition of chromatin

In 1879, Walter Flemming coined the term chromatin, defining it as the portion of the nucleus that could be stained by aniline dyes. Not long after this discovery, C. Rabl was the first to describe a non-random distribution of chromosomes within the nucleus. He observed that chromosome position from mitosis was maintained in interphase (Rabl configuration (Rabl, 1885)). P. Baccarini described further subnuclear structures in 1908. He noticed dark stained bodies in the interphase nucleus of plants, which could be distinguished from nucleoli, and called them “chromocentri” (Baccarini, 1908) Figure 1. Not much was known about those structures, now termed chromocenters, and they where postulated to be condensed parts of chromosomes by Rosenberg (1908). Emil Heitz finally introduced the term heterochromatin in 1928 on a morphological basis. He used this term to describe chromosomal regions that remained more compact and therefore intensely stained during the cell cycle. Chromosome regions, which always stained very light, and disappeared during telophase were named euchromatin (Heitz, 1928). He also observed a spatial relation between heterochromatin and nucleoli Figure 1. Since chromocenters always stained strongly it was rightly assumed that they consist of heterochromatin.

Figure 1. Original Chromocenter drawings. A Original drawings of P. Baccarini showing chromocenters in root cells of *Cynomorium coccineum* L. **B** Original drawing of E. Heitz showing heterochromatin, also called chromocenters, accumulated around nucleoli and euchromatin in *Pellia epiphylla*.

Despite more than century of efforts to define the biological roles of these structures, many of the basic mechanisms that govern the formation of different chromatin types remain poorly understood. Especially, how the compaction of heterochromatin is achieved is largely unknown.

Composition and structure of chromatin

The key to understanding all processes in chromatin biology lies in R. Kornberg's finding that chromatin does not only contain DNA but also proteins (Kornberg, 1974). R. Kornberg was also the first to use the term nucleosome to describe the basic building block of chromatin. The nucleosomes consist of the core histone proteins H2A, H2B, H3 and H4, which are each present twice and form an octamer and were first described by (Bentley *et al.*, 1984). The DNA double helix is wrapped around the octamer 1.7 times (146 bp), forming the nucleosome. This so-called "beads on a string" structure (Olins & Olins, 1974) is agreed to be the first level of chromatin structure and has a diameter of 11 nm (Luger *et al.*, 1997). The next level of compaction is achieved by linker histones H1, which strongly interact with DNA, leading to the so-called 30 nm fiber (Robinson *et al.*, 2006). The structure of the 30 nm fiber however is controversial, being described as either zigzag or solenoid conformations (reviewed by (Woodcock & Ghosh, 2010)). Many higher order structures besides the 30 nm fiber have been discussed but none have been verified on a molecular level. The highest visible and undisputed compaction level of chromatin is the mitotic chromosome.

Histone and DNA modifications

Even though the structure of higher order chromatin is still not resolved many factors are known to influence chromatin compaction. As already mentioned, chromatin is distinguished into euchromatin and heterochromatin (Figure 2). A negative correlation of transcriptional activity and compaction could be observed i.e. higher compaction leading to less transcriptional activity. To distinguish chromatin into two groups might be oversimplified since recent work in *Drosophila* and human has proposed 5 distinct groups of chromatin based on localization of many chromatin proteins (Filion *et al.*, 2010; Ernst & Kellis, 2010). We will however concentrate on the generally accepted model of hetero- and euchromatin and its modifications. Heterochromatin itself can be separated into two different types, facultative and constitutive heterochromatin. Facultative heterochromatin contains genes, which are differently expressed during development and are not permanently silenced. By contrast, constitutive heterochromatin

contains permanently silenced genes in defined chromosomal regions such as the telomeres and centromeres including adjacent regions.

Distinct histone and DNA modifications and their binding proteins, which influence their compaction level, define each chromatin type.

Histone modifications

In histones, the N-terminal tail is the major site for posttranslational modifications that control the compaction and activation level of chromatin. Common modifications are acetylation of lysines, methylation of lysines and arginines, phosphorylation of serines, threonines and tyrosines as well as ADP ribosylation, ubiquitylation and sumoylation. Acetylation and methylation are the two best-studied marks relative to chromatin compaction.

Acetylation of histones is catalyzed by histone acetyl transferases (HATs) that transfer an acetyl group from acetyl-CoA to the ε-amino group of lysine. The negative charge of lysines is neutralized through this acetylation which has been proposed to weaken the interactions between histones and DNA. This in return leads to a more open chromatin structure explaining why euchromatin is usually associated with acetylated histones. Although acetylation is most common at the N-terminal tail of histones (histone 3 Lysine 9 (H3K9), H3K14, H4K8 and H4K12, among others), it has been shown that lysines in the core region of linker histone 5 (H5K56) are also acetylated in humans (Tjeertes *et al.*, 2009). The reverse reaction is carried out by histone deacetylases (HDACs) that re-establish the positive charge of lysine, leading to tighter packing of chromatin as in heterochromatin. HDACs are therefore crucial for the establishment of facultative heterochromatin during differentiation (Dovey *et al.*, 2010).

The main targets of histone methylation are the side chains of arginines and lysines. In contrast to acetylation, methylation does not change the charge of a histone. Another difference to acetylation is that lysines can be mono-, di- or tri-methylated (Paik & Kim, 1971) and arginines mono-and di-methylated. As in acetylation, usually the histone tail is methylated (H3K9, H3K4, H4K20 and more). An exception is the methylation of H3K79 within the histone core. Methylation is carried out by lysine or arginines methyltransferases. Methylation is not *per se* defining the conformation of chromatin. H3K9 trimethylation (H3K9me3) is a mark of heterochromatin whereas H3K4 mono methylation is a sign of an open euchromatic conformation.

Facultative and constitutive heterochromatin can be distinguished by their histone marks. Facultative heterochromatin is enriched in H3K27 trimethylation (H3K27m3) as can be seen in the inactive X-chromosome (Trojer & Reinberg, 2007). By contrast, constitutive heterochromatin contains higher levels of H3K9 trimethylation (H3K9m3) (Trojer & Reinberg, 2007). Since a growing number of histone modifications influencing chromatin structure have been identified, a "histone code" controlling epigenetics has been proposed (Strahl & Allis, 2000). The histone code hypothesis states that gene transcription regulation is encoded in histone modifications

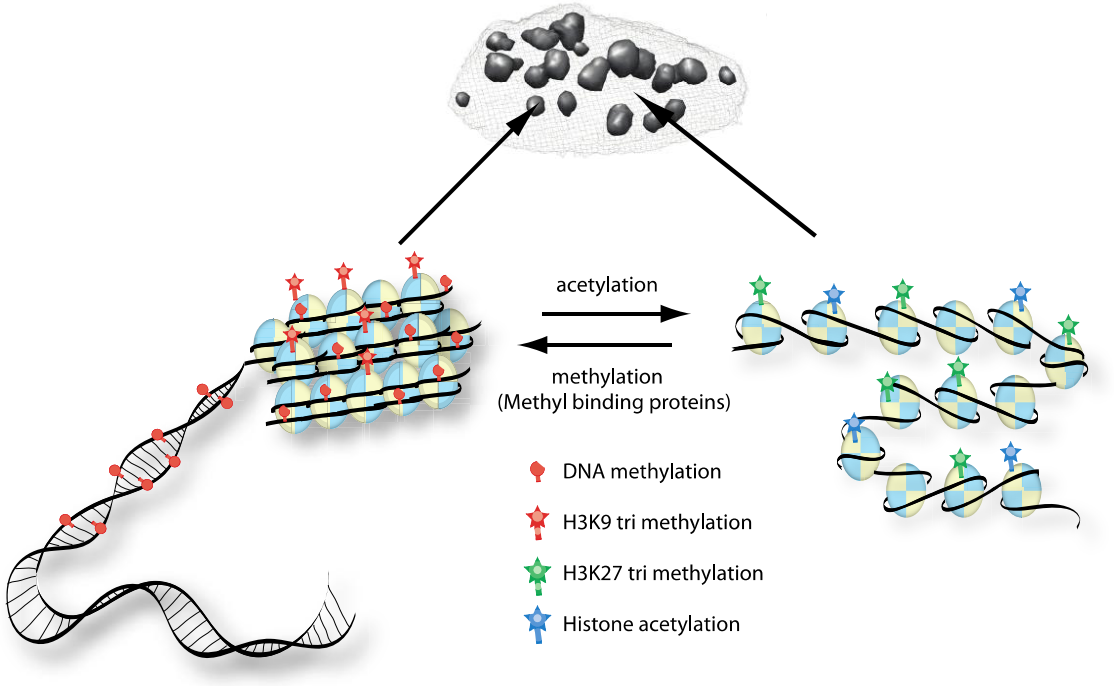


Figure 2. Chromatin states and associated modifications. Hetero- and Euchromatin can be differentiated by their compaction level and associated Histone modifications.

similar to the tripled code of bases in the linear DNA structure. Potential readers of this histone code are proteins as heterochromatin protein 1, polycomb group proteins and many others. The bromodomain was the first to be shown to selectively bin to lysines in histone tails whereas the chromodomain seem to target methylation marks (Jenuwein & Allis, 2001).

DNA modifications

In mammals, the most common chemical modification of the DNA itself is the methylation of cytosins. In the mammalian genome, 80% of all CpG di-nucleotides are methylated on the C5 position of the cytosine. Statistically, a CpG nucleotides should occur every 16 bp (considering the GC content, down to 40 bp), but studies have revealed that they are only found every 100 bp (Swartz *et al.*, 1962). This distribution is disrupted by so called CpG-islands which contain the expected frequency of CpGs at every 25 bp and stretch for around 200-1000 bp (Gardiner-Garden & Frommer, 1987). CpGs in CpG-islands are not methylated and coincide with active gene promoters. Additionally, methylated CpGs are found in heterochromatin (Razin & Cedar, 1977) leading to the conclusion that methylation is a silencing defining mark (Bird & Wolffe, 1999). Importantly, this post replicative modification pattern is maintained throughout

DNA replication, making methylation one of the key epigenetic regulators controlling the inheritance of gene expression.

Methylation marks are maintained by the DNA methyltransferase enzymes (Dnmts) Dnmt1, Dnmt3a and Dnmt3B. Unmethylated CpGs are methylated by the *de novo* methyltransferases Dnmt3A and Dnmt3b, establishing methylation patterns during development. The reversal of developmental DNA methylation patterns is crucial for reprogramming cells (Chin *et al.*, 2009). During replication, the newly synthesized DNA strand is unmethylated resulting in a hemimethylated DNA which is recognized by Dnmt1, the so-called maintenance methyltransferase (Leonhardt *et al.*, 1992), leading to the maintenance of the heritable methylation pattern.

More DNA modification, as 5-hydroxymethylcytosine (5hmC), which has been long known in bacteriophages (Wyatt & Cohen, 1952), was recently discovered in human and murine brain (Kriaucionis & Heintz, 2009; Tahiliani *et al.*, 2009). Additionally the modifications fro,5hmC to formylcytosine as well as carboxylcytosine have recently been reported (Ito *et al.*, 2011; Pfaffeneder *et al.*, 2011). The function and mechanisms of those modifications is the subject of intense research.

How methylation controls gene expression is still an open question and direct as well as indirect mechanisms are discussed. A model of direct control is the modification of transcription factor binding sites by methylation, hindering binding of the factor and therefore activation of the gene (Watt & Molloy, 1988). One model of indirect control is that methyl-CpG binding proteins (MBDs) bind to methylated cytosines and sterically prevent any binding of transcription factors (Leonhardt & Cardoso, 2000). It has also been suggested that MBDs could recruit transcriptional regulators (e.g. HDACs) inducing conformational chromatin changes leading to repression (Nan *et al.*, 1998; Zhang *et al.*, 1999). The most studied MBD involved in this process is methyl-cytosine binding protein 2 (MeCP2) and will be discussed in more detail in subsequent sections.

Chromocenters

After mitotic chromosomes, chromocenters represent the highest visible (by DNA staining) level of DNA compaction. It contains the so called satellite DNA repeats which can be distinguished into two types: the major satellite repeats (6 megabases of 234 bp units) and minor satellite repeats (~600 kb of 120 bp units) (Guenatri *et al.*, 2004), the first being the main DNA component in chromocenters. Minor satellite repeats are found directly at the centromeres

whereas major satellite repeats are flanking this region. Due to its position in the chromosome major satellite DNA is also referred to as pericentric heterochromatin. In Interphase of mouse cells these regions cluster together to form chromocenters (between 2-6 chromosomes per chromocenter). Since pericentric heterochromatin does not contain many active genes these heterochromatic regions, were referred to as “junk DNA” with no function (Ohno, 1972). However more recent studies support the idea of functional roles of heterochromatin by the detection of noncoding RNA arising from theses regions which influence gene regulation (Zaratiegui *et al.*, 2007).

Heterochromatin and transcriptional regulation

One of the most important epigenetic roles of heterochromatin was recognized very early on. In 1930, Muller (Muller, 1930) discovered that *Drosophila* flies treated with x-rays develop random color patterns of white and brown patches in the eyes. He could show that by random mutation the *white* gene locus responsible for white color was translocated adjacent to heterochromatic regions and thereby silenced. This effect was named position effect variegation (PEV). Further studies (Demerec & Slizynska, 1937) broadened the knowledge about PEV showing that genes in direct heterochromatic neighborhood were silenced first before more distal genes. These experiments showed that usually active genes get silenced just by being in the vicinity of heterochromatin which led to the development of the concept of heterochromatin spreading. Similar observations could be made after the relocation to telomeric regions - referred to as telomeric position effect variegation (TPEV)- and was observed in different species (Gehring *et al.*, 1984; Horn & Cross, 1995; Gottschling *et al.*, 1990). The PEV is based on *cis* interactions i.e. genes affected by heterochromatin on the same chromosome in a linear way. Whether heterochromatin can also influence gene expression in *trans* has been investigated by many groups (Brown *et al.*, 1997; Delaire *et al.*, 2004; Sabbattini *et al.*, 2001; Meaburn & Misteli, 2008; Szczerbal *et al.*, 2009; Takizawa *et al.*, 2008), but remains controversial.

As heterochromatin and telomeres are often associated with the nuclear lamina (Mathog *et al.*, 1984; Rae & Franke, 1972), it has also been suggested that these interactions could contribute to gene silencing, but different studies have failed to produce unambiguous results. A more detailed summary of the effects of chromocenters and the nuclear lamina on gene expression is made in the next two sections.

Chromocenters as silencing compartment

In some species e.g. mice and plants (as described in the historical section) heterochromatin is clustered into chromocenters during interphase. Chromocenters contain so called pericentric heterochromatin which resides close to the centromeres and is enriched in satellite repeat sequences of 234 base pairs. Chromocenters are often—but not exclusively—found associated with nucleoli and the nuclear lamina. It has been suggested that chromocenters as well as the heterochromatin at the lamina might be involved in gene silencing in a manner resembling the PEV but silencing genes that are not on the same chromosome. Initial studies in hematopoietic cells showed a correlation between gene activity and gene-chromocenter association (Brown *et al.*, 1997). In this study, B-lymphocytes were observed during inactivation and a relocation of active genes away from chromocenters and Ikaros proteins was observed whereas silenced genes were colocalizing with chromocenters (Brown *et al.*, 1997). The number of genes studied were usually relatively limited (not more then 6) and concentrate on genes relevant in the hematopoietic system, therefore a general effect cannot be ensured. Other early studies mainly concentrated on the hematopoietic cell system but soon others organisms and cell types were examined as well (Table 1).

These experiments were performed using fixed cells, capturing only a brief moment, and chromatin mobility (> 2 µm in G1 time frame: around 4h (Walter *et al.*, 2003)) was not taken into account. Harmon and Sedat performed the first study that took into consideration the role of chromatin mobility on *Drosophila* eye disks comparing distances from genes to heterochromatin in cells with active versus inactive gene loci (Harmon & Sedat, 2005). They could show a clear relation of gene silencing and association to heterochromatin of the same or different chromosome.

Although these and other studies established a correlation between gene position to pericentric heterochromatin and expression, this does not always seems to be true. Experiments using more sophisticated methods to look at more genes showed no clear correlation of expression and distance to chromocenters. A key player in determining the fate of genes that are in close proximity to chromocenters seems to be the Ikaros protein, as genes lacking the Ikaros binding site are still relocated to pericentric heterochromatin but not silenced (Sabbattini *et al.*, 2001). In addition, it is not clear if the observed relocation is the cause or the consequence of silencing/activation.

A challenging experimental problem posed by the different data sets is the variability of nuclear morphology. Different cells show a great variety of shapes and sizes that is often altered upon differentiation. This challenge is acknowledged by many researchers in the field stating that observations might be biased by shape and size (Meaburn & Misteli, 2008). When looking for correlations between genes and pericentric heterochromatin not only the size of the nucleus plays an important role but also the distribution of chromocenters within it. The heterochromatic distribution can vary greatly between cells and needs to be taken into consideration (described in the following section: Cellular differentiation and heterochromatin remodeling). As no general method for accounting for differences in nuclear morphology has been developed so far, its influence on the analysis of gene distance data remains unclear.

Table 1 Overview of studies correlating gene expression and heterochromatin proximity

Subnuclear compartment	Cells	Number of Genes tested	System studied	Method	Outcome	Reference
Chromocenter	Mouse B-lymphocyte	6	Activation of B-lymphocytes	FISH gene colocalization to chromocenter	Correlation between position and transcription	(Brown *et al.*, 1997)
Chromocenter	Mouse T-cells	2	Comparison of CD4 and CD8 (non) expressing cells	FISH gene colocalization to chromocenter	Correlation between position and transcription	(Delaire *et al.*, 2004)

Chromocenter	Mouse B-lymphocytes	1	Comparing B-lymphocytes with integrated transgene within and out of pericentric heterochromatin	Integration of transgene into chromocenters	Transgene active even inside pericentric heterochromatin	(Sabbattini *et al.*, 2001)
Chromocenter	*Drosophila* eye disk	3	Comparing activated vs. silenced genes	FISH gene distance to heterochromatin	Relocation of genes away from chromocenters upon activation	(Harmon & Sedat, 2005)
Lamina/ Chromocenter	Mouse neuronal precursor cells	1	Astrocyte differentiation	FISH gene colocalization to structure Radial distance to lamina	No relation to chromocenters for active or inactive loci; movement away from lamina upon activation	(Takizawa *et al.*, 2008)
Lamina	Pig Mesenchymal stem cells	9	Adipogenesis	FISH and 3D distance from signal to nuclei center as ratio of the nuclear radius	Correlation between position and transcription	(Szczerbal *et al.*, 2009)
Lamina	Human Mammary epithelial cells	11	Tumorigenesis	FISH and radial positioning from signal to lamina by template ellipse	For 7 out of 11 genes correlation between position and transcription	(Meaburn & Misteli, 2008)
Lamina	Myogenic mouse cells	2	Comparison of myogenic vs. non myogenic cell lines	FISH gene colocalization with lamina	Movement of locus away from lamina upon activation	(Lee *et al.*, 2006)
Lamina	Human HT1080 cells	Whole chromosomal region	Comparing expression of genes tethered to the lamina and non tethered	Tethering of gene to lamina	Some genes remain active some get repressed	(Finlan *et al.*, 2008)
Lamina	Mouse fibroblasts	Chromosomal region	Comparing expression of genes tethered to the lamina and non tethered	Tethering of gene to lamina	Some genes remain active some get repressed	(Reddy *et al.*, 2008)

The nuclear lamina as silencing compartment

Since heterochromatin does not only accumulate in form of chromocenters but also at the periphery of mammalian nuclei, many recent studies have concentrated on the nuclear periphery as a silencing compartment. This is of special interest, as human cells do not exhibit clear heterochromatic clusters comparable to those found in mouse cells.

The first indication of the periphery being a transcriptional regulator was found in yeast. In yeast, telomeres are enriched at the nuclear periphery leading to an increased concentration of silent information regulator (Sir) proteins which cause silencing of genes (Maillet *et al.*, 1996). Sir2 deacetylates histone tails recruiting other Sir proteins leading to silencing (reviewed in Gasser & Cockell, 2001). Silencing at the periphery through tethering of genes could also be observed in mouse fibroblasts (Reddy *et al.*, 2008). Gene repositioning in relation to the nuclear periphery has also been reported in astrocyte differentiation in mouse (Takizawa *et al.*, 2008), in tumorigenesis in human cells (Meaburn & Misteli, 2008) as well as in other examples (Table 1).

In general the parts of the genome mapping to lamina, called lamina-associated domain (LADs) are usually gene-poor, enriched in repressive chromatin modifications and exhibit low levels of transcription (Peric-Hupkes *et al.*, 2010). The genetic content of these LADs varies in different cell types which underscores the regulation potential of lamin association (Peric-Hupkes *et al.*, 2010).

As in the case of chromocenters, these studies have weaknesses regarding their analysis. They either score association with the lamina by simple colocalization (Lee *et al.*, 2006) or by using radial positioning within the cell (Meaburn & Misteli, 2008; Szczerbal *et al.*, 2009; Takizawa *et al.*, 2008). Scoring colocalization is problematic because it does not take gene mobility into account (as described above). Using a shell based analysis poses the additional problem of how the shells should be placed. In general, two possibilities of shell placement can be found in literature: equidistant (Szczerbal *et al.*, 2009) or equivolume shells (de Nooijer *et al.*, 2009). Using one or the other can change the data substantially.

In general the first mechanistic insight in developmental heterochromatic silencing came from studies in *Drosophila* describing the heterochromatin protein 1 (HP1) and the polycomb protein (Pc) (James & Elgin, 1986; Paro & Hogness, 1991). In recent years, many more polycomb proteins have been discovered and named polycomb group proteins (PcG). PcG are involved in the gene expression by altering chromatin modifications and nuclear organization of their target (Schuettengruber *et al.*, 2007). PcG homologs have also been described in higher eukaryotes and seem to be related to heterochromatin silencing. However, it could be shown that they distribute in foci throughout the nuclei, not always in close relationship with heterochromatin. However the existence of polycomb bodies is controversial, since no nuclear compartment could be correlated to the fluorescent CpG bodies in correlative light-electron microscopy (Smigova *et al.*, 2011).

Cellular differentiation and heterochromatin remodeling

As discussed in the previous chapter heterochromatin plays an important role in genome activity due to its silencing potential. Interestingly heterochromatin and especially chromocenters are remodeled during cell differentiation as different sets of genes get activated and repressed. During differentiation drastic changes in heterochromatin positioning and/or clustering can occur. An extreme example is the heterochromatin organization in rod photoreceptor cells of nocturnal animals. During differentiation all heterochromatin accumulates in the center of the nucleus which minimizes light scattering by the dense heterochromatin (Solovei *et al.*, 2009) illustrating the functional plasticity of heterochromatin.

Cellular (retro) differentiation

Embryonic stem cells (ES cells) are the transcriptionally most active cells, which is reflected in their transcription even of so-called junk DNA such as short/long interspersed elements (SINEs and LINEs) (Efroni *et al.*, 2008). In general, the heterochromatin distribution of ES cells is thought to reflect an open and transcriptionally active chromatin with less heterochromatin (Gaspar-Maia *et al.*, 2011). Differentiated cells however have a much more restrictive chromatin environment leading to silencing of many genes.

Many differences in heterochromatin organization are visible when comparing ES cells and differentiated cells. Electron microscopy (EM) studies showed that heterochromatin does not form clear aggregates in human ES cells and is distributed evenly throughout the nucleus (Park *et al.*, 2004). Other EM-based analyzes also showed that heterochromatin is present in mouse ES cells but only as a thin layer around nucleoli and the lamina (Baharvand & Matthaei, 2003). The same pattern was shown *in vivo* using the inner mass cells from blastocysts (Ahmed *et al.*, 2010). Contradicting studies studies using light microscopy and antibodies against heterochromatin markers (e.g. H3K9me3) could show that small heterochromatic foci can be found in ES cells. The number of foci significantly increased during differentiation in neuronal progenitor cells (Meshorer & Misteli, 2006). Clusters of heterochromatin seem to be a general feature of differentiation (Park *et al.*, 2004). In contrast to differentiated cells, centromeric clusters are localized to the interior of the nucleus in ES cells and not associated with the lamina (Bartova *et al.*, 2008; Butler *et al.*, 2009).

The different condensation states of chromatin in ES and differentiated cells are also reflected on the molecular level. Chromatin immunoprecipitation (ChIP) followed by microarray showed that H3K9me3 and H3K27me3, both marks for heterochromatin, increase from 4% genome coverage in ES cells to 12% and 16% in differentiated cells (Hawkins *et al.*, 2010). Consistent with these data, histone acetylation -a mark for active chromatin- is increased in undifferentiated ES cells (Krejci *et al.*, 2009). Despite this large-scale chromatin remodeling it has been shown that chromosomes still occupy similar 3D regions within the nucleus, called chromosome territories (CT), in ES cells (human and murine) and somatic cells (Bartova *et al.*, 2008; Wiblin *et al.*, 2005).

Recently, induced pluripotent stem cells (iPS) have come into focus as an alternative to ES cells. They are derived from somatic cells and are reprogrammed using a set of transcription factors (Takahashi & Yamanaka, 2006). An important question is how similar iPS cells are to ES cells and whether they have the same pluripotent capabilities (Okita *et al.*, 2007; Zhao *et al.*, 2009).

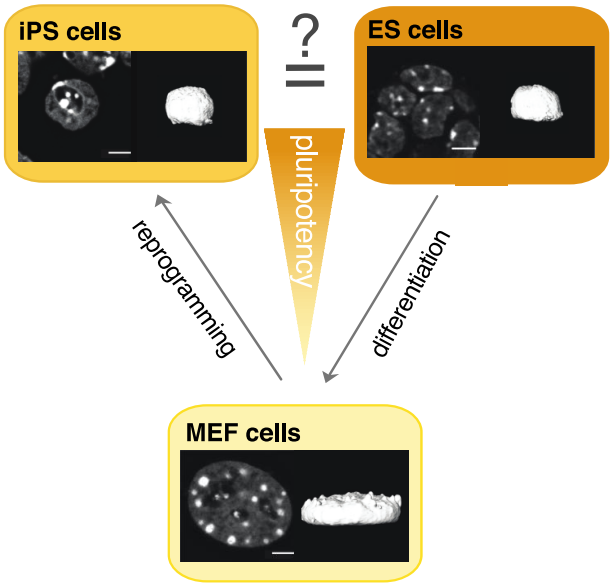


Figure 3. Morphological differences and similarities between ES, iPS and MEF cells. Schematic representation of the pluripotency state of MEF, iPS and ES cells. Confocal optical mid section of a DAPI stained representative MEF (lower), iPS (left) and ES (right) cell nucleus and corresponding 3D reconstructions highlighting their morphological variability. Scale bar 5 µm

From a morphological point of view, they have a similar shape and show the same distribution of heterochromatin around nucleoli and at the lamina as ES cells (Zeuschner *et al.*, 2010), which underscores the importance of decondensed chromatin for a pluripotent state. Evidence on the molecular level underlines the importance of chromatin rearrangement for pluripotency. HDAC inhibitors increase the efficiency of reprogramming (Huangfu *et al.*, 2008) as well as DNA demethylation, both factors lead to a general more open chromatin structure (Tsuji-Takayama *et al.*, 2004).

Myogenesis

As described in the previous section heterochromatin remodeling is a common feature of differentiation. Myogenesis is a well-established and studied model system for adult stem cell

differentiation. Myoblasts are adult stem cells, which differentiate into polynucleated myotubes (Figure 4) upon serum withdrawal (Yaffe & Saxel, 1977).

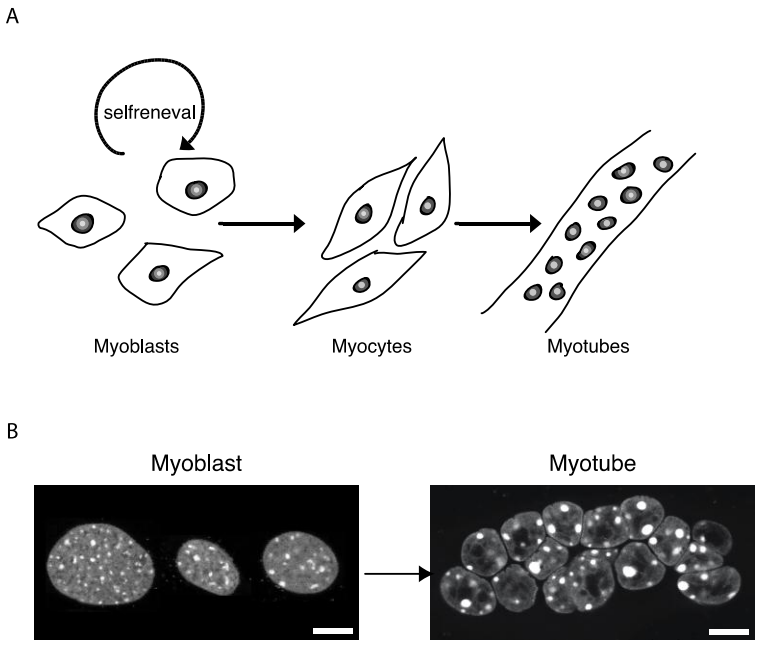


Figure 4. Schematic representation of myogenesis. A Myoblasts are adult muscle stem cells which differentiate into polynucleated myotubes upon serum withdrawal. Myocytes are an intermediated state during myogenesis. **B** DAPI stained nuclei of myoblast and polynucleated myotubes underlining morphological differences. Scale bar 5 µm

Myoblasts already display clear heterochromatic foci showing a more differentiated heterochromatic structure compared to pluripotent ES cells, but they undergo further large-scale chromatin remodeling during myogenesis (Brero *et al.*, 2005). The number of chromocenters drops from 20 in myoblasts to 11 in differentiated myotubes and the variability of chromocenter numbers is also reduced. This process seems to be continuous since myocytes, which are in a differentiation status between myoblasts and myotubes, display an intermediate number of 15 chromocenters (Brero *et al.*, 2005). This reduction in chromocenter numbers during myogenesis was also shown in *ex vivo* studies on satellite cells obtained from mice, demonstrating that the effect is not a cell culture artifact. *Ex vivo* myoblasts contained 18-27 chromocenters whereas myotube nuclei only showed 4-12 (Terranova *et al.*, 2005). This chromocenter reorganization is not limited to mouse myogenesis but has been shown in rat myoblasts (Chaly & Munro, 1996), mouse and human neurons as well as during ES cell differentiation, indicating that it is a general feature of terminal differentiation (Meshorer & Misteli, 2006). In C2C12 myoblasts, clustering of chromocenters could be blocked by trichostatin A (TSA), an inhibitor of HDACs, pointing to a role of histone deacetylation for clustering. Even after removal of TSA clustering could no longer be observed. However, TSA treated cells were still able to differentiate into myotubes but showed altered nuclear architecture (Terranova *et al.*, 2005). These observations still leave some questions open, an important one being if this clustering reflects also a compaction of DNA (reduction of chromocenter volume). A second question of interest is how

this clustering of chromocenters (reduction in chromocenter numbers) is induced and what functional consequence it has.

MeCP2 as modulator of genome expression

As described earlier, DNA methylation is a key regulator of epigenetic information and therefore gene expression. Gene silencing by methylation can be mediated through physically blocking transcription factors or other interacting proteins from binding DNA (Bell & Felsenfeld, 2000; Szabo *et al.*, 2000). Additionally, methylation marks can be recognized by transcriptional repressors (Nan *et al.*, 1997) and DNA binding proteins that can cause chromatin compaction (Brero *et al.*, 2005).

Methyl CpG binding protein 2 (MeCP2) was the second methyl CpG binding protein to be discovered and the first to be cloned. In interphase mouse nuclei, MeCP2 is prominently localized at heterochromatic foci and in metaphase chromosomes MeCP2 localizes strongly at pericentric heterochromatin, highly enriched in methylated major satellite DNA repeats.

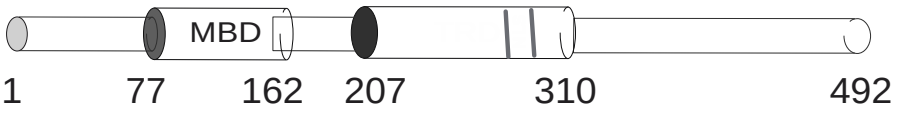


Figure 5. Schematic representation of MeCP2. A schematic representation of the rat MeCP2 protein and its functional domains is shown below. MBD: methyl CpG binding domain; TRD: transcriptional repression domain; NLS: nuclear localization signal.

MeCP2 (Figure 5) consists of a conserved methyl-CpG binding domain (MBD) that binds to 5-methyl cytosine with high affinity, and is the shared characteristic with the other MBD protein family members; a transcriptional repression domain (TRD), which interacts with histone deacetylases and the transcriptional corepressor SIN 3A (Jones *et al.*, 1998); a nuclear localization sequence; and a C-terminal part, which binds nucleosomes (Brero *et al.*, 2006; Georgel *et al.*, 2003).

MeCP2 seems to have important regulatory functions as demonstrated in the disease Rett syndrome which is caused by MeCP2 mutations. Rett syndrome (RTT) was first described in 1966 by Andreas Rett (Rett, 1966) and affects one in every 10,000-15,000 female births (Amir *et al.*, 1999; Hagberg, 1985; Hagberg *et al.*, 1983). Affected girls seem to develop normally until six to 18 months, subsequently they enter a developmental arrest, which is followed by strongly impaired motor skills, stereotypic hand movements, loss of speech, seizures, abnormal breathing, microcephaly, ataxia and other symptoms. Mutations within the *MECP2* gene located on chromosome Xq28 are found in approximately 80% of all classic RTT cases (Amir *et al.*, 1999; Amir & Zoghbi, 2000).

Since MeCP2 does not only have a methyl CpG binding site but also a transcriptional repression domain, it has a strong ability to act as an epigenetic regulator of gene expression in different modi.

MeCP2 as remodeler of heterochromatin

Georgel et al could demonstrate that MeCP2 is able to induce compaction of oligonucleosomes *in* vitro establishing MeCP2 remodeling function (Georgel *et al.*, 2003). Further studies to elucidate which proteins could regulate or initiate heterochromatin clustering were conducted by Brero *et al.* (Brero *et al.*, 2005). They tested three different proteins on their ability to cluster heterochromatin. Heterochromatin protein 1 (HP1) since it was one of the first proteins to be characterized to associate with heterochromatin, centromere protein B (CenpB) involved in centromere formation and MeCP2 as methyl CpG binding protein. The proteins HP1 and CenpB did not induce heterochromatin clustering upon transient expression in myoblasts, therefore excluding heterochromatic proteins as general clusterers. By contrast, fluorescently-tagged MeCP2 expression was able to induce large-scale chromatin compaction in a dose dependent manner.

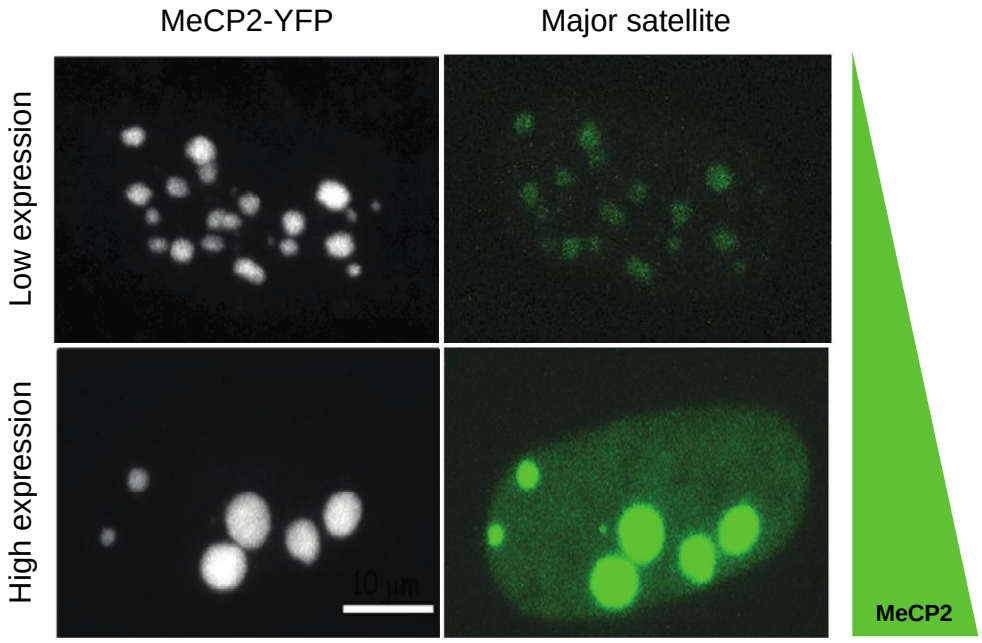


Figure 6. MeCP2-YFP overexpression induces clustering of pericentric heterochromatin. Pmi28 myoblasts expressing different levels of MeCP2-YFP and exhibiting extensive chromatin remodeling. Bar, 5 µm (Brero *et al.*, 2005)

If MeCP2 was only expressed at low levels, myoblasts exhibited the same chromocenter pattern as in untransfected cells (small and many) whereas cells with a higher expression level showed reduced number and larger chromocenters as in myotubes. Transfected cells were able to differentiate leading to myotube nuclei with even more reduced numbers of chromocenters (9.5). In further experiments, Brero *et al.* could show that the MBD domain of MeCP2 is necessary and sufficient for induction of chromatin clustering. Even though this suggests a critical role of MeCP2 in muscle development, MeCP2 knock-out mice and also patients suffering from RTT do not show defects of myogenesis and form differentiated tissues (Kriaucionis & Bird, 2003). This might be due to MeCP2 not being involved in the establishment of differentiation but in the maintenance of the transcriptional silencing. Another possibility could be a functional redundancy between MDB proteins as they all share the highly conserved

methyl binding domain of MeCP2. Brero *et al.* could further show that MBD2 can also induce clustering in myoblasts and is expressed in a differentiation dependent manner (Brero *et al.*, 2005), which hints at a functional redundancy between MBD proteins. However more recent studies showed a malfunction in cardiac muscle development upon MeCP2 over-expression (Alvarez-Saavedra *et al.*, 2010), providing evidence in favor of involvement of MeCP2 in development and leaving the discussion open.

MeCP2 as transcriptional modulator

Due to its transcriptional repressor domain and association with pericentric heterochromatin MeCP2, was until recently only recognized as a repressor of transcription. This hypothesis has been proven with many studies and underlines the importance of MeCP2. However, recent studies suggest that MeCP2 can act not only as a repressor but also under certain circumstances as an activator as well (Chahrour *et al.*, 2008)

MeCP2 as repressor

The first indication of MeCP2s ability to repress transcription *in vitro* was through studies with transcription extracts. In these extracts MeCP2 enriched from mouse liver was able to block transcription (Meehan *et al.*, 1992). MeCP2s ability to block transcription was confirmed by Nan *et al.* (Nan *et al.*, 1997) using purified MeCP2. In this study MeCP2 was extracted from rat brain by affinity chromatography to exclude false positives by impurities of other proteins. Addition of purified MeCP2 to an *in vitro* transcription reaction with methylated DNA was able to repress the reaction strongly (Nan *et al.*, 1997). This repression mechanism seems to be partially functional without DNA methylation.

In the light of MECP2 role in neurological disease, special interest has been given to its role as a repressor during neurodevelopment. Early studies though were not able to detect large-scale differences of gene expression in knock-out mouse or RTT patient brain (Colantuoni *et al.*, 2001; Tudor *et al.*, 2002). It became clear that especially in the brain various regions might be affected to a different degree and a whole brain analysis might not be accurate. One of the first genes found to be directly repressed by MeCP2 was the brain derived neurotrophic factor (BDNF). In neuronal cultures, MeCP2 binds to the BDNF III promoter to repress transcription (Martinowich *et al.*, 2003). Since then, more genes have been identified in brain to be directly repressed by MeCP2 including Dlx5 and Fxyd1 in the front cortex (Horike *et al.*, 2005; Jordan *et al.*, 2007) and also GABRB3 which are necessary for normal development and dendritic morphology (Samaco *et al.*, 2005). Interestingly not only MeCP2 mutation but also duplications of WT MeCP2 lead to severe mental retardations similar to RTT but affects only males in contrast to RTT syndrome (Ramocki *et al.*, 2010). The importance of correct MeCP2 expression could also been shown by over expression of MeCP2 in heart and skeletal muscle. Overexpression leads to a repression of Tbx5, a transcription factor that plays an important role in the development of the heart resulting in severe malformation of the heart and leading to embryonic lethality (Alvarez-Saavedra *et al.*, 2010).

MeCP2 as transcriptional activator

Early studies using an *in vitro* reporter construct could demonstrate enhanced transcriptional activity by the TRD of MeCP2 (Yu *et al.*, 2001). This positive effect was also observed using full length MeCP2 (Yu *et al.*, 2001). However, this observation was not seriously considered, as it did not match the popular view of MeCP2 as a repressor. As described above, first genome-wide studies of gene expression comparing wt and MeCP2 knock-outs did not show large differences (Tudor *et al.*, 2002). A study concentrating on the mouse hypothalamus found the astonishing result that 2184 out of 2582 regulated genes (85%) were up and not down regulated comparing *mecp2*-null males with wt (Chahrour *et al.*, 2008). At least for six selected genes, it could be shown that binding of MeCP2 to their promoters causes significant upregulation of their expression. Additionally, the study revealed an interaction of MeCP2 with the transcriptional activator CREB1 (Chahrour *et al.*, 2008). Up to now, not many studies have been conducted about MeCP2s role as an activator but considering the heterogeneous data, the true function of MeCP2 and its mechanism remains to be uncovered.

REFERENCES

Ahmed, K., Dehghani, H., Rugg-Gunn, P., Fussner, E., Rossant, J. and Bazett-Jones, D.P. (2010) Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. *PLoS ONE*, **5**, e10531.

Alvarez-Saavedra, M., Carrasco, L., Sura-Trueba, S., Demarchi Aiello, V., Walz, K., Neto, J.X. and Young, J.I. (2010) Elevated expression of MeCP2 in cardiac and skeletal tissues is detrimental for normal development. *Hum. Mol. Genet.*, **19**, 2177-2190.

Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U. and Zoghbi, H.Y. (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.*, **23**, 185-188.

Amir, R.E. and Zoghbi, H.Y. (2000) Rett syndrome: methyl-CpG-binding protein 2 mutations and phenotype-genotype correlations. *Am J Med Genet*, **97**, 147-152.

Baccarini, P. (1908) Sulle cinesi vegeative del "Cynomorium coccineum L.". *Nuovo Girornale botanico italiano Nuova serie*, **15**, 189-203.

Baharvand, H. and Matthaei, K.I. (2003) The ultrastructure of mouse embryonic stem cells. *Reprod Biomed Online*, **7**, 330-335.

Bartova, E., Galiova, G., Krejci, J., Harnicarova, A., Strasak, L. and Kozubek, S. (2008) Epigenome and chromatin structure in human embryonic stem cells undergoing differentiation. *Dev Dyn*, **237**, 3690-3702.

Bell, A.C. and Felsenfeld, G. (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*, **405**, 482-485.

Bentley, G.A., Lewit-Bentley, A., Finch, J.T., Podjarny, A.D. and Roth, M. (1984) Crystal structure of the nucleosome core particle at 16 Å resolution. *J. Mol. Biol.*, **176**, 55-75.

Bird, A.P. and Wolffe, A.P. (1999) Methylation-induced repression--belts, braces, and chromatin. *Cell*, **99**, 451-454.

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.

Brero, A., Leonhardt, H. and Cardoso, M.C. (2006) Replication and translation of epigenetic information. *Curr. Top. Microbiol. Immunol.*, **301**, 21-44.

Brown, K.E., Guest, S.S., Smale, S.T., Hahm, K., Merkenschlager, M. and Fisher, A.G. (1997) Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell*, **91**, 845-854.

Butler, J.T., Hall, L.L., Smith, K.P. and Lawrence, J.B. (2009) Changing nuclear landscape and unique PML structures during early epigenetic transitions of human embryonic stem cells. *J Cell Biochem*, **107**, 609-621.

Chahrour, M., Jung, S.Y., Shaw, C., Zhou, X., Wong, S.T., Qin, J. and Zoghbi, H.Y. (2008) MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science*, **320**, 1224-1229.

Chaly, N. and Munro, S.B. (1996) Centromeres reposition to the nuclear periphery during L6E9 myogenesis in vitro. *Exp Cell Res*, **223**, 274-278.

Chin, M.H., Mason, M.J., Xie, W., Volinia, S., Singer, M., Peterson, C., Ambartsumyan, G., Aimiuwu, O., Richter, L., Zhang, J. *et al.* (2009) Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell*, **5**, 111-123.

Colantuoni, C., Jeon, O.H., Hyder, K., Chenchik, A., Khimani, A.H., Narayanan, V., Hoffman, E.P., Kaufmann, W.E., Naidu, S. and Pevsner, J. (2001) Gene expression profiling in postmortem Rett Syndrome brain: differential gene expression and patient classification. *Neurobiol Dis*, **8**, 847-865.

de Nooijer, S., Wellink, J., Mulder, B. and Bisseling, T. (2009) Non-specific interactions are sufficient to explain the position of heterochromatic chromocenters and nucleoli in interphase nuclei. *Nucleic Acids Res*, **37**, 3558-3568.

Delaire, S., Huang, Y.H., Chan, S.W. and Robey, E.A. (2004) Dynamic repositioning of CD4 and CD8 genes during T cell development. *J Exp Med*, **200**, 1427-1435.

Demerec, M. and Slizynska, H. (1937) Mottled White 258-18 of Drosophila Melanogaster. *Genetics*, **22**, 641-649.

Dovey, O.M., Foster, C.T. and Cowley, S.M. (2010) Histone deacetylase 1 (HDAC1), but not HDAC2, controls embryonic stem cell differentiation. *Proc Natl Acad Sci U S A*, **107**, 8242-8247.

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, 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.

Ernst, J. and Kellis, M. (2010) Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol.*, **28**, 817-825.

Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J. *et al.* (2010) Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell*, **143**, 212-224.

Finlan, L.E., Sproul, D., Thomson, I., Boyle, S., Kerr, E., Perry, P., Ylstra, B., Chubb, J.R. and Bickmore, W.A. (2008) Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet*, **4**, e1000039.

Gardiner-Garden, M. and Frommer, M. (1987) CpG islands in vertebrate genomes. *J Mol Biol*, **196**, 261-282.

Gaspar-Maia, A., Alajem, A., Meshorer, E. and Ramalho-Santos, M. (2011) Open chromatin in pluripotency and reprogramming. *Nat Rev Mol Cell Biol*, **12**, 36-47.

Gasser, S.M. and Cockell, M.M. (2001) The molecular biology of the SIR proteins. *Gene*, **279**, 1-16.

Gehring, W.J., Klemenz, R., Weber, U. and Kloter, U. (1984) Functional analysis of the white gene of Drosophila by P-factor-mediated transformation. *EMBO J*, **3**, 2077-2085.

Georgel, P.T., Horowitz-Scherer, R.A., Adkins, N., Woodcock, C.L., Wade, P.A. and Hansen, J.C. (2003) Chromatin compaction by human MeCP2. Assembly of novel secondary chromatin structures in the absence of DNA methylation. *The Journal of biological chemistry*, **278**, 32181-32188.

Gottschling, D.E., Aparicio, O.M., Billington, B.L. and Zakian, V.A. (1990) Position effect at S. cerevisiae telomeres: reversible repression of Pol II transcription. *Cell*, **63**, 751-762.

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.

Hagberg, B. (1985) Rett's syndrome: prevalence and impact on progressive severe mental retardation in girls. *Acta Paediatr Scand*, **74**, 405-408.

Hagberg, B., Aicardi, J., Dias, K. and Ramos, O. (1983) A Progressive Syndrome of Autism, Dementia, Ataxia, and Loss of Purposeful Hand Use in Girls - Retts Syndrome - Report of 35 Cases. *Ann. Neurol.*, **14**, 471-479.

Harmon, B. and Sedat, J. (2005) Cell-by-cell dissection of gene expression and chromosomal interactions reveals consequences of nuclear reorganization. *PLoS Biol.*, **3**, e67.

Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S. *et al.* (2010) Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell*, **6**, 479-491.

Heitz, E. (1928) Das Heterochromatin der Moose. I. *Jahrbücher für wissenschaftlitsche Botanik*, **69**, 762-818.

Horike, S., Cai, S., Miyano, M., Cheng, J.F. and Kohwi-Shigematsu, T. (2005) Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat Genet*, **37**, 31-40.

Horn, D. and Cross, G.A. (1995) A developmentally regulated position effect at a telomeric locus in Trypanosoma brucei. *Cell*, **83**, 555-561.

Huangfu, D., Maehr, R., Guo, W., Eijkelenboom, A., Snitow, M., Chen, A.E. and Melton, D.A. (2008) Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nat Biotechnol*, **26**, 795-797.

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.

James, T.C. and Elgin, S.C. (1986) Identification of a nonhistone chromosomal protein associated with heterochromatin in Drosophila melanogaster and its gene. *Mol Cell Biol*, **6**, 3862-3872.

Jenuwein, T. and Allis, C.D. (2001) Translating the histone code. *Science*, **293**, 1074-1080.

Jones, P.L., Jan Veenstra, G.C., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J. and Wolffe, A.P. (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet*, **19**, 187-191.

Jordan, C., Li, H.H., Kwan, H.C. and Francke, U. (2007) Cerebellar gene expression profiles of mouse models for Rett syndrome reveal novel MeCP2 targets. *BMC Med Genet*, **8**, 36.

Kornberg, R.D. (1974) Chromatin structure: a repeating unit of histones and DNA. *Science*, **184**, 868-871.

Krejci, J., Uhlirova, R., Galiova, G., Kozubek, S., Smigova, J. and Bartova, E. (2009) Genome-wide reduction in H3K9 acetylation during human embryonic stem cell differentiation. *J Cell Physiol*, **219**, 677-687.

Kriaucionis, S. and Bird, A. (2003) DNA methylation and Rett syndrome. *Hum. Mol. Genet.*, **12 Spec No 2**, R221-227.

Kriaucionis, S. and Heintz, N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, **324**, 929-930.

Lee, H., Quinn, J.C., Prasanth, K.V., Swiss, V.A., Economides, K.D., Camacho, M.M., Spector, D.L. and Abate-Shen, C. (2006) PIAS1 confers DNA-binding specificity on the Msx1 homeoprotein. *Genes Dev*, **20**, 784-794.

Leonhardt, H. and Cardoso, M.C. (2000) DNA methylation, nuclear structure, gene expression and cancer. *J Cell Biochem Suppl*, **Suppl 35**, 78-83.

Leonhardt, H., Page, A.W., Weier, H.U. and Bestor, T.H. (1992) A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell*, **71**, 865-873.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. and Richmond, T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, **389**, 251-260.

Maillet, L., Boscheron, C., Gotta, M., Marcand, S., Gilson, E. and Gasser, S.M. (1996) Evidence for silencing compartments within the yeast nucleus: a role for telomere proximity and Sir protein concentration in silencer-mediated repression. *Genes Dev.*, **10**, 1796-1811.

Martinowich, K., Hattori, D., Wu, H., Fouse, S., He, F., Hu, Y., Fan, G. and Sun, Y.E. (2003) DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. *Science*, **302**, 890-893.

Mathog, D., Hochstrasser, M., Gruenbaum, Y., Saumweber, H. and Sedat, J. (1984) Characteristic folding pattern of polytene chromosomes in Drosophila salivary gland nuclei. *Nature*, **308**, 414-421.

Meaburn, K.J. and Misteli, T. (2008) Locus-specific and activity-independent gene repositioning during early tumorigenesis. *J Cell Biol*, **180**, 39-50.

Meehan, R.R., Lewis, J.D. and Bird, A.P. (1992) Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Res.*, **20**, 5085-5092.

Meshorer, E. and Misteli, T. (2006) Chromatin in pluripotent embryonic stem cells and differentiation. *Nature reviews. Molecular cell biology*, **7**, 540-546.

Muller, H. (1930) Types of visible variations induced by X-rays in Drosophila. *Journal of Genetics*, **22**, 299-334.

Nan, X., Campoy, F.J. and Bird, A. (1997) MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell*, **88**, 471-481.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N. and Bird, A. (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, **393**, 386-389.

Ohno, S. (1972) So much "junk" DNA in our genome. *Brookhaven Symp Biol*, **23**, 366-370.

Okita, K., Ichisaka, T. and Yamanaka, S. (2007) Generation of germline-competent induced pluripotent stem cells. *Nature*, **448**, 313-317.

Olins, A.L. and Olins, D.E. (1974) Spheroid chromatin units (v bodies). *Science*, **183**, 330-332.

Paik, W.K. and Kim, S. (1971) Protein methylation. *Science*, **174**, 114-119.

Park, S.H., Kook, M.C., Kim, E.Y., Park, S. and Lim, J.H. (2004) Ultrastructure of human embryonic stem cells and spontaneous and retinoic acid-induced differentiating cells. *Ultrastruct Pathol*, **28**, 229-238.

Paro, R. and Hogness, D.S. (1991) The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. *Proc Natl Acad Sci U S A*, **88**, 263-267.

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W., Solovei, I., Brugman, W., Graf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M. *et al.* (2010) Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell*, **38**, 603-613.

Pfaffeneder, T., Hackner, B., Truss, M., Munzel, M., Muller, M., Deiml, C.A., Hagemeier, C. and Carell, T. (2011) The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew Chem Int Ed Engl*, **50**, 7008-7012.

Rabl, C. (1885) *Morphologisches Jahrbuch*. Akademische Verlagsgesellschaft.

Rae, M.M. and Franke, W.W. (1972) The interphase distribution of satellite DNA-containing heterochromatin in mouse nuclei. *Chromosoma*, **39**, 443-456.

Ramocki, M.B., Tavyev, Y.J. and Peters, S.U. (2010) The MECP2 duplication syndrome. *American journal of medical genetics. Part A*, **152A**, 1079-1088.

Razin, A. and Cedar, H. (1977) Distribution of 5-methylcytosine in chromatin. *Proc Natl Acad Sci U S A*, **74**, 2725-2728.

Reddy, K.L., Zullo, J.M., Bertolino, E. and Singh, H. (2008) Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature*, **452**, 243-247.

Rett, A. (1966) [On a unusual brain atrophy syndrome in hyperammonemia in childhood]. *Wien Med Wochenschr*, **116**, 723-726.

Robinson, P.J., Fairall, L., Huynh, V.A. and Rhodes, D. (2006) EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. *Proc Natl Acad Sci U S A*, **103**, 6506-6511.

Sabbattini, P., Lundgren, M., Georgiou, A., Chow, C., Warnes, G. and Dillon, N. (2001) Binding of Ikaros to the lambda5 promoter silences transcription through a mechanism that does not require heterochromatin formation. *EMBO J*, **20**, 2812-2822.

Samaco, R.C., Hogart, A. and LaSalle, J.M. (2005) Epigenetic overlap in autism-spectrum neurodevelopmental disorders: MECP2 deficiency causes reduced expression of UBE3A and GABRB3. *Hum Mol Genet*, **14**, 483-492.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. and Cavalli, G. (2007) Genome regulation by polycomb and trithorax proteins. *Cell*, **128**, 735-745.

Smigova, J., Juda, P., Cmarko, D. and Raska, I. (2011) Fine structure of the "PcG body" in human U-2 OS cells established by correlative light-electron microscopy. *Nucleus*, **2**, 219-228.

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.

Strahl, B.D. and Allis, C.D. (2000) The language of covalent histone modifications. *Nature*, **403**, 41-45.

Swartz, M.N., Trautner, T.A. and Kornberg, A. (1962) Enzymatic synthesis of deoxyribonucleic acid. XI. Further studies on nearest neighbor base sequences in deoxyribonucleic acids. *J Biol Chem*, **237**, 1961-1967.

Szabo, P.E., Pfeifer, G.P., Miao, F., O'Connor, T.R. and Mann, J.R. (2000) Improved in vivo dimethyl sulfate footprinting using AlkA protein: DNA-protein interactions at the mouse H19 gene promoter in primary embryo fibroblasts. *Anal Biochem*, **283**, 112-116.

Szczerbal, I., Foster, H.A. and Bridger, J.M. (2009) The spatial repositioning of adipogenesis genes is correlated with their expression status in a porcine mesenchymal stem cell adipogenesis model system. *Chromosoma*, **118**, 647-663.

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.

Takahashi, K. and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, **126**, 663-676.

Takizawa, T., Meaburn, K.J. and Misteli, T. (2008) The meaning of gene positioning. *Cell*, **135**, 9-13.

Terranova, R., Sauer, S., Merkenschlager, M. and Fisher, A.G. (2005) The reorganisation of constitutive heterochromatin in differentiating muscle requires HDAC activity. *Exp Cell Res*, **310**, 344-356.

Tjeertes, J.V., Miller, K.M. and Jackson, S.P. (2009) Screen for DNA-damage-responsive histone modifications identifies H3K9Ac and H3K56Ac in human cells. *EMBO J*, **28**, 1878-1889.

Trojer, P. and Reinberg, D. (2007) Facultative heterochromatin: is there a distinctive molecular signature? *Mol Cell*, **28**, 1-13.

Tsuji-Takayama, K., Inoue, T., Ijiri, Y., Otani, T., Motoda, R., Nakamura, S. and Orita, K. (2004) Demethylating agent, 5-azacytidine, reverses differentiation of embryonic stem cells. *Biochem Biophys Res Commun*, **323**, 86-90.

Tudor, M., Akbarian, S., Chen, R.Z. and Jaenisch, R. (2002) Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. *Proc Natl Acad Sci U S A*, **99**, 15536-15541.

Waddington, C. (1942) Canalization of development and the inheritance of acquired characters. *Nature*, **150**, 557-584.

Walter, J., Schermelleh, L., Cremer, M., Tashiro, S. and Cremer, T. (2003) Chromosome order in HeLa cells changes during mitosis and early G1, but is stably maintained during subsequent interphase stages. *J Cell Biol*, **160**, 685-697.

Watt, F. and Molloy, P.L. (1988) Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev*, **2**, 1136-1143.

Wiblin, A.E., Cui, W., Clark, A.J. and Bickmore, W.A. (2005) Distinctive nuclear organisation of centromeres and regions involved in pluripotency in human embryonic stem cells. *J. Cell Sci.*, **118**, 3861-3868.

Woodcock, C.L. and Ghosh, R.P. (2010) Chromatin higher-order structure and dynamics. *Cold Spring Harb Perspect Biol*, **2**, a000596.

Wyatt, G.R. and Cohen, S.S. (1952) A new pyrimidine base from bacteriophage nucleic acids. *Nature*, **170**, 1072-1073.

Yaffe, D. and Saxel, O. (1977) Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature*, **270**, 725-727.

Yu, F., Zingler, N., Schumann, G. and Stratling, W.H. (2001) Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription. *Nucleic Acids Res*, **29**, 4493-4501.

Zaratiegui, M., Irvine, D.V. and Martienssen, R.A. (2007) Noncoding RNAs and gene silencing. *Cell*, **128**, 763-776.

Zeuschner, D., Mildner, K., Zaehres, H. and Scholer, H.R. (2010) Induced pluripotent stem cells at nanoscale. *Stem Cells Dev*, **19**, 615-620.

Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A. and Reinberg, D. (1999) Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev*, **13**, 1924-1935.

Zhao, X.Y., Li, W., Lv, Z., Liu, L., Tong, M., Hai, T., Hao, J., Guo, C.L., Ma, Q.W., Wang, L. *et al.* (2009) iPS cells produce viable mice through tetraploid complementation. *Nature*, **461**, 86-90.

Aims of this study

The aim of this work was to test the hypothesis, suggested by past studies, that gene expression is directly influenced by proximity to chromatin in the form of chromocenters or as peripheral heterochromatin. Furthermore, I wanted to elucidate whether chromatin reorganization by itself influences gene expression.

To address these questions we:

1. Generated and characterized MeCP2 antibodies suitable for these experiments.

   MeCP2 antibodies are essential tools for our studies of gene repositioning in a system with chromatin reorganization but without the gene expression profile change of differentiation.

2. Established a method for unbiased analysis and quantification of the data.

   As large-scale morphological changes take place during differentiation we needed to develop a singe cell based normalization to establish if gene repositioning is an artifact of the changed morphology or chromatin reorganization influences gene expression.

3. Analyzed the role of nuclear architecture on gene expression and the influence of genomic context.

   Using the newly developed tools we were able to address the above question in a larger scale and establish if gene proximity to heterochromatin regulates gene expression or is influenced by other factors.

Results

Chapter 1: Generation of biological tools: Antibodies

Generation and characterization of rat and mouse monoclonal antibodies specific for MeCP2 and their use in X-inactivation studies

Generation and characterization of rat and mouse monoclonal antibodies specific for MeCP2 and their use in X-inactivation studies

The data presented in this chapter has been accepted for publication by PLoS ONE:
PLoS ONE 6(11): e26499. doi:10.1371/journal.pone.0026499

The results presented in this chapter are the work of multiple people and not only myself. In order to facilitate comprehension of the data, it is presented as one work.

The contributions of all people are listed below and acknowledged by them.

K. Laurence Jost: Carried out the species reactivity test, Immunofluorescence on tissue and the X chromosome inactivation skewing experiment, writing of the manuscript and figure preparation.
Andrea Rottach: Carried out the epitope mapping
Manuela Milden: Carried out the slot blot analysis and the epitope mapping
Bianca Bertulat: Carried out the immunofluorescence on cells
Annette Becker: Carried out the antigen preparation and IP analysis
Patricia Wolf: Carried out the epitope mapping and Immunofluorescence on cells
Elisabeth Kremmer: Carried out the immunization and monoclonal antibody selection
Heinrich Leonhardt: Designed the project
M. Cristina Cardoso: Designed the project and supervised the manuscript writing

Figures contributions:
Figure 7-13 and Supplement figures: Designed and put together by K. Laurence Jost with the respective data provided by the people listed in the contributions.

INTRODUCTION

Methyl CpG binding protein 2 (MeCP2) was the second methyl CpG binding protein to be discovered (Meehan *et al.*, 1992) and the first to be cloned (Lewis *et al.*, 1992). In interphase mouse nuclei, MeCP2 is prominently localized at heterochromatic foci (Lewis *et al.*, 1992). In metaphase chromosomes, the association of MeCP2 with euchromatic arms is rather weak compared to a strong localization at pericentric heterochromatin (Lewis *et al.*, 1992), highly enriched in heavily methylated major satellite DNA repeats (Jones, 1970). MeCP2 consists of a conserved methyl CpG binding domain (MBD) that binds to 5-methyl cytosine with high affinity and is shared with the other MBD protein family members. The transcriptional repression domain (TRD), which carries a nuclear localization sequence interacts with histone deacetylases and the transcriptional corepressor Sin3A (Jones *et al.*, 1998; Nan *et al.*, 1997; Nan *et al.*, 1998). Finally, the C-terminal domain binds nucleosomes (Figure 7).

Even though MeCP2 is ubiquitously expressed, it is genetically linked to a neurological disease called Rett syndrome (RTT, OMIM 312750). RTT was first described in 1966 by Andreas Rett (Rett, 1966) and affects one in every 10,000-15,000 female births (Amir *et al.*, 1999; Hagberg, 1985; Hagberg *et al.*, 1983). Affected girls seem to develop normally until six to 18 months, subsequently they enter a developmental arrest, which is followed by strongly impaired motor skills, stereotypic hand movements, loss of speech, seizures, abnormal breathing, microcephaly, ataxia and other symptoms. Mutations within the *MECP2* gene located on chromosome Xq28 are found in approximately 80% of all classic RTT cases (Amir *et al.*, 1999; Amir & Zoghbi, 2000). Since *MECP2* is located on the X chromosome it is subjected to random X chromosome inactivation. Thus, depending on which chromosome was inactivated, a mosaic pattern of healthy (wild type allele expressing) and affected (mutant allele expressing) cells is created (Chahrour & Zoghbi, 2007). A further important aspect is the stark discrepancy between MeCP2 mRNA expression levels compared to protein levels (e.g. (Shahbazian *et al.*, 2002)), which highlights the need for highly specific antibodies detecting MeCP2 on a protein level.

Up to now rabbit polyclonal and mouse monoclonal antibodies have been raised against MeCP2 but the available antibodies are limited in their application range. Here, we describe the generation of the first rat monoclonal antibodies against MeCP2 being capable of reacting specifically in most common immunological applications. To complete the collection, we generated two mouse monoclonal antibodies and a rabbit polyclonal antibody. We could demonstrate the suitability of these high affinity and specific antibodies for immunoblotting, immunoprecipitation, and immunofluorescence stainings of cells and tissues. Additionally, we used one of our anti-MeCP2 rat monoclonal antibodies on MeCP2 heterozygous null mouse brain to analyze and quantify X chromosome inactivation skewing.

RESULTS AND DISCUSSION

Generation of rat/mouse monoclonal antibodies against MeCP2

To generate new rat and mouse antibodies potentially detecting different domains of MeCP2, we generated a baculovirus expression plasmid coding for the full length rat MeCP2 with a double strep-tag and transfected/infected Sf9 insect cells with this construct. The recombinant protein was purified using strep-tactin sepharose leading to a single band in SDS-PAGE analysis (Figure 7). The protein was used to immunize Lou/C rats and CBL mice, leading to the generation of a panel of clonal hybridomas by fusion of lymphocytes from immunized animals with the myeloma cell line P3X63-Ag8.653. All antibodies generated by the hybridomas were initially screened in a solid-phase enzyme linked immunosorbent assay (ELISA, data not shown). Positive hybridoma supernatants from clones 4H7, 4G10 and 4E1 (rat monoclonal) as well as 4B4 and 8D11 (mouse monoclonal) were stably subcloned and used for further characterization. In parallel, we immunized rabbits with untagged human MECP2 protein to generate polyclonal antibodies and used the resulting antiserum directly.

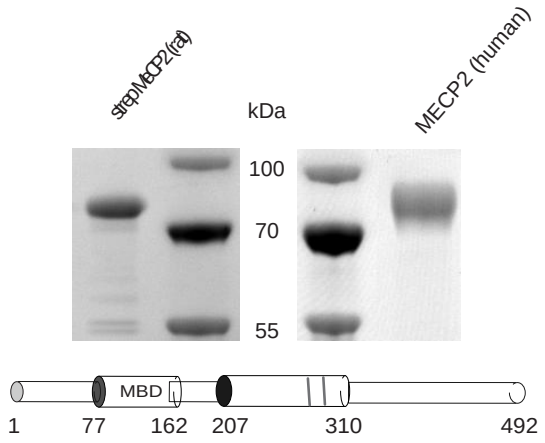


Figure 7. Antigen preparation. Purified strep-tagged MeCP2 (rat) and purified intein tagged MECP2 (human) were subjected to a SDS-PAGE and stained with Coomassie. The molecular weight markers are labeled in the middle. A schematic representation of the rat MeCP2 protein and its functional domains is shown below. MBD: methyl CpG binding domain; TRD: transcriptional repression domain; NLS: nuclear localization signal.

Sensitivity of the rat and mouse antibodies

To test the sensitivity of the antibodies we performed slot blot analysis with native rat MeCP2-GFP protein. The protein was applied in decreasing amounts ranging from 25 ng down to 0.78 ng. All monoclonal antibodies showed clear signals down to 1.56 ng of native protein and he rat monoclonal antibody 4E1 was still able to detect 0.78 ng of native protein (Figure 8). The rabbit anti MeCP2 polyclonal antiserum was also able to detect down to 0.78 ng of native protein (Figure 8). The last column contained 500 ng BSA as negative control and none of the antibodies reacted with it.

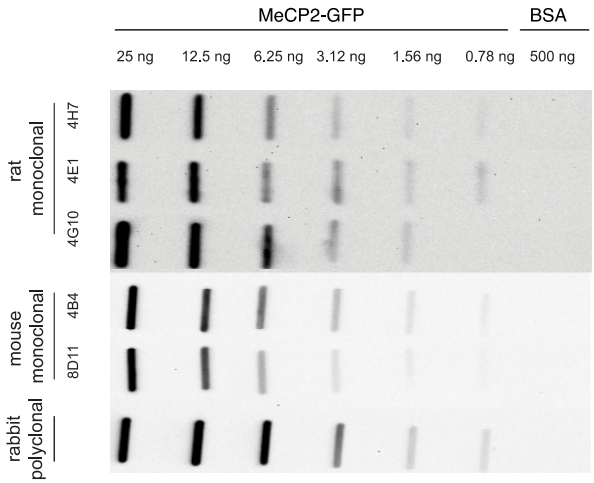


Figure 8. Antibody sensitivity. The detection limit of the anti MeCP2 antibodies was tested on native rat MeCP2 GFP by slot blotting analysis and lies between 1.58 and 0.78 ng of recombinant purified rat MeCP2.

Epitope mapping

To determine the binding domain of the new monoclonal antibodies within the MeCP2 protein, we used different constructs of GFP/YFP tagged MeCP2 expressed in mammalian cells. The cell lysates were analyzed by SDS-PAGE, blotted on a nitrocellulose membrane and incubated with the different antibodies. All fusions were expressed as controlled by incubation of the membranes with anti GFP mouse monoclonal antibody. The results (Figure 9 and Figure S1) show that the rat monoclonal antibodies 4G10 and 4H7 reacted against the C-terminus of MeCP2 and 4E1 against the N-terminus. Both mouse monoclonal antibodies 4B4 and 8D11 showed specific binding to the C-terminus. Since none of the antibodies detected the MBD domain, which is highly conserved in all MBD proteins no cross-reaction with these proteins is expected. Additionally, the polyclonal rabbit antibody detected all fragments except the TRD.

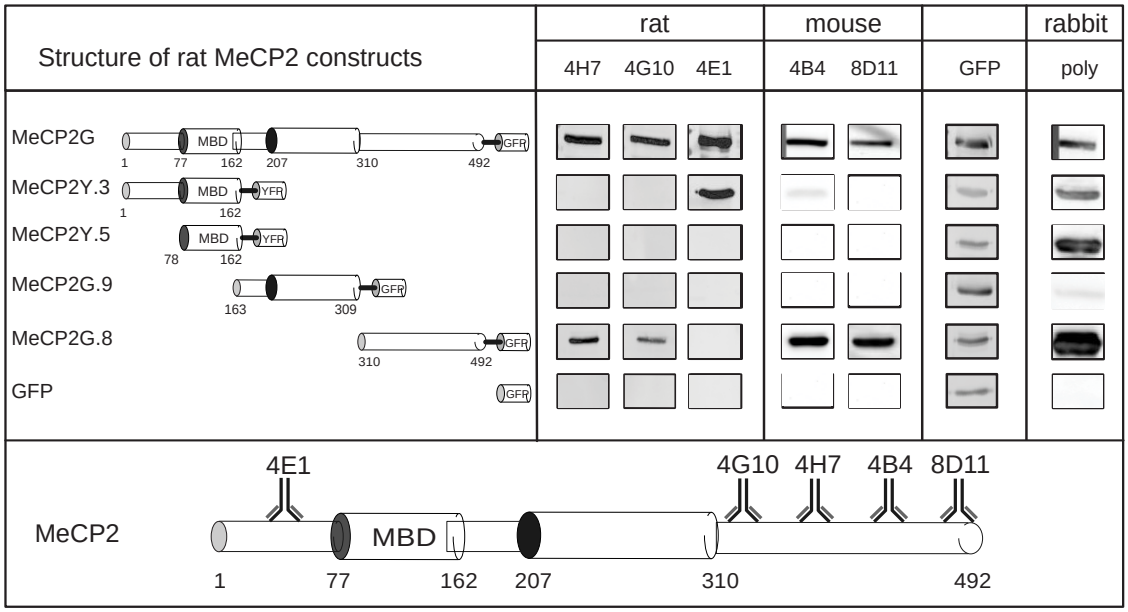


Figure 9. Epitope mapping. To determine the binding site of the new monoclonal antibodies within the MeCP2 protein, we probed extracts of mammalian cells expressing different MeCP2 constructs fused to GFP/YFP as indicated. To control for the level of the fusion proteins, the membranes were reprobed with anti GFP mouse monoclonal antibody. A summary of the epitope mapping results for the different antibodies is shown below. MeCP2 functional domains are as in Figure 7.

Specificity and cross species reactivity

MeCP2 is highly conserved throughout different species (Figure 10A). To test for cross species reactivity, nuclei from pig, mouse and rat brain tissue were isolated and extracts analyzed by western blot. As shown in Figure 10B all antibodies detected endogenous MeCP2 in mouse and rat. Remarkably, the rat monoclonal antibody 4E1 was the only one that did not detect MeCP2 in nuclear extracts from pig brain. This coincides with the fact that it is the only antibody in our tests to react with the N-terminal part (amino acids 1-78) of MeCP2 (see below). Only five amino acids are not identical in this domain of MeCP2 in the pig compared to mouse and rat. We, thus, conclude that the epitope recognized by rat monoclonal antibody 4E1 must include one or more of these residues (Figure 10A, highlighted in red).

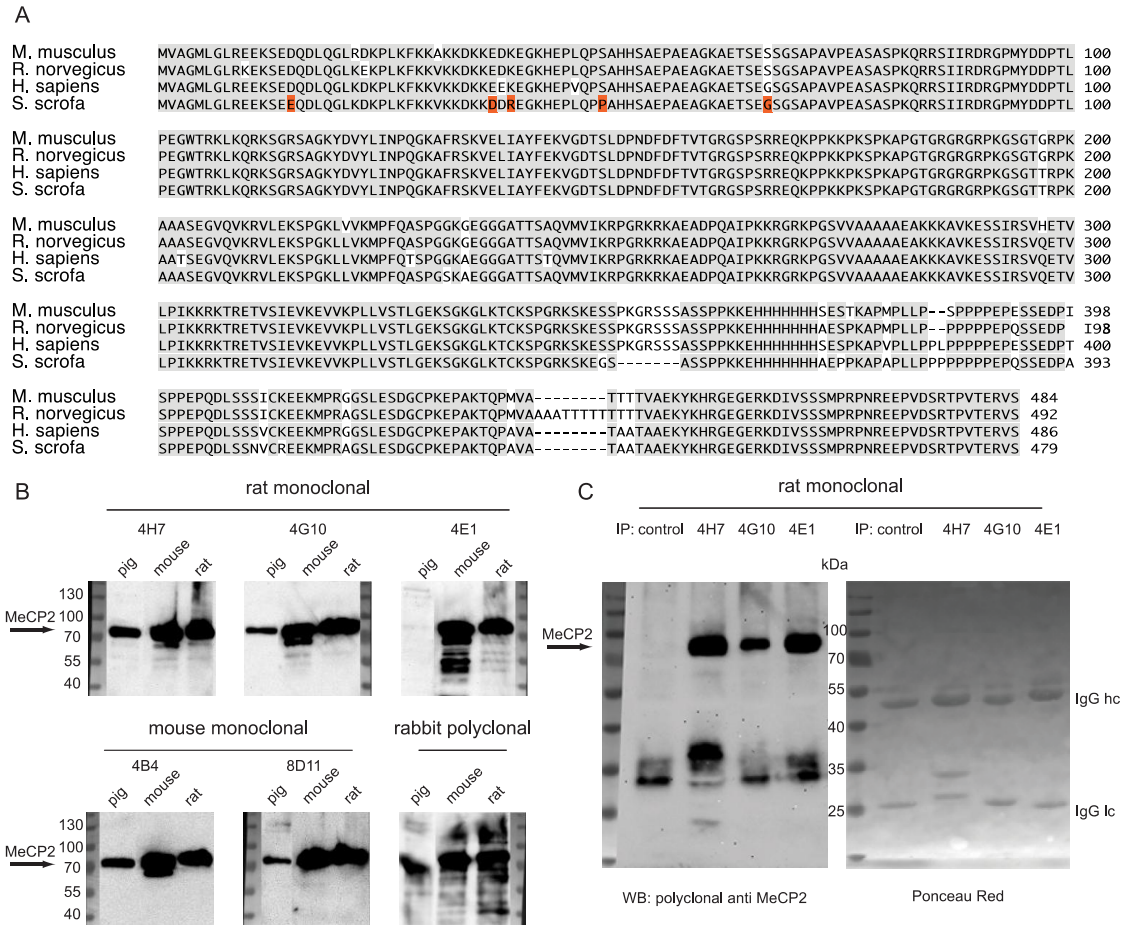


Figure 10. Antibody specificity. A Sequence alignment of MeCP2 from different species. Identical residues are shaded in gray. The identities range from 93 % (human-mouse) to 97 % (rat-mouse). **B** For a multi-species immunoblot nuclear extracts from pig, mouse and rat brain (10^6 nuclei) were loaded and probed with the antibodies as indicated. **C** For immunoprecipitation (IP) analysis, mouse brain extracts were incubated with the rat monoclonal MeCP2 antibodies as indicated. As a negative control, equal amount of anti-RFP mix rat monoclonal antibody was used. After further incubation with protein G agarose beads, the samples were loaded on a SDS-PAGE and western blot (WB) analysis was performed using the rabbit polyclonal anti MeCP2 antiserum. To control for equal amounts of monoclonal antibody used in the IP, the membrane was stained with Ponceau S. In all immunoprecipitations, an equivalent amount of the rat monoclonal IgG was detected. hc: heavy chain; lc: light chain.

An important and commonly used method for studying protein interaction partners is immunoprecipitation (IP). Thus, we tested next the ability of the monoclonal antibodies to

specifically immunoprecipitate MeCP2 from mouse brains. We could show that our three rat antibodies were able to specifically pull down MeCP2 from whole brain extract (Figure 10C). Additional unspecific bands at 35 kDa were also detected in the negative control (rat anti RFP antibody mix) (Rottach *et al.*, 2008) and are most probably due to unspecific binding to the beads. In contrast, our mouse monoclonal antibodies did not successfully immunoprecipitate MeCP2 and are thus not suitable for this application (data not shown).

In situ analysis of MeCP2 in cells and in tissue

Western blot techniques usually deal with denatured protein and do not give information about the localization of the protein in the cell. It is therefore important to test whether the new antibodies correctly detect MeCP2 localization *in situ*. MeCP2 is predominantly localized at pericentric heterochromatic regions in mouse cells, which are highly enriched in strongly methylated major satellite DNA repeats and tend to form clusters known as chromocenters (Brero *et al.*, 2005). Immunostainings were thus performed on mouse myoblasts expressing GFP tagged human MECP2 using formaldehyde and methanol as fixation reagents (Figure 11A and Figure S2).

Our three rat monoclonal antibodies revealed strong signals colocalizing with the ectopically expressed GFP-MECP2 and worked in both fixation conditions. Untransfected cells did not give a signal, consistent with undetectable endogenous levels of MeCP2 in those cells (Brero *et al.*, 2005). Using the mouse monoclonals, only 4B4 gave a signal for ectopically expressed protein. 8D11 exhibited high background noise and no specific binding in both fixation conditions. Our polyclonal rabbit antiserum showed strong and specific binding even when used at a dilution of 1:500. DAPI was used as a counterstain and additional control in all cells since DAPI's preference for AT rich regions strongly highlights chromocenters.

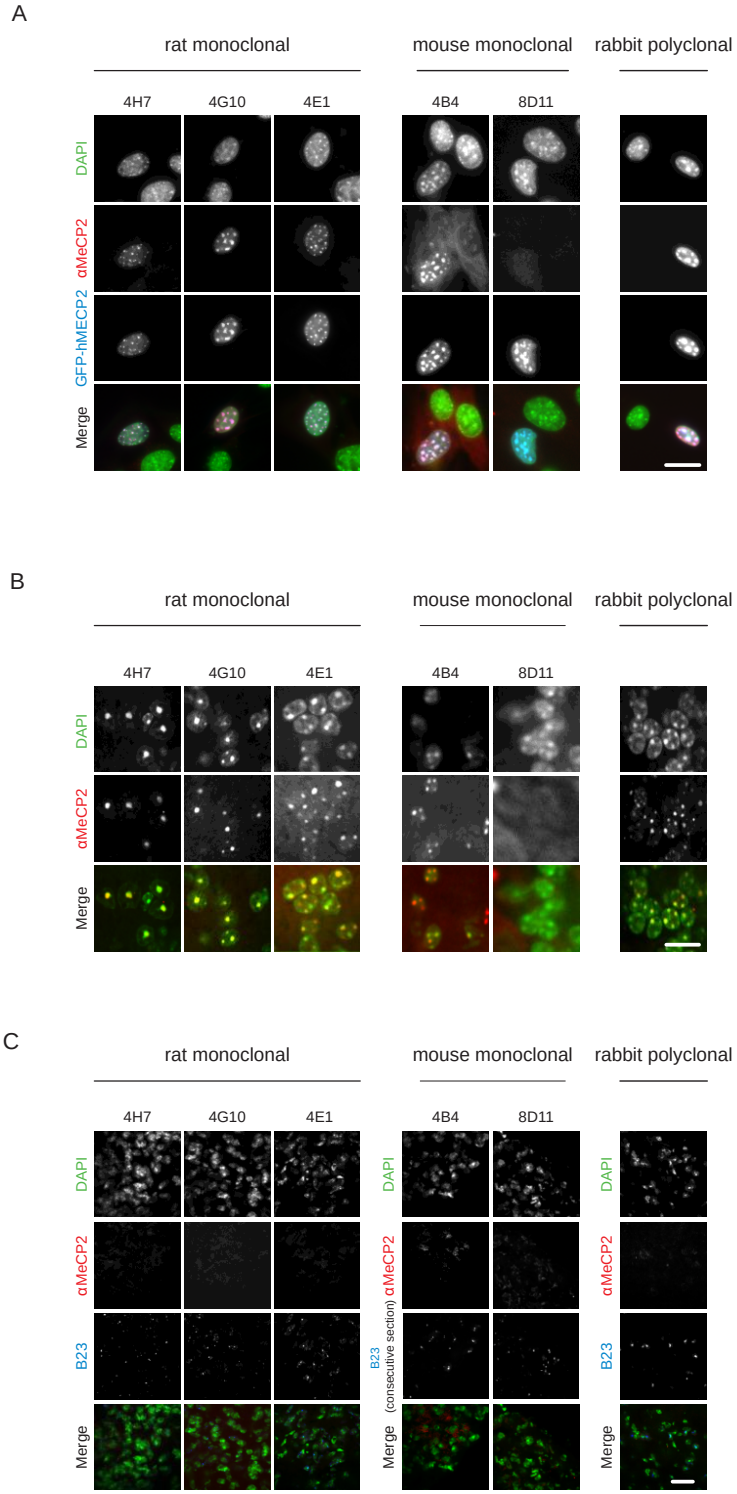


Figure 11. In situ analysis of MeCP2 in cells and tissue. A Mouse myoblasts (C2C12 cells) were transiently transfected with GFP-MECP2 (human) and fixed using formaldehyde. MeCP2 was then detected with our monoclonal antibodies (undiluted) and our rabbit polyclonal antibody (1:500). The first row shows the DNA counterstain (DAPI) of transfected and untransfected cells (green). The row underneath shows the signal obtained by our antibody staining (red). The third row shows the localization of the transfected GFP-MECP2 (blue). The merge contains an overlay of the antibody staining, the fluorescent signal of GFP-MECP2 and the DNA counterstain. Scale bar 20 µm. **B** Mouse wild type brain sections (25 µm) were stained using our antibodies. The first row shows the DNA counterstain with DAPI highlighting heterochromatic regions. The central row shows the signal obtained by immunofluorescence with our antibodies. The last row shows an overlay of DAPI and MeCP2. Scale bar 20 µm. **C** Mouse MeCP2 hemizygous null brain sections (25 µm) were stained as described above as a negative control. Mouse anti B23 antibody was used as a positive control (consecutive section when testing mouse monoclonal anti MeCP2). Scale bar 40 µm.

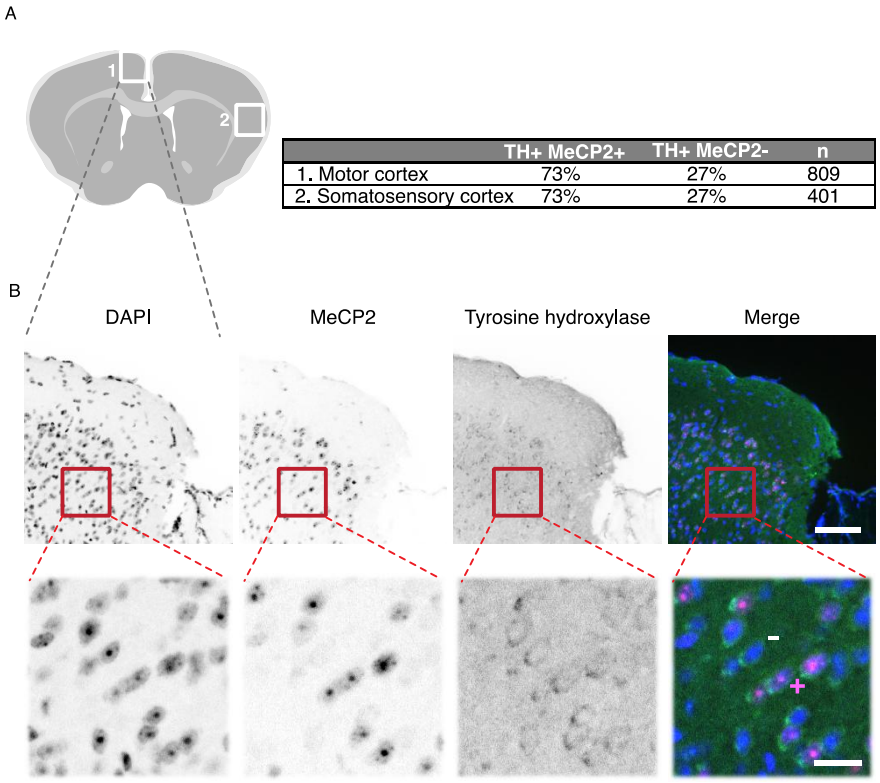
Since MeCP2 plays a crucial role in RTT syndrome one of the most important goals for us was to test whether the antibodies work on brain tissue detecting MeCP2 in its native conformation. We, therefore, prepared cryosections of wild type mouse brain and also MeCP2 hemizygous null male mouse brain as negative control. The 25 µm-thick wild type brain sections were stained with the anti MeCP2 antibodies and counterstained with DAPI as marker for chromocenters. As demonstrated in Figure 11B our rat monoclonal antibodies show a strong

and specific staining of chromocenters colocalizing with DAPI. Our mouse monoclonal antibody 4B4 shows a less intense but still specific staining of MeCP2. Unfortunately, the antibody 8D11 was not able to detect endogenous MeCP2 in brain, as it had failed to do with ectopic MeCP2 expression in cells, and is therefore not suitable for immunofluorescence. The strongest signals were achieved with our polyclonal rabbit antiserum, which was used as a positive control (Figure 11B). To verify the specificity to MeCP2 we performed the same stainings in MeCP2 null mouse brain sections. We added anti B23 nucleoli marker antibody as a positive staining control. As shown in Figure 11C none of the anti MeCP2 antibodies showed any significant signal in the knock out brain sections whereas B23 showed a clear and specific signal. The double staining with anti-B23 mouse monoclonal antibody was facilitated by the combination with our rat monoclonals, whereas, in the case of the mouse monoclonals, we had to perform the anti B23 staining using an adjacent tissue section. Additionally, using rat monoclonal antibodies obviated the cross reaction with endogenous mouse immunoglobulins present in the tissue, whereas these were readily detected when using the mouse monoclonal antibodies.

X chromosome inactivation skewing in MeCP2 heterozygous mouse brain

Since MeCP2 lies on the X chromosome it is subjected to random X chromosome inactivation in early development. In RTT, X chromosome inactivation leads to a mosaic pattern of all the cells, theoretically in a 50:50 ratio of healthy (active X containing wild type *MeCP2* allele) and affected (active X containing mutant *Mecp2* allele) cells. Deviations from this ratio indicate skewed inactivation of the X chromosome and affect the severity of RTT symptoms. Our antibodies should be highly suitable for studies concerning X chromosome inactivation as well as other studies on RTT affected brain and other tissues.

To test this, we evaluated X chromosome inactivation skewing in female MeCP2 heterozygous brain. We performed a double staining with rabbit anti tyrosine hydroxylase (TH) and our rat MeCP2 antibody (4H7). TH is the first enzyme in the biosynthesis of dopamine and norepinephrine from tyrosine and is, therefore, a marker for dopaminergic and noradrenergic neurons. Roux *et al.* (Roux *et al.*, 2008) showed that TH positive cells always co-expressed MeCP2 and, hence, X chromosome inactivation skewing can be obtained by counting TH positive cells with and without MeCP2 signal. We focused on two areas of the cortex, the motor cortex and the somatosensory cortex (Figure 12A).



	TH+ MeCP2+	TH+ MeCP2-	n
1. Motor cortex	73%	27%	809
2. Somatosensory cortex	73%	27%	401

Figure 12. X chromosome inactivation skewing in brain from heterozygous MeCP2 null mouse. A Schematic overview of a cryosection of a (female) heterozygous MeCP2 null brain with regions analyzed for X chromosome inactivation skewing marked with white squares. The results of the quantification of tyrosine hydroxylase positive, MeCP2 positive or negative cells are shown. N indicates the number of tyrosine hydroxylase positive neurons scored. **B** Representative images of a section co stained with DAPI (DNA), anti MeCP2 (4H7) and anti tyrosine hydroxylase antibodies. The motor cortex region is depicted in an overview (upper panels; scale bar 80 µm). A magnification corresponding to the red square is shown in the lower panels (scale bar 20 µm). To illustrate the scoring strategy, an example of tyrosine hydroxylase and MeCP2 positive neuron is marked by + and of tyrosine hydroxylase positive and MeCP2 negative neuron is marked by -.

In both cases we could observe a pronounced X chromosome inactivation skewing favoring wild type MeCP2 expression (73%). Previously published mouse data suggest that X chromosome inactivation skewing in brain is the reason for very different phenotypes in RTT (Young & Zoghbi, 2004). The degree of skewing is controversial and might dependent on the tissue analyzed or the method applied (Gibson *et al.*, 2005; Young & Zoghbi, 2004). Our antibodies could help to elucidate the state of X chromosome inactivation in RTT tissue in particular also with respect to truncated versus full length MeCP2 by a combination of the N and C-terminal specific antibodies.

Figure 13 summarizes the characterization of the novel anti MeCP2 antibodies. The antibodies recognize MeCP2 from different species, including human, mouse, rat and pig. Whereas the two new mouse antibodies are suitable for western blot and to a lesser extend for immunofluorescence, the rabbit polyclonal as well as the rat monoclonal antibodies performed very well in immunoblotting, immunoprecipitation, and immunofluorescence analysis of ectopic and endogenous MeCP2 making them a very valuable set of tools for studies of MeCP2 pathophysiology *in situ* and *in vitro*.

	rat monoclonal			mouse monoclonal		rabbit polyclonal
	4H7	4G10	4E1	4B4	8D11	
Subtype	IgG2a	IgG2a	IgG2a	IgG2a	IgG2b	
Native immuno (slot) blot	✓	✓	✓	✓	✓	✓
Immunoprecipitation	✓	✓	✓	–	–	✓
Western blot	✓	✓	✓	✓	✓	✓
Immunofluorescence (cells) [1]	✓	✓	✓	✓	–	✓
Immunofluorescence (tissue)	✓	✓	✓	✓	–	✓
Epitope	C	C	N	C	C	
Species reactivity [2]	H, P, R, M	H, P, R, M	H, R, M	H, P, R, M	P, R, M	H, P, R, M

[1] Methanol and formaldehyde fixation [2] Not all species were tested in all methods listed above H = Human; P = Pig; R = Rat; M = Mouse

Figure 13. Summary of the characterization of the rabbit, rat and mouse anti MeCP2 antibodies.

MATERIALS AND METHODS

Plasmids

Mammalian expression constructs (Figure 9 and Figure S1) coding for GFP or YFP-tagged rat MeCP2 full length (MeCP2G) and domain constructs (MeCP2Y.3 and MeCP2Y.5) were previously described (Agarwal *et al.*, 2007; Brero *et al.*, 2005). The mammalian expression constructs MeCP2G.9 and MeCP2G.8 were generated from the above plasmids by PCR amplification using the following primers:

pMeCP2G.9 ss ccgctcgaggccatggggagcccttccaggagagaaca

as cgcggatccttccgggtcttgcgcttcttgatggggagcac

pMeCP2G.8 ss ggaagatctgccatggaaaccgtcagcattgaggtcaag

as ataagaatgcggccgcttacttgtacagctcgtccatgcc

The mammalian expression construct (Figure 11 and Figure S2) expressing GFP-tagged human MECP2 was described before (Kudo *et al.*, 2003) and was provided by S. Kudo (Hokkaido Institute of Public Health, Sapporo, Japan). For expression in Sf9 (Invitrogen Paisley PA4 9RF, UK) insect cells the Bac-to-Bac baculovirus expression system (Invitrogen Paisley PA4 9RF, UK) was used. To express MeCP2 with a N-terminal double strep-tag (Figure 7), a sequence encoding the strep-tactin target peptide strep tag III (MWSHPQFEKGGGSTGGGSGGGSWSHPQFEK) was synthesized (Entelechon, Bad Abbach, Germany) flanked by BamHI and NotI sites and subcloned into pFastBac1 (Invitrogen, Paisley PA4 9RF, UK) using the same sites. Rat MeCP2 full length was generated by PCR amplification from MeCP2G (described above) using the following primers:

MeCP2 ss: ggaagatctgccatggaaaccgtcagcattgaggtcaag

as: ataagaatgcggccgcttacttgtacagctcgtccatgcc

with NotI and XhoI sites and subcloned in frame with the strep-tag in the pFastBac1 vector. For expression of rat MeCP2-GFP in Sf9 insect cells (Figure 8) the mammalian expression construct coding for MeCP2G full length (described above) was cut using NotI and XhoI and cloned in frame in the pFastBac1 vector.

A prokaryotic expression construct coding for intein tagged human MECP2 (Figure 7) (Yusufzai & Wolffe, 2000) was obtained from C.L. Woodcock (University of Massachusetts, Amherst, USA).

Tissues

Male mouse MeCP2 hemizygous brains (Figure 11) (Guy *et al.*, 2001) were kindly provided by the group of P. Huppke (Georg August University, Göttingen, Germany). Female mouse MeCP2 null heterozygous brains (Figure 12) (Guy *et al.*, 2001) were kindly provided by the group of L. Villard (Faculte de Medecine La Timone, Marseille, France). Wild type mouse brains

(C57BL/6N; Charles River Laboratories International, Inc., Wilmington, MA 01887, USA) were used as control. Mice were over 10 months old.

Cell culture and transfection

For immunofluorescence (IF) experiments mouse C2C12 myoblasts (Yaffe & Saxel, 1977) were cultured using standard conditions described previously (Cardoso *et al.*, 1993). For subsequent IF experiments (Figure 11 and Figure S2), C2C12 cells were transiently transfected with human GFP-MECP2 expression construct (Kudo *et al.*, 2003) using Transfectin (Bio Rad, München, Germany) according to manufacturer's advice.

For the epitope mapping (Figure 9 and Figure S1), human embryonic kidney (HEK) 293T (Suetake *et al.*, 2011) cells were cultured in DMEM supplemented with 10 % fetal calf serum and 50 µg/ml gentamicin and transfected with full length rat MeCP2 and domain constructs (described above) using polyethylenimine (Sigma, St. Louis, MO, USA).

MeCP2 antigens for immunization (Figure 7) and slot blot applications (Figure 8) were produced using the baculovirus system in Sf9 insect cells. Sf9 cells were maintained in EX-CELL 420 Insect Serum Free (SAFC) medium supplemented with 10 % fetal bovine serum shaking at 100 rpm and 28 °C. Transfection of Sf9 cells to produce recombinant baculovirus, was performed using Cellfectin (Invitrogen, Paisley PA4 9RF, UK) according to the manufacturer's instructions.

Antigen purification

Sf9 insect cells were infected with the recombinant baculovirus (coding for MeCP2 with N-terminal double strep-tag; (Gloeckner *et al.*, 2007)) and incubated at 28 °C with shaking for 5 days. The cells were pelleted by centrifugation (200 x g, 5 min, 4 °C) and resuspended in a buffer containing 25 mM Tris-HCl, pH 8.0; 1 M NaCl; 50 mM glucose; 10 mM EDTA; 0.2 % Tween-20; 0.2 % NP40. The buffer was supplemented with protease inhibitors (Complete mini; Roche, Mannheim, Germany). After incubation on ice for 10 min, cells were disrupted with a high-pressure homogenizer (EmulsiFlex-C5, Avestin) followed by centrifugation at 14,000 x g for 30 min.

Strep-tagged recombinant rat MeCP2 protein was purified by incubating the supernatant with 500 µl of strep-tactin sepharose beads (IBA, Göttingen, Germany) for 4 h at 4 °C on a rotary shaker. To elute strep-tagged proteins, the beads were incubated with D-Desthiobiotin (0.5 mg/ml; IBA, Göttingen, Germany), dissolved in 1x PBS, for 30 min at 4 °C. After centrifugation (200 x g, 2 min), beads were separated from the eluate containing the purified proteins. The elution step was performed three successive times.

Intein tagged human MECP2 protein was purified as previously described resulting in untagged MECP2 through protein splicing (Georgel *et al.*, 2003).

Immunizations, generation of hybridomas and ELISA screening

Monoclonal antibodies specific for MeCP2 were generated via the hybridoma technology as described by Rottach *et al.* (Rottach *et al.*, 2008). 80 µg of a N-terminal, strep-tagged full length rat MeCP2 were injected both intraperitoneally and subcutaneously into Lou/C rats and CBL mice using CpG2006 (TIB MOLBIOL, Berlin, Germany) as adjuvant. 8 weeks later and 3 days before fusion a boost was given intraperitoneally and subcutaneously. Spleen cells were isolated and fused to the myeloma cell line P3X63-Ag8.653 (ATCC, Rockville, MD, USA) using polyethylene glycol 1500 (PEG 1500, Roche, Mannheim, Germany). After fusion, cells were cultured in 96-well plates using RPMI 1640 with 20% fetal calf serum, penicillin/streptomycin, glutamine, pyruvate, and non-essential amino acids (PAA, Cölbe, Germany) supplemented by aminopterin (Sigma, St. Louis, MO, USA). The hybridoma supernatants were tested in a solid-phase enzyme linked immunosorbent assay (ELISA). Microtiter plates were coated over night with strep-tagged rat MeCP2 at a concentration of 3–5 g/ml in 0.1 M sodium carbonate buffer (pH 9.6) and blocked with non-fat milk (Frema, Neuform, Zarrentin, Germany). The hybridoma supernatants were added and the bound monoclonal antibodies were detected using a cocktail of biotinylated mouse monoclonal antibodies against the rat IgG heavy chains, thus avoiding the detection of IgM mouse monoclonal antibodies (anti IgG1, anti IgG2a, anti IgG2b [ATCC, Manassas, VA], anti IgG2c [Ascenion, Munich, Germany]). For visualization, peroxidase-labeled avidin (Alexis, San Diego, CA) antibodies were applied and o-phenylenediamine was used as chromogen in the peroxidase reaction. The clones 4H7, 4G10 and 4E1 (rat monoclonal) as well as 4B4 and 8D11 (mouse monoclonal) were stably subcloned and used for further characterization.

The rabbit polyclonal antibody was generated using the untagged human MECP2 according to the Express rabbit protocol from PickCell (PickCell, Amsterdam, Netherlands) and used in form of antiserum.

Ethics statement

Immunizations of mice and rats for the purpose of generating monoclonal antibodies were approved by the Government of Upper Bavaria, according to the animal experimentation law § 8a, permit number 209.1/211-2531.6-4/99.

Sensitivity assay via slot blot analysis

Purification of MeCP2-GFP

Sf9 insect cells were infected with the recombinant baculovirus (coding for rat MeCP2-GFP) and incubated at 28 °C with shaking for 5 days. The cells were pelleted and resuspended as explained above for strep-tag MeCP2 and disrupted by sonication (three times each for 25 seconds, 70% power; Bandelin Sonopuls GM70, Sonontrode HD70, Berlin, Germany) on ice. Lysates were cleared by centrifugation at 15,000 x g for 30 min at 4 °C.

Recombinant rat MeCP2-GFP protein was purified by incubating 200 ml whole cell lysate with 1 ml (1.5 mg/ml) GBP nanotrap according to the manufacturer's advice (Chromotek, Planegg-Martinsried, Germany). After transfer of the GBP nanotrap beads containing lysate to a Bio-Rad Poly-Prep chromatography column (Cat: 731-1550, Bio-Rad Laboratories, Hercules CA 94547, USA) the column was washed three times with 10 ml PBS. To elute the MeCP2-GFP protein, the beads were incubated with 5 ml of a high salt buffer. Buffer exchange was done with PBS using Amicon ultra centrifugal filters (Ultracel 10 kDa molecular weight cutoff; Millipore, Ireland). Eluted protein was quantified with Pierce the 660 nm Protein Assay (Thermo Scientific; Pro: #1861426, Schwerte, Germany) and checked by SDS-PAGE analysis (data not shown).

Slot blotting analysis

Native MeCP2-GFP was spotted directly onto a nitrocellulose membrane (GE Healthcare, München, Germany). Membranes were incubated in blocking buffer, 5% (w/v) non-fat dry milk in PBS (PBSM), for 20 min at room temperature. Primary antibodies were used undiluted and incubated for 2 h at room temperature, followed by three washes in PBS/0.1% Tween-20. Subsequently, membranes were incubated for 1 h at room temperature with horseradish peroxidase conjugated anti-rabbit IgG (Sigma, St. Louis, MO, USA) diluted 1:10,000 or anti-mouse (GE Healthcare, München, Germany) and rat IgG (Sigma, St. Louis, MO, USA) 1:5,000 in 5% (w/v) PBSM. After three washing steps in PBS/0.1% Tween-20, signals were detected with ECL (GE Healthcare, München, Germany).

Epitope mapping

For epitope mapping, different constructs of rat MeCP2 with C-terminal GFP or YFP tag were used for transient transfection of HEK 293T cells. After cell lysis (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5%NP40, 2 mM PMSF, 0.5 mM EDTA, 1x mammalian protease inhibitor mix, 1 mg/ml DNase, 2 mM $MgCl_2$) the concentration of the GFP fusion proteins was calculated using a fluorescent read out of GFP and YFP (Infinite® M1000, TECAN), respectively (GFP: excitation wavelength: 490 nm, emission wavelength: 511 nm, YFP: excitation wavelength: 525 nm, emission wavelength: 538 nm). The protein concentration was normalized to the construct with the lowest expression rate and lysates were diluted accordingly (228 nM GFP or YFP). The samples were boiled in Laemmli sample buffer at 95 °C for 10 min and loaded on a 10% SDS-PAGE. Western blot analysis was performed as described above. In addition to the polyclonal and monoclonal anti MeCP2 antibodies, anti GFP mouse monoclonal antibody (Cat: 11814460001, Roche Diagnostics GmbH, Mannheim, Germany) was used to control for expression level of the different deletion proteins.

Cross-species reactivity assay via western blot analysis

For western blot analysis brain cell nuclei were extracted from pig (obtained fresh from the local slaughterhouse), mouse and rat (Charles River Laboratories International, Inc., Wilmington, MA

01887, USA) as described (Prusov & Zatsepina, 2002) and lysed in RIPA buffer (50 mM Tris/HCl pH 8, 150 mM NaCl, 1% Tween, 0.5% Doc, 0.1% SDS). For each gel lane, lysates from 10^6 nuclei were loaded.

Samples were separated on a 10% SDS-PAGE and transferred to a nitrocellulose membrane (GE Healthcare, München, Germany). The following primary antibodies were used for western blot analysis: rabbit polyclonal anti MeCP2 (1:500), mouse monoclonal and rat monoclonal anti MeCP2 (undiluted). Secondary antibodies were as above for slot blot analysis.

Immunoprecipitation

Mechanically disrupted mouse brain tissue (3 - 4 brains) was dissolved in buffer A (20mM Tris pH 7.9, 0.6 M NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.4% NP-40), and then diluted with buffer B (20mM Tris pH 7.9, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.4% NP-40) to obtain an NaCl concentration of 200 mM. Mouse brain extracts were incubated with 400 µl of the monoclonal MeCP2 antibody indicated, at 4 °C for 2 h with shaking. As negative control, anti RFP mix rat monoclonal antibody (Rottach *et al.*, 2008) of equal amount was used. 100 µl protein G agarose beads, that were equilibrated with buffer B, were added and incubated with the extract for 1 h at 4 °C with shaking. After three washes with buffer B immuno complexes were dissolved in 60 µl 1x Laemmli sample buffer.

Samples were separated on a 10% SDS-PAGE, transferred to a nitrocellulose membrane (GE Healthcare, München, Germany) and western blot analysis was performed using rabbit polyclonal MeCP2 antiserum (1:500) as primary antibody. Signals were detected with ECL (GE Healthcare, München, Germany). As loading and blotting control, the membrane was stained with Ponceau S after transfer.

Immunofluorescence analysis

Cells

For immunofluorescence staining, C2C12 cells were seeded on glass coverslips and transiently transfected with GFP tagged MECP2 (human). Cells were fixed with 3.7% formaldehyde in PBS and incubated with the undiluted rat/mouse anti MeCP2 antibodies for 1 h at room temperature. After incubation with the secondary Alexa 647 conjugated goat anti rat/mouse IgG antibody (Invitrogen Paisley PA4 9RF, UK) diluted 1:400 in PBS containing 2% BSA, the cells were counterstained with DAPI (2 µg/ml) and mounted in Vectashield medium (Vector Labs, Burlingame, CA, USA).

Tissues

Mouse brains were fixed by overnight immersion in PBS-buffered 4% paraformaldehyde. The brains were embedded in Tissue Tek (Sakura, Zoeterwoude, Netherlands) and cryosectioned (25 µm) using a cryostat HM 560 (Microm, Walldorf, Germany).

Sections were air dried at room temperature for 30 min, re-hydrated in 10 mM sodium citrate buffer (pH 6.0) for 5 min, pulse-heated (80 °C) for 30 min in the microwave. The slides were equilibrated in PBS after heating and incubated with the following antibodies: anti MeCP2 mouse monoclonal (undiluted), rat monoclonal (undiluted), rabbit polyclonal (1:500), anti B23 mouse monoclonal (Sigma, St. Louis, MO, USA, 1:1,000) and anti tyrosine hydroxylase rabbit antibody (AB152, Millipore, Billerica, MA, USA) Both, primary and secondary antibodies were complemented with 0.1% Triton X-100 and 1% BSA. No additional blocking step was performed. Incubation was done under a glass chamber (made of coverslips) in a humid box for 12-24 h at room temperature (Solovei *et al.*, 2009). Washings between antibody incubations and after incubation with secondary antibodies were performed with PBS with 0.05% Triton X-100 at 37 °C, 3 x 20 min. In order to stabilize preparations, immunostained sections were post-fixed with 2% paraformaldehyde for 10 min before counterstaining with DAPI (2 µg/ml) for 1 h and mounted in Vectashield medium (Vector Labs, Burlingame, CA, USA).

Microscopy

Epifluorescence images were obtained on a Zeiss Axiovert 200 microscope equipped with Plan-Apochromat x63/1.4 numerical aperture (NA) oil immersion objective lenses and a Sensicam (PCO) CCD camera. Confocal images were collected using an UltraVIEW VoX spinning disc system (Perkin Elmer) on a Nikon Ti microscope equipped with an oil immersion Plan-Apochromat x40/1.3 NA objective lens (pixel size in XY = 186 nm, Z-step = 0.3 µm).

Scoring of tyrosine hydroxylase and MeCP2 positive cells was done by eye in z-stacks.

REFERENCES

Agarwal, N., Hardt, T., Brero, A., Nowak, D., Rothbauer, U., Becker, A., Leonhardt, H. and Cardoso, M.C. (2007) MeCP2 interacts with HP1 and modulates its heterochromatin association during myogenic differentiation. *Nucleic Acids Res*, **35**, 5402-5408.

Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U. and Zoghbi, H.Y. (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature Genetics*, **23**, 185-188.

Amir, R.E. and Zoghbi, H.Y. (2000) Rett syndrome: methyl-CpG-binding protein 2 mutations and phenotype-genotype correlations. *Am J Med Genet*, **97**, 147-152.

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.

Cardoso, M.C., Leonhardt, H. and Nadal-Ginard, B. (1993) Reversal of terminal differentiation and control of DNA replication: cyclin A and Cdk2 specifically localize at subnuclear sites of DNA replication. *Cell*, **74**, 979-992.

Chahrour, M. and Zoghbi, H.Y. (2007) The story of Rett syndrome: from clinic to neurobiology. *Neuron*, **56**, 422-437.

Georgel, P.T., Horowitz-Scherer, R.A., Adkins, N., Woodcock, C.L., Wade, P.A. and Hansen, J.C. (2003) Chromatin compaction by human MeCP2. Assembly of novel secondary chromatin structures in the absence of DNA methylation. *J Biol Chem*, **278**, 32181-32188.

Gibson, J.H., Williamson, S.L., Arbuckle, S. and Christodoulou, J. (2005) X chromosome inactivation patterns in brain in Rett syndrome: implications for the disease phenotype. *Brain Dev*, **27**, 266-270.

Gloeckner, C.J., Boldt, K., Schumacher, A., Roepman, R. and Ueffing, M. (2007) A novel tandem affinity purification strategy for the efficient isolation and characterisation of native protein complexes. *Proteomics*, **7**, 4228-4234.

Guy, J., Hendrich, B., Holmes, M., Martin, J.E. and Bird, A. (2001) A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat Genet*, **27**, 322-326.

Hagberg, B. (1985) Rett's syndrome: prevalence and impact on progressive severe mental retardation in girls. *Acta Paediatr Scand*, **74**, 405-408.

Hagberg, B., Aicardi, J., Dias, K. and Ramos, O. (1983) A Progressive Syndrome of Autism, Dementia, Ataxia, and Loss of Purposeful Hand Use in Girls - Retts Syndrome - Report of 35 Cases. *Annals of Neurology*, **14**, 471-479.

Jones, K.W. (1970) Chromosomal and nuclear location of mouse satellite DNA in individual cells. *Nature*, **225**, 912-915.

Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J. and Wolffe, A.P. (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet*, **19**, 187-191.

Kudo, S., Nomura, Y., Segawa, M., Fujita, N., Nakao, M., Schanen, C. and Tamura, M. (2003) Heterogeneity in residual function of MeCP2 carrying missense mutations in the methyl CpG binding domain. *J Med Genet*, **40**, 487-493.

Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F. and Bird, A. (1992) Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell*, **69**, 905-914.

Meehan, R.R., Lewis, J.D. and Bird, A.P. (1992) Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Res*, **20**, 5085-5092.

Nan, X., Campoy, F.J. and Bird, A. (1997) MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell*, **88**, 471-481.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N. and Bird, A. (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, **393**, 386-389.

Prusov, A.N. and Zatsepina, O.V. (2002) Isolation of the chromocenter fraction from mouse liver nuclei. *Biochemistry (Mosc)*, **67**, 423-431.

Rett, A. (1966) [On a unusual brain atrophy syndrome in hyperammonemia in childhood]. *Wien Med Wochenschr*, **116**, 723-726.

Rottach, A., Kremmer, E., Nowak, D., Boisguerin, P., Volkmer, R., Cardoso, M.C., Leonhardt, H. and Rothbauer, U. (2008) Generation and characterization of a rat monoclonal antibody specific for PCNA. *Hybridoma*, **27**, 91-U99.

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.

Roux, J.C., Dura, E. and Villard, L. (2008) Tyrosine hydroxylase deficit in the chemoafferent and the sympathoadrenergic pathways of the Mecp2 deficient mouse. *Neurosci Lett*, **447**, 82-86.

Shahbazian, M.D., Antalffy, B., Armstrong, D.L. and Zoghbi, H.Y. (2002) Insight into Rett syndrome: MeCP2 levels display tissue- and cell-specific differences and correlate with neuronal maturation. *Hum Mol Genet*, **11**, 115-124.

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.

Suetake, I., Mishima, Y., Kimura, H., Lee, Y.H., Goto, Y., Takeshima, H., Ikegami, T. and Tajima, S. (2011) Characterization of DNA-binding activity in the N-terminal domain of the DNA methyltransferase Dnmt3a. *Biochem J*, **437**, 141-148.

Yaffe, D. and Saxel, O. (1977) Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature*, **270**, 725-727.

Young, J.I. and Zoghbi, H.Y. (2004) X-chromosome inactivation patterns are unbalanced and affect the phenotypic outcome in a mouse model of rett syndrome. *Am J Hum Genet*, **74**, 511-520.

Yusufzai, T.M. and Wolffe, A.P. (2000) Functional consequences of Rett syndrome mutations on human MeCP2. *Nucleic Acids Res*, **28**, 4172-4179.

SUPPLEMENTARY INFORMATION

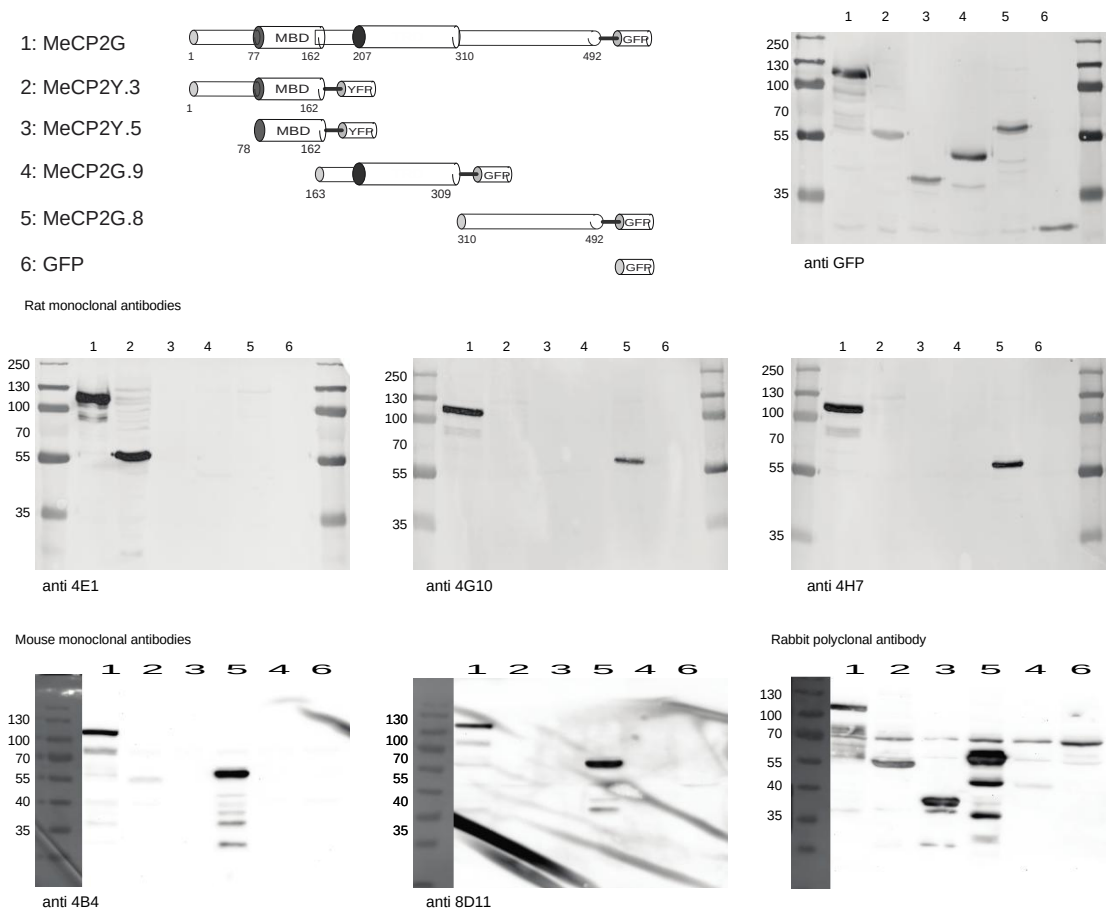


Figure S1. Epitope mapping. Complete blots of the epitope mapping presented in Figure 9 together with a schematic representation of the constructs.

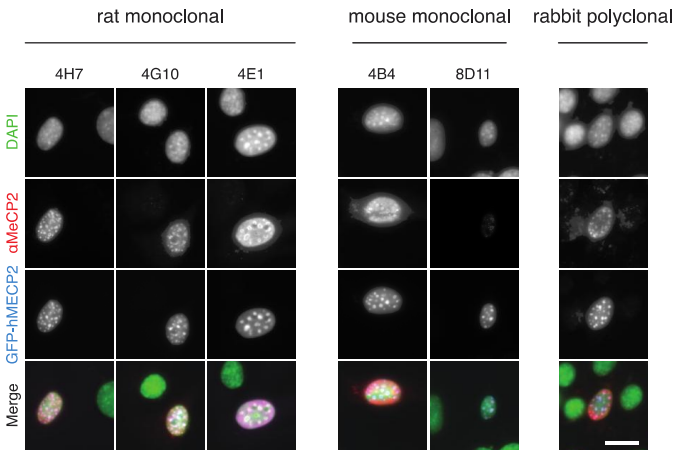


Figure S2. *In situ* analysis of MeCP2 in cells. Mouse myoblasts (C2C12 cells) were transiently transfected with GFP-MECP2 (human) and fixed with methanol. MECP2 was then detected with our monoclonal antibodies (undiluted) and our rabbit polyclonal antibody (1:500). The first row shows the DNA counterstain (DAPI) of transfected and untransfected cells (green). The row underneath shows the signal obtained by our antibody staining (red). The third row shows the localization of the transfected GFP-MECP2 (blue). The merge contains an overlay of the antibody staining, the fluorescent signal of GFP-MECP2 and the DNA counterstain. Scale bar 20 µm.

Chapter 2: Generation of computational tools: Software

3D-Image analysis platform monitoring relocation of pluripotency genes during reprogramming

The data presented in this chapter has been published:
Nucl. Acids Res. (2011) 39 (17): e113. doi: 10.1093/nar/gkr486

The results presented in this chapter are the work of multiple people and not only myself. In order to facilitate comprehension of the data, it is presented as one work.

The contributions of all people are listed below and acknowledged by them.

K. Laurence Jost: Designed the project, analyzed the data, tested and supported the software development and wrote the manuscript.
Sebastian Haase: Developed the software and supported the manuscript writing
Daniel Smeets: Carried out the cell culturing and FISH experiments
Nadine Schrode: Carried out the cell culturing and FISH experiments
Jörn M. Schmiedel: Developed the normalization procedure
Bianca Bertulat: Supported the software development
Hanspeter Herzel: Developed the normalization procedure
Marion Cremer: Designed the project and supported the manuscript writing
M. Cristina Cardoso: Designed the project and supervised the manuscript writing

Figures contributions:

Figure 14A: Designed by Bianca Bertulat
Figure 14bB-16C: Designed and put together by K. Laurence Jost
Figure S3: Designed by Marion Cremer
Figure S4-S9 and movies: Designed and put together by K. Laurence Jost

INTRODUCTION

In the last years, nuclear organization of chromatin has increasingly come into focus as an important level of genome regulation (Lanctot *et al.*, 2007; Meaburn *et al.*, 2007). During differentiation, distinct positional changes of specific genes upon transcriptional activation have been reported. One of the first correlations between gene repositioning and its activation level was observed in mouse B-lymphocyte maturation. Brown *et al.* (Brown *et al.*, 1997) showed that gene repositioning away from centromeric regions could be seen upon activation. In this study, distances were not measured but colocalization with the centromeric region was evaluated. A more recent study, on *Drosophila*, reported distance measurements taking chromatin mobility into account and also concluded a close association of silenced genes with centromeric heterochromatin (Harmon & Sedat, 2005). Similar observations have been reported for relocation of genes to the nuclear lamina (Reddy *et al.*, 2008). However, it is unclear whether proximity to heterochromatin, in general, regulates gene (in)activity (Deniaud & Bickmore, 2009; Joffe *et al.*, 2010). The comparison of different biological data sets is complicated by variability of their framework parameters such as drastic changes of nuclear morphology during differentiation. Up to now, software for distance measurement has not taken into account such morphological changes or do not measure in 3D (Gue *et al.*, 2005; Shirley *et al.*, 2011). The application of sophisticated normalization procedures becomes mandatory to produce meaningful and unbiased objective data evaluation. Prominent morphological and transcriptional changes occur upon gain and loss of pluripotency, making it an ideal test system for studying repositioning of genes during cellular differentiation. The generation of induced pluripotent stem (iPS) cells requires the expression of the pluripotency genes Oct4 and Sox2 (Huangfu *et al.*, 2008). Additional factors such as Nanog have been shown to increase the efficiency of the process (Silva *et al.*, 2006). Accordingly these three genes are up regulated at the transcriptional level in embryonic stem (ES) cells and iPS cells.

In this study, we describe a novel method to investigate whether changes in gene positioning within the nucleus correlate with their transcriptional status and/or genomic context. Our computational analysis includes a series of filters to segment the objects to be analyzed, in our application, the nuclear periphery, the chromocenters and the gene loci. This is followed by a single cell-based normalization procedure, which permits the comparison of data sets exhibiting large morphological variability. In brief, using this approach, we could show that gene position relative to heterochromatin does not correlate with silencing, but internal gene positioning is compatible with expression in pluripotent cells and may be influenced by the surrounding gene density in mouse embryonic fibroblasts (MEFs).

RESULTS AND DISCUSSION

Signal detection and segmentation are compromised by biological and technical parameters

For monitoring gene position in 3D during induced pluripotency, we measured the shortest distance between the gene locus and nuclear landmarks i.e. (1) the nearest pericentric heterochromatin (chromocenter) surface and, (2) the nuclear periphery (Figure 14A). Data were compared between MEF, iPS and ES cells in order to determine differences and similarities (Figure 14B). These measurements require the definition of the gene position and correct segmentation of compartments of interest (chromocenters and periphery). To achieve high throughput, signal detection and segmentation were automated. The first step consists of the segmentation of chromocenters and nuclear periphery. Gray scale image stacks are processed using a series of software filters (Figure S5) resulting in a binary (segmented) image for heterochromatic regions (chromocenters) and the nuclear outline (nuclear periphery). The settings of the filtering step are set once and can be applied for a complete set of images stacks and are automatically saved. A challenging problem during segmentation is thresholding since the apparent size of chromocenters is drastically changed by this procedure (Figure 14D). To minimize subjective bias on threshold levels with subsequent significant impact on absolute distance measurements (lower threshold leading to shorter distances and vice versa), we have applied a basic global threshold approach calculating the threshold for each image individually used minimizing bias by using fix thresholds. The detection of the relevant fluorescence signals (FISH gene signals) was also automated in order to increase image analysis throughput. Gene signals can vary in size and intensity based on hybridization efficiency and antibody sensitivity. We, therefore, added a feature to set the size and intensity of signals in order to decrease false detection of background noise. The settings can be chosen individually for each image stack to ensure best possible definition of gene signals. After segmentation the signal is reduced to a 3D pixel with sub-pixel localization precision, which is subsequently used for measurements.

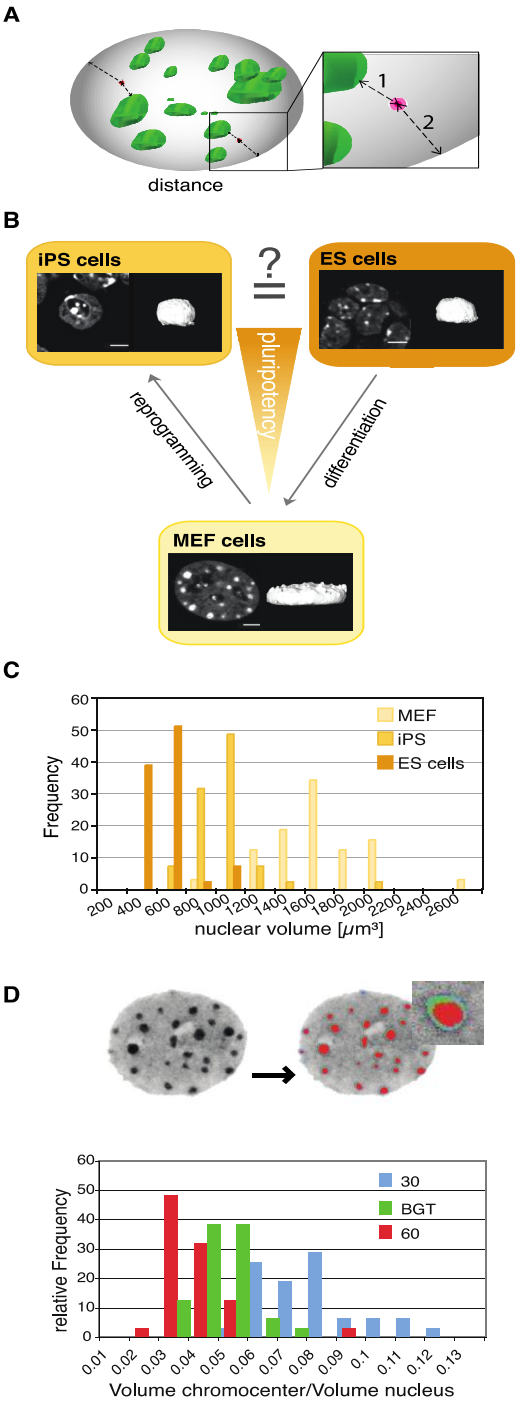


Figure 14. Correlation of subnuclear gene topology versus gene expression is hampered by drastic morphological variability. A 3D representation of a mouse mammalian nucleus (gray) with chromocenters highlighted in green (DNA staining with DAPI) and gene loci visualized by FISH in red. Nearest 3D distances are measured from the signal of interest (gene locus) to the chromocenter (1) and to the nuclear periphery (2). **B** Schematic representation of the pluripotency state of MEF, iPS and ES cells (ES). Confocal optical mid section of a DAPI stained representative MEF (lower), iPS (left) and ES (right) cell nucleus and corresponding 3D reconstructions highlighting their morphological variability. Scale bar 5mm. **C** Bar histograms of nuclear volumes measured in MEFs, iPS and ES cells (n=30). **D** The ratios between chromocenter versus nuclear volumes were compared using either BGT depicted in green, or fluorescent intensity threshold levels of 30 (blue) and 60 (red), respectively, and their relative frequency plotted. The largely diverging volumes demonstrate the dependency of these measurements from a given threshold setting (N = 133).

Distance measurements and threshold independent single cell-based normalization

Distance measurements were performed using the gene signal channel (Cy5) and the binary segmented image from the DNA (DAPI) channel to calculate the shortest distance in 3D from the signal to the nearest chromocenter surface and towards the nuclear periphery. The nonnormalized distance plots are shown in Figure S6. The direct comparison of absolute measurements and distance analyses was complicated by the drastic changes of nuclear shape and size concomitant with the process of cellular reprogramming. Primary fibroblasts, commonly used as the source for reprogramming, have flat ellipsoid nuclei with length diameters that

significantly exceed those of the much smaller spherically shaped ES and iPS nuclei, respectively, but have much smaller z-diameters (Figure 14A and B). Flat and spherically shaped nuclei thus yield shorter distance distributions to the periphery exclusively due to their shape. To solve this problem, we established a single cell-based normalization. This enabled the determination of changes of gene positioning in relation to gene expression and to elucidate similarities or differences between the three cell types examined irrespective of their diverging nuclear morphologies. We utilized a Monte Carlo approach: for each cell the software created 10 000 random points within the nucleus in 3D (Supplementary Movie 2) and calculated the distance of each point to the nearest relevant structure resulting in a cell-specific reference distribution (Figure 15A). The number of simulations was determined by balancing the requirement for speed and data reproducibility (Figure S7). Since chromocenters are known to harbor very few genes, we further adapted the simulation to exclude chromocenters from our reference distribution. As this additional step did not alter the outcome, we neglected it in our further evaluation (Figure S8).

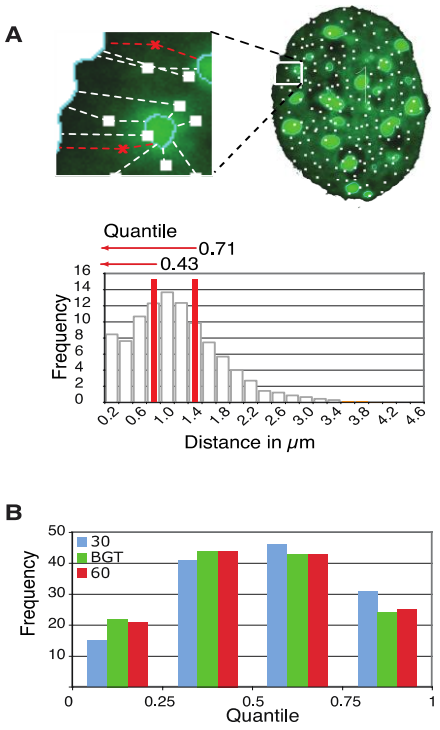


Figure 15. Single cell-based normalization overcomes threshold dependent variability. A For data normalization, 10,000 random points were set throughout the 3D nuclear volume (Supplementary Movie 2) and all distances from these points (white) to the chosen nuclear landmarks measured. The target gene locus measurement (red) is set in relation to the random distribution obtained by the simulation. The fractions of random point measurements, which are smaller or equal to the gene locus distance measurements, are defined as quantile. **B** Gene loci distance measurements in MEF were normalized to a random distribution and the resulting quantiles plotted for each threshold setting (N = 133)

Combining this reference distribution with our non-normalized gene to heterochromatin distance measurement, we obtained a quantile. This is defined as the fraction of reference distances smaller or equal to the non-normalized measurement and can take any number between zero and one. A quantile of zero represents a real measurement, which is shorter than all randomly distributed points, and a quantile of one means that the real measurement is more distal than all simulated points. By this, all measured distances are set into the context of their individual cell (Figure 15A) and give an unbiased view of any deviation from a random distribution.

Using quantiles has several major advantages. First, each quantile between zero and one is equiprobable in value if measurements are sampled from the random background distribution. A deviation from an equal distribution (each quantile bin in a frequency of 25%) indicates a non-random positioning. Deviations from random positioning can be determined by Pearson's chi-square test. Second, they are independent from morphological variations and distributions. Last but not least, a major advantage lies in threshold robustness (Figure 15B), a challenge for quantitative image analysis (Ronneberger *et al.*, 2008). The same cells were analyzed using two fixed thresholds (30 and 60) and our calculated basic global threshold for segmentation. The differences in frequency of quantiles are minimal (Figure 15B), which emphasizes the superiority of this normalization method.

When assembling quantiles obtained from numerous nuclei in a frequency distribution, one can expect three types of distribution: (i) no preference for any position within the nucleus results in an uniform distribution of quantiles (Figure S9A); (ii) an accumulation of small quantiles would represent short distances (Figure S9B), while (iii) accumulation of larger quantiles would indicate large distances to the respective reference nuclear landmark (Figure S9C). The comparison of quantile distributions in data sets from different cell types allows the disclosure of differences in gene positioning (Figure S9D–F).

Internal nuclear gene positioning is correlated with expression and gene relocation is restricted by high gene density

Using our normalization, we compared the gene positions within the three different cell types. In Figure 16A, we show the quantile distribution as relative frequencies, i.e. data points represent all quantiles in the intervals 0–0.25, 0.25–0.5, 0.5–0.75 and 0.75–1.

We observed only two cases of a random subnuclear distribution for the pluripotency genes using our randomizer tool. In MEF cells, Sox2 positioning did not differ from a random pattern with regard to the nuclear periphery and Oct4 relative to the distance to the nearest chromocenter (Figure 16A, double dagger). All other distributions differed highly ($P<10^4$ two-tailed t-test) from randomness.

Sox2 and Oct4 did not change their distance to chromocenters between cell types but Nanog showed reduced distances in iPS cells (Figure 16A). It is not clear why Nanog exhibits different positioning in iPS cells but might be influenced by partial reprogramming of the cells. We did not observe a general change of distance to chromocenters indicating that they may not act as silencing compartments as proposed before. (Brown *et al.*, 1997; Harmon & Sedat, 2005).

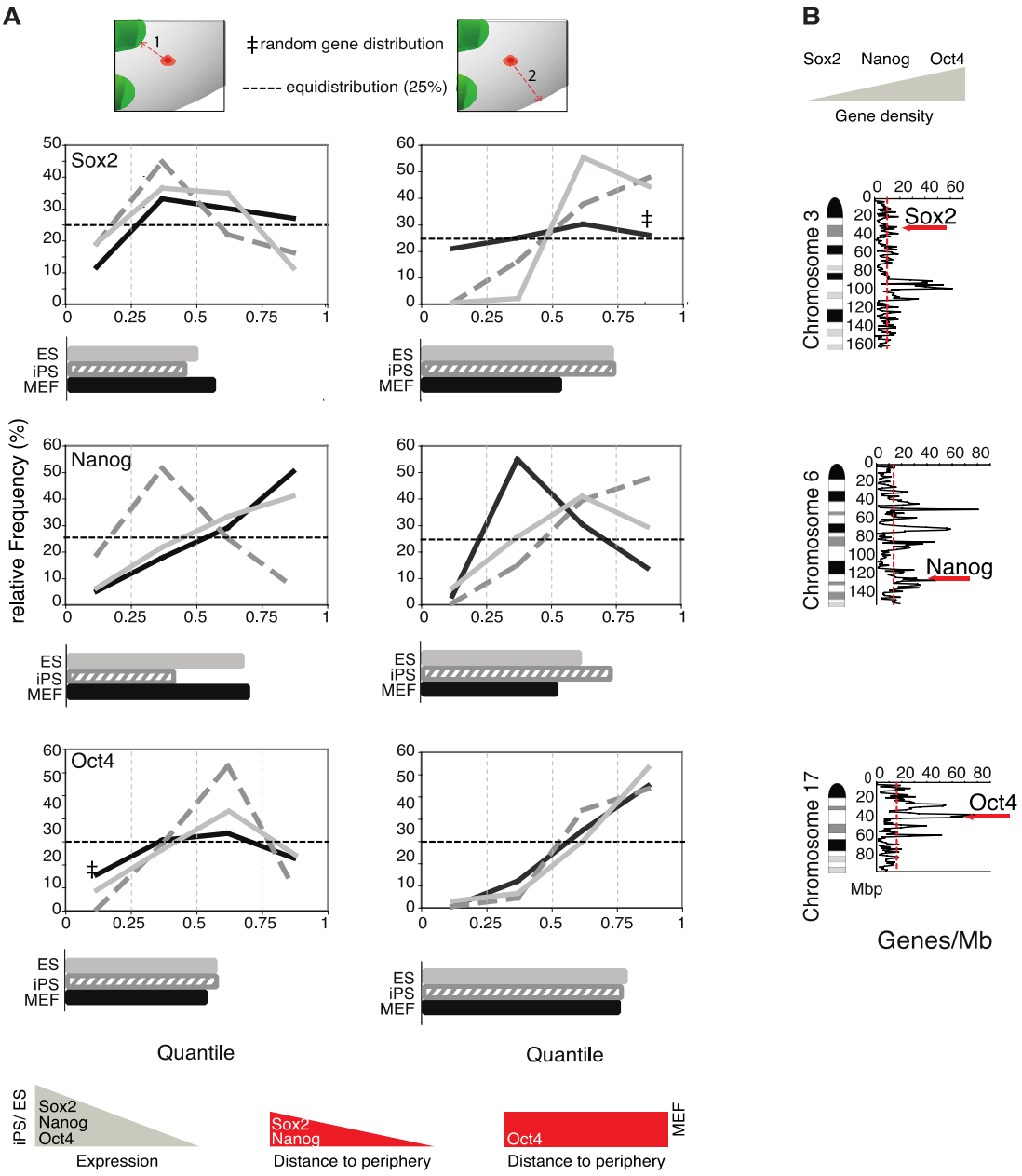


Figure 16. Gene repositioning during gain/loss of pluripotency relative to chromosomal context and to distance from heterochromatin. Distances of pluripotency genes Sox2, Nanog and Oct4 to chromocenters and to the nuclear periphery in MEF, ES and iPS cells were normalized as shown in Figure 2A. **A** Quantile distribution relative to chromocenters (1) and to nuclear periphery (2) with histogram of the average quantiles underneath. A summary of expression and re-localization of the genes in the nucleus is added at the bottom. **B** Position of the genes within the corresponding mouse chromosome ideogram (Giemsa banding) with gene density plotted along the chromosome.

Nanog and Sox2 exhibited larger quantiles, reflecting larger distances from the nuclear periphery in iPS and ES cells where they are actively transcribed, compared to MEFs where these genes are silent (Figure 16A). In contrast, no relocation of Oct4 was detected when comparing its positioning in MEFs and iPS/ES (Figure 16A). In all three cell types, Oct4 occupied an internal nuclear position. This observation is consistent with previous data on

human and mouse ES cells and a few differentiated cells (Bartova *et al.*, 2008; Hepperger *et al.*, 2008; Wiblin *et al.*, 2005), where an internal Oct4 position is maintained. Taken together these results do not support the hypothesis that the nuclear periphery acts as a general transcriptional silencing compartment (Reddy *et al.*, 2008) since in MEFs no relocation to the periphery is observed. An alternative hypothesis would be that an internal nuclear position is compatible with gene expression although not sufficient for it.

All three genes show a preferential internal nuclear localization in iPS and ES cells but differ in MEFs, which could reflect their chromosomal context. Oct4 is located within a chromosomal region of particularly high gene density (75 genes/Mb), whereas Nanog resides in a region of intermediate gene density (33 genes/Mb) and Sox2 in a region of low gene density (six genes/Mb) (Figure 16B). The very high gene density around Oct4 might restrict its relocation. The interior positioning in the active state (Figure 16A) may facilitate spatial interactions of the extended transcription network of these genes (Loh *et al.*, 2006).

Although there are many similarities between the three genes regarding their nuclear (re)localization, there are also differences even when comparing iPS versus ES cells. One possible explanation for these differences, could be the inclusion of iPS cells in our data sets that were only partially reprogrammed although SSEA-1 positive.

Importantly, this computational analysis tool can be used for unbiased analysis of any topological change occurring in cells during cell cycle, differentiation or other physiological processes. Accompanying changes are taken into account as any observed data sets are tested against computer generated random distributions. This approach thus allows a rigorous test for functional topological relationships in cells and tissues.

MATERIALS AND METHODS

Biological system and microscopy

Cells and culture conditions.

Transgenic MEF cells were kindly provided by R. Jaenisch (Wernig *et al.*, 2008) (Whitehead Institute, Cambridge, MA, USA). These cells had been infected with lentiviral vectors containing the four reprogramming factors Oct4, Klf-4, Sox2 and c-Myc expressed in a doxycycline (DOX)-inducible manner (Brambrink *et al.*, 2008). Non-induced cells were used as controls. MEF cells were grown in cell culture flasks in Dulbecco's Modified Eagle's Medium (DMEM) GlutaMAX (Invitrogen) supplemented with 10% fetal calf serum (FCS) (PAA Laboratories) and seeded on gelatine-coated coverslips 36–48 h before fixation. Reprogramming towards iPS was achieved by adding DOX (2 ng/ml) (Sigma-Aldrich) to the medium for ~4 weeks. After induction cells were cultured in DMEM GlutaMAX supplemented with 15% FCS, leukemia inhibitor factor (LIF) (1000 U/ml; Chemicon/ Millipore), non-essential amino acids (0.1mM; PAA Laboratories) and b-mercaptoethanol (1 mM) (Merck). Distinct colonies of cells were observed after a few weeks. Reprogramming on a single cell level was assessed by positive immunofluorescence (IF) staining against the murine pluripotency marker SSEA-1 (mouse IgM anti SSEA-1, Millipore).

Mouse embryonic stem (ES) cells of the CCE line (129/ Sv-derived mouse ES cell line) (kindly provided by C. Bonifer, Leeds, UK) were cultured without feeder cells on gelatinized glass slides under the same culture conditions as described for iPS cells. Medium was changed daily and cells were split before individual colonies touched each other. Pluripotency was tested on a single cell level by positive staining for the SSEA-1 marker as for iPS cells (Figure S3).

DNA probes. Specific bacterial artificial chromosome (BAC) clones were used for the delineation of the genomic regions covering the pluripotency transcription factors Oct4 (MMU 17; clone RP23-75C13), Sox2 (MMU 3; RP23-425G5) and Nanog (MMU 6; RP23-474F18). The DNA probes were labeled with biotin-coupled dUTPs by nick-translation and labeled DNA probes were suspended at a final concentration of 50 ng/ml together with a 20-fold excess of unlabeled mouse COT-1 DNA (Invitrogen) in 50% formamide/ 10% dextransulfate/2xSSC (saline-sodium citrate buffer) following standard protocols. Hybridization solely of the endogenous gene loci was ensured by specific visualization of the addressed genes on the expected chromosomal position.

3D immuno-FISH analysis. Cells were fixed with 4% paraformaldehyde for 10 min. Subsequent treatment included permeabilization with 0.5% Triton X-100/PBS for 30 min at room temperature (RT), incubation in 20% glycerol/PBS for 30 min, repeated freezing/thawing steps in liquid nitrogen, protein degradation with 0.1N HCl and storage in 50% formamide/2xSSC at 4 °C over night (Cremer *et al.*, 2008). This approach allows the preservation of morphological structures of single nuclei as well as of whole iPS colonies. Probes and cell samples were allowed to pre-anneal for 2–3 h at RT and thereafter denatured simultaneously for 3 min at 76°C. Hybridization was performed in a 37°C water bath for 3 days. Stringent washing was

done in 0.1xSSC at 62°C for 3-5 min followed by incubation in 4% BSA/PBST (bovine serum albumin in PBS with 0.05% Tween) blocking solution for 10 min. Hybridized probes were detected with Cy5-conjugated streptavidin (1:200 in 2% BSA/PBST) (Rockland) together with immunodetection of SSEA-1 using an anti-SSEA-1 mouse monoclonal antibody (IgM) for 1 h at RT and subsequent incubation with Alexa594-conjugated goat anti-mouse IgM antibody (1:500) (Sigma-Aldrich). DNA counterstain was performed with 40,6-diamidino-2-phenylindole (DAPI) (200 ng/ml) (Sigma-Aldrich) for 10 min and samples were mounted in Vectashield Antifade Medium (Vector Laboratories).

Image acquisition. Only SSEA-1-positive cells were used for image acquisition. 3D image stacks of single nuclei were acquired using a Leica SP5 confocal laser scanning microscope (Leica Microsystems) with a Plan-Achromat 63x oil objective (voxel size 50:50:200nm [x:y:z]).

Computational analysis

User interface. The user interface enables the easy adjustment of various settings to respond to different image qualities (Figure S4) and step-by-step audio visual tutorial Supplementary Movie 1.

Vid: viewer identification determines the active window.

DNA col.channel: define which channel contains the DNA counterstain from which the chromocenters and nuclei are detected.

BAC col.channel: define which channel contains the gene or other point-like signals from which distances should be measured.

Nuclear_threshold: depending on the fluorescence image the threshold for the nuclear periphery can be adjusted, the output can be simply controlled visually.

minNucPixelSize: in order to filter out false signals a minimal size for a nucleus can be set.

fill nuc holes: DNA counterstain does not stain nucleoli which results in very dark nucleoli which may be wrongly segmented as periphery. Using this setting, small holes within the nucleus, resulting from incorrect segmentation can be filled.

BAC minPixelsize: this setting defines the smallest size of the reference structure for visualization.

Baseline: using this setting the minimal fluorescent intensity of the wanted signal is set. By adjustments, larger weakly stained unwanted particles can be filtered out.

nSim (number of simulations): sets the number of simulated points in the nucleus. Depending on the size of the nucleus this number might need to be adjusted

nBins (number of bins): defines how many bins should be used for the output graph.

nucID (nucleus identification): If more than one nucleus is present per image, this identifies the nucleus used for further analysis. The number of the relevant nucleus can be entered in this field.

z-min overlap: values larger than 0 allow a certain degree of z overlap to occur without merging two separate nuclei.

Batch process multiple files for CC (chromocenter) segmentation: this button opens up a window to perform preprocessing on the raw data image files. The program automatically detects similar file names in order to carry out batch processing. The median and radius value for Gaussian blur can be set at this point. If 8 bit images are used, the 'fix saturation value' box with a saturation value of 255 should be kept to avoid problems due to overexposure in the image. The batch processing can be started to apply all filters to all selected files resulting in new binary (3D) image files containing the segmented nuclei and chromocenters (Figure S5).

Load CC and show CC outlines: load preprocessed chromocenter segmentation and show chromocenter outline overlaid to the microscopic image. Additionally the gene signal centroid is marked.

Filter and label nuclei: the raw data are processed to segment nuclei, which are outlined and numbered on the viewer window.

Draw nuclei cut lines: if nuclei are very close to each other automated segmentation might fail. This button enables one to manually draw a separating line between two adjacent structures.

Paint cut: this function enables one to draw into the image without changing the raw data. This can be used in order to mask false signals that are disturbing correct segmentation. If manual segmentation has been performed it can be documented and recalled by the save and load button.

Calc volumes: The volumes of segmented chromocenters and nuclei are printed.

Simulate 3D points: Performs the measurement and automated simulation of the random points. Output can be directly pasted into excel and contains the real distance to chromocenters and nuclear periphery and normalized quantile values.

Mark 3D sect: Visualizes the random points in each section.

Parameter settings for measurements

For the segmentation of the nucleus, a constant threshold of 40 was used across almost all data sets. Only for very light and dark images, a small deviation was required. The quality of segmentation was visually controlled.

For the segmentation of the chromocenters, a series of filtering steps (Figure S5) was performed followed by basic global thresholding (BGT) algorithm. These settings were kept constant throughout all data sets. The settings for gene signal detection (BAC minPixelSize and baseline) were adjusted individually for each image depending on background noise and signal intensities. The settings varied between 40 and 70 for the pixel size and between 40 and 100 for the baseline. If segmentation was altered by manual correction, the altered image was stored to keep it for control. To account for chromatic aberration, xyz correction (20nmx50nmx950 nm)

based on bead measurements was performed. This is especially important for dyes with large spectral distances such as DAPI and Cy5, which were used in our experiments.

Image data processing, measurements and normalization

Our analysis software is geared towards optimized throughput when analyzing multiple 3D data sets. The program is implemented using the Priithon image analysis platform (http://priithon.googlecode.com), which allows fast interactive visualization and comparison of the non-normalized data versus the resulting segmentation shown as outlines.

First, the image processing and 'chromocenter' segmentation steps illustrated in Figure S5 are applied to a group of files. An unsharp mask type filter is applied using the same user-specified parameters (a 4x4 pixel median filter and a 6x6 uniform filter) for all data sets. To get a robust measure for thresholding chromocenters, the volume is projected along Z using the maximum intensity method before applying the BGT algorithm (Gonzalez & Woods, 2008). Here, the threshold is iteratively calculated from the intensity histogram so that the image is split into bright and dark pixels in such a way that their respective mean intensity is symmetrically far from the chosen threshold.

Nuclei are segmented in a similar way to chromocenters. First, a 3x3 pixel median filter and a 5x5 uniform filter are averaged and applied to each section of the volume. After this smoothing step, the data were thresholded using a user-specified constant threshold as described in the parameter settings. Then, we perform a 2D hole filling step using a specified radius to solve the problem posed by nucleoli or other unstained regions within the nucleus, followed by segmentation and enumeration of 3D-connected structures, in this case nuclei. Fully automated and unsupervised segmentation of nuclei often fails especially in the case of cell colonies, which are densely packed. For this case, our program provides the possibility to manually draw multiple cut-lines. These are section-wise spline-interpolated polygons (Bezier curves), which can be saved and reloaded to make these changes reproducible and documentable. The cut-lines assist the segmentation algorithm by forcing a larger area to be split into two or more nuclei. The user can furthermore specify a fractional parameter determining to what degree overlap is allowed.

Gene loci are calculated based on the fluorescence intensity distribution of the respective channel. A user-specified base intensity is subtracted to account for background signal and electronic amplification offsets. Pixels with higher intensities are weighted and used for the centroid position calculation with sub-pixel resolution.

The distance analysis is done on a nucleus-by-nucleus basis. Random points are generated uniformly within the 3D nucleus. To improve performance for distance measurements, distance maps are pre-calculated for the chromocenter and for the nuclear envelope distances, respectively. This is done using the Euclidean Distance Transform based on the 3D-grid given by the raw data voxel size.

The acquired data were collected and documented using Excel. All distances given in micrometer from one gene and one cell type were binned in steps of 0.1 mm and their relative frequency calculated. Quantile normalized distances were binned in 0.25 steps. Graphs were created by plotting the relative frequency as y-axis and the bins as x-axis. To test for divergence from a random distribution (relative frequency of 25% for each bin) the chi-squared value was calculated using Excel.

REFERENCES

Bartova, E., Krejci, J., Harnicarova, A. and Kozubek, S. (2008) Differentiation of human embryonic stem cells induces condensation of chromosome territories and formation of heterochromatin protein 1 foci. *Differentiation*, **76**, 24-32.

Brambrink, T., Foreman, R., Welstead, G.G., Lengner, C.J., Wernig, M., Suh, H. and Jaenisch, R. (2008) Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell*, **2**, 151-159.

Brown, K.E., Guest, S.S., Smale, S.T., Hahm, K., Merkenschlager, M. and Fisher, A.G. (1997) Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell*, **91**, 845-854.

Cremer, M., Grasser, F., Lanctot, C., Muller, S., Neusser, M., Zinner, R., Solovei, I. and Cremer, T. (2008) Multicolor 3D fluorescence in situ hybridization for imaging interphase chromosomes. *Methods Mol Biol*, **463**, 205-239.

Deniaud, E. and Bickmore, W.A. (2009) Transcription and the nuclear periphery: edge of darkness? *Curr Opin Genet Dev*, **19**, 187-191.

Gonzalez, R.C. and Woods, R.E. (2008) *Digital image processing*. 3rd ed. Prentice Hall, Upper Saddle River, N.J.

Gue, M., Messaoudi, C., Sun, J.S. and Boudier, T. (2005) Smart 3D-FISH: automation of distance analysis in nuclei of interphase cells by image processing. *Cytometry A*, **67**, 18-26.

Harmon, B. and Sedat, J. (2005) Cell-by-cell dissection of gene expression and chromosomal interactions reveals consequences of nuclear reorganization. *PLoS Biol*, **3**, e67.

Hepperger, C., Mannes, A., Merz, J., Peters, J. and Dietzel, S. (2008) Three-dimensional positioning of genes in mouse cell nuclei. *Chromosoma*, **117**, 535-551.

Huangfu, D., Osafune, K., Maehr, R., Guo, W., Eijkelenboom, A., Chen, S., Muhlestein, W. and Melton, D.A. (2008) Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. *Nat Biotechnol*, **26**, 1269-1275.

Joffe, B., Leonhardt, H. and Solovei, I. (2010) Differentiation and large scale spatial organization of the genome. *Curr Opin Genet Dev*, **20**, 562-569.

Lanctot, C., Cheutin, T., Cremer, M., Cavalli, G. and Cremer, T. (2007) Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat Rev Genet*, **8**, 104-115.

Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J. *et al.* (2006) The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet*, **38**, 431-440.

Meaburn, K.J., Misteli, T. and Soutoglou, E. (2007) Spatial genome organization in the formation of chromosomal translocations. *Semin Cancer Biol*, **17**, 80-90.

Reddy, K.L., Zullo, J.M., Bertolino, E. and Singh, H. (2008) Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature*, **452**, 243-247.

Ronneberger, O., Baddeley, D., Scheipl, F., Verveer, P.J., Burkhardt, H., Cremer, C., Fahrmeir, L., Cremer, T. and Joffe, B. (2008) Spatial quantitative analysis of fluorescently labeled nuclear structures: problems, methods, pitfalls. *Chromosome Res*, **16**, 523-562.

Shirley, J.W., Ty, S., Takebayashi, S., Liu, X. and Gilbert, D.M. (2011) FISH Finder: a high-throughput tool for analyzing FISH images. *Bioinformatics*, **27**, 933-938.

Silva, J., Chambers, I., Pollard, S. and Smith, A. (2006) Nanog promotes transfer of pluripotency after cell fusion. *Nature*, **441**, 997-1001.

Wernig, M., Lengner, C.J., Hanna, J., Lodato, M.A., Steine, E., Foreman, R., Staerk, J., Markoulaki, S. and Jaenisch, R. (2008) A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nat Biotechnol*, **26**, 916-924.

Wiblin, A.E., Cui, W., Clark, A.J. and Bickmore, W.A. (2005) Distinctive nuclear organisation of centromeres and regions involved in pluripotency in human embryonic stem cells. *J Cell Sci*, **118**, 3861-3868.

SUPPLEMENTARY INFORMATION

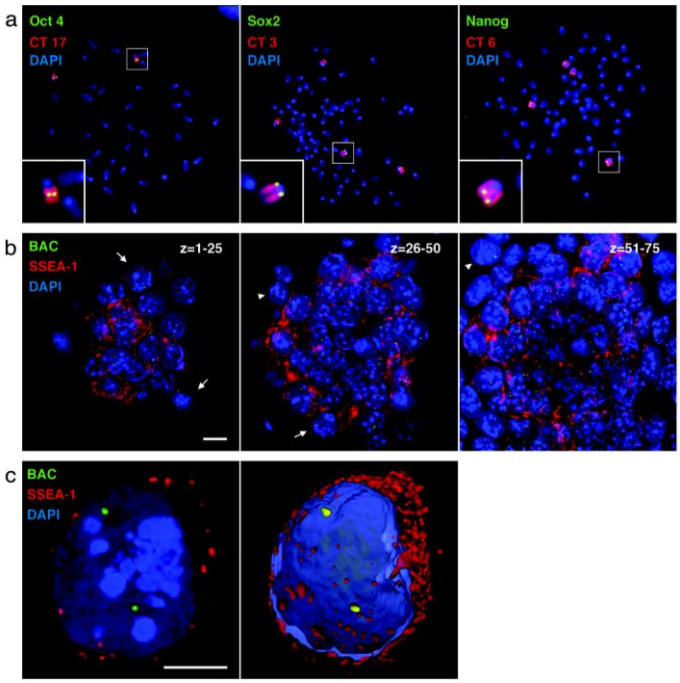


Figure S3. Probe specificity control and iPS selection strategy. (**a**) Hybridization of BACs containing *Oct4* (left), *Sox2* (middle) or *Nanog* (right) together with the respective chromosome paints on metaphase spreads of MEF cells stably transfected with lentiviral vectors containing *Oct4*, *Klf-4*, *Sox2*, and *c-Myc* show distinct signals only at the expected genomic sites. (**b**) Three subsequent blocks of z-projections (optical serial sections 1-25, 26-50, 51-75) of a typical iPS cell colony. Arrows exemplify SSEA-1 positive cells that were chosen for further analysis. Arrowheads show SSEA-1 negative cells that were excluded from analysis. Scale bar indicates 10 µm. (**c**) Example of a post-processed cell, which has been cut out of its neighboring cells (left). Right image shows a 3D reconstruction of the same cell. Scale bars 5 µm.

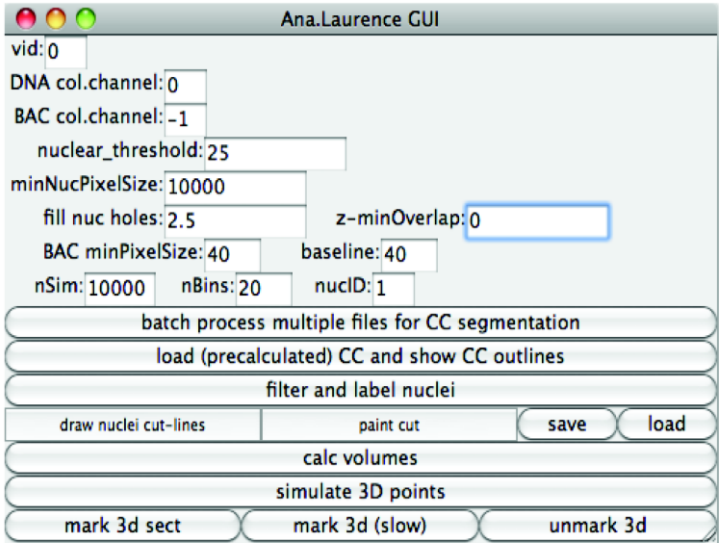


Figure S4. User interface of randomizer software. The user interface window of the randomizer software provides many options to adjust parameters to meet the needs of individual experiments (detailed information in methods and step-by-step tutorial Movie 1).

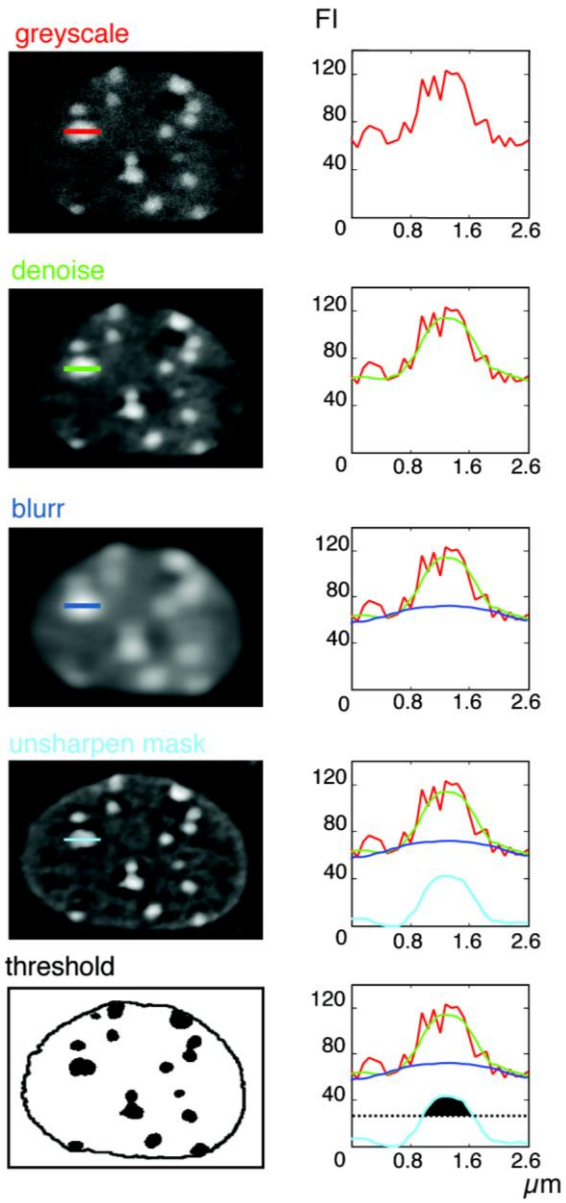


Figure S5. 3D image segmentation procedure. Starting with a grayscale image stack of DAPI stained mouse nuclei (left column) image processing routine included denoise and blurring as a prerequisite for an unsharpen mask calculation. Basic global thresholding results in a 3D binary picture of chromocenter structures (detailed information in methods). The effect of each filtering step is highlighted exemplarily over a line scan across a chromocenter by overlaying the respective fluorescent intensity (FI) profile with the ones of the previous steps (right column).

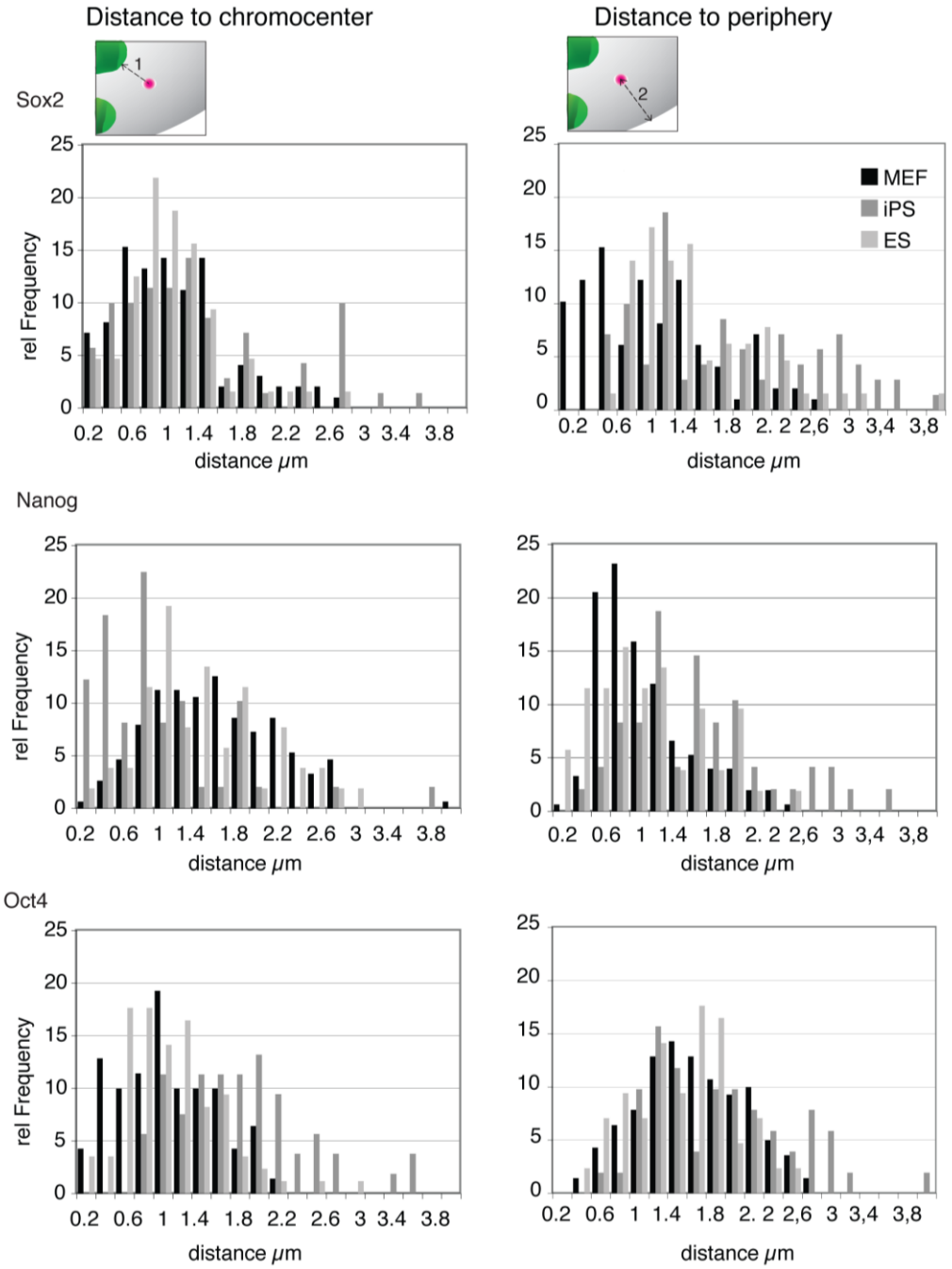


Figure S6. Non-normalized distance distribution for *Sox2*, *Nanog* and *Oct4* relative to the chromocenters and periphery. The graphs represent non-normalized distances of the genes to the nuclear landmarks in MEF (black), iPS (dark gray) and ES (light gray) cells. (N=52-151)

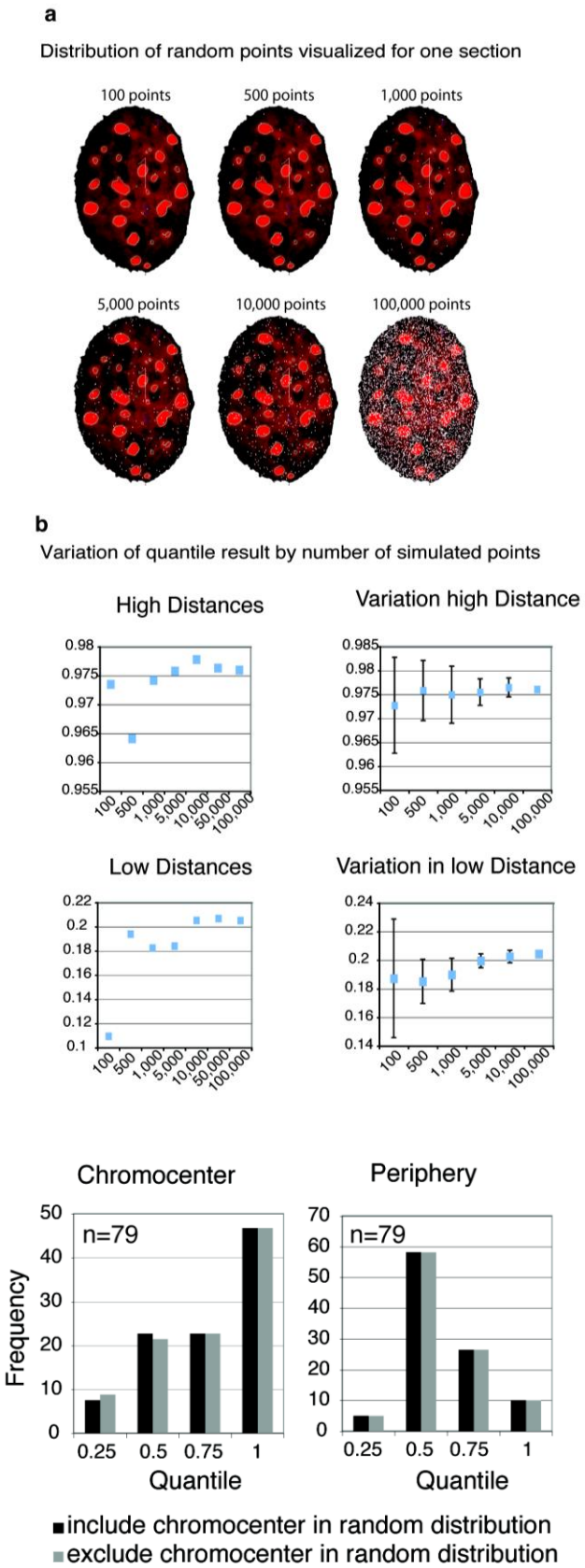


Figure S7. Effect of number of simulated points on quantile variation. (**a**) Visualization of random distribution using increasing numbers of simulated points. For simplicity only one z-section is shown. See Movie 2 for full z-stack with 10,000 simulated points. (**b**) Calculating the quantile for a small and a large distance measurement with different numbers of simulated points for the random distribution. A plateau is seen at 10,000. Repeating these measurements 30 times and plotting the average quantile with standard deviation shows that the error with 10,000 simulated points is minimal.

Figure S8. Impact of chromocenter inclusion/exclusion on data normalization. Simulations were performed excluding and including random points from the chromocenter volume and normalized data compared. Including/excluding chromocenter volumes for random positioning does not affect normalized data.

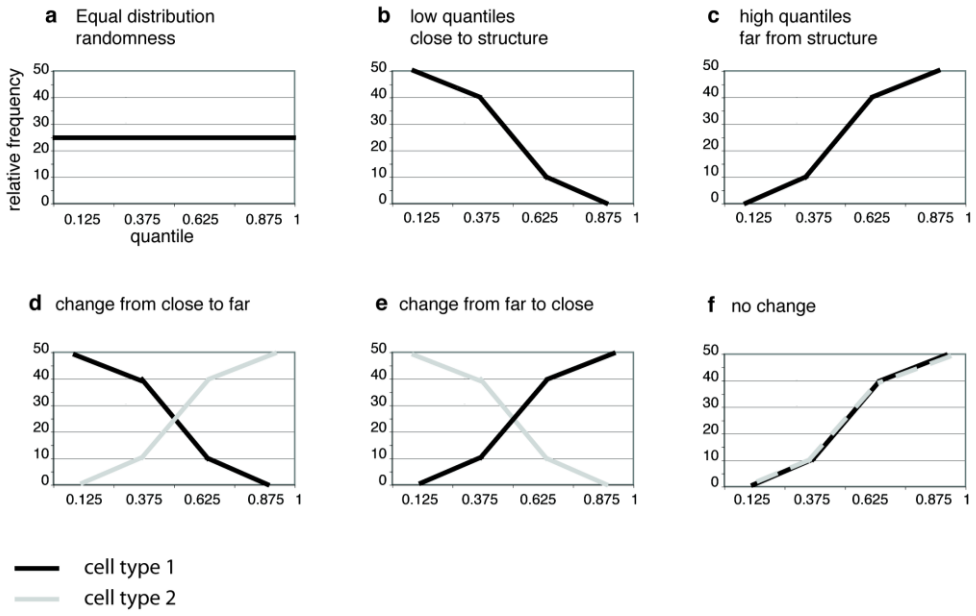


Figure S9. Examples of different quantile distributions. (**a**) Example of a gene showing no position preference within the nucleus. All quantiles are represented in the same proportion. (**b**) Example of a gene with high preference for proximity to the measured nuclear landmark. The majority of quantiles are in the lower quart. (**c**) Example of a gene with aversion to the measured nuclear landmark. The majority of quantiles are in the upper quart. (**d**) Example of increased distances of a gene in cell type 2 compared to cell type 1. (**e**) Example of decreased distances of a gene in cell type 2 compared to cell type 1. (**f**) Example of no positional change compared to the measured structure. In both situations the gene presents a distal position.

Movie 1. Randomizer software step-by-step tutorial.

Movie 2. Random points simulation in 3D. Visualization of the 10,000 random points simulated for each individual nucleus in 3D. DNA is false colored in red, the nuclear periphery in white and chromocenter borders in light blue. Random points are marked in white. The two yellow circles represent the gene signal whose centroid position is marked by a cross.

Chapter 3: Nuclear topology and gene expression

Role of heterochromatin in gene silencing

The data presented in this chapter is prepared for publication.

The results presented in this chapter are the work of multiple people and not only myself. In order to facilitate comprehension of the data, it is presented as one work.

The contributions of all people are listed below and acknowledged by them.

K. Laurence Jost: Designed the project, performed *in silico* analysis of affymetrix data, performed the FISH and Immuno-FISH experiments, imaged and analyzed the data, prepared most of the figures and wrote the manuscript
Bianca Bertulat: Helped with image acquisition and figure preparation
Alessandro Brero: Sample preparation for gene expression profiling
Tanja Hardt: Sample preparation for gene expression profiling and BAC DNA extraction
Claudia Gösele: RNA and Affymetrix screen
Herbert Schulz: Statistical analysis of Affymetrix screen
Norbert Hübner: Affymetrix screen
M. Cristina Cardoso: Designed the project and supervised the manuscript writing

Figures contributions:

Figure 17B (left): Bianca Bertulat
Figure 17A, B (right), C - 25 and supplement figures: Designed and put together by K. Laurence Jost

INTRODUCTION

In recent years nuclear topology has come into focus as an epigenetic regulator of gene expression with heterochromatin as a main player. The first line of evidence for heterochromatin as a silencing compartment were the position effect variegation experiments in *Drosophila* by Mueller in 1930 (Muller, 1930). In these experiments, silencing of genes that were randomly relocated next to heterochromatin of the same chromosome could be observed. Further experiments could show that this effect is not only observable for genes on the same chromosome (*cis*) but also for genes on different chromosomes (trans). Shown in mouse (Brown *et al.*, 1997; Delaire *et al.*, 2004), *Drosophila* (Harmon & Sedat, 2005), yeast (Simmer *et al.*, 2010) and even in plasmodium (Duraisingh *et al.*, 2005). Heterochromatin can be found in essentially all eukaryotes but its distribution and composition differs from species to species. Most commonly, heterochromatin can be found lining the lamina at the inside of the nucleus. This position has been well documented by electron microscopy in different species including murine and human embryonic stem cells (Baharvand & Matthaei, 2003; Zeuschner *et al.*, 2010). Additionally, some species form clusters of constitutive heterochromatin from multiple chromosomes, called chromocenters which are found in mouse (Brero *et al.*, 2005), plants (Baccarini, 1908) and *Drosophila* (Harmon & Sedat, 2005). Both forms of heterochromatin (chromocentric and peripheral) have been hypothesized to act as silencing compartments. Brown *et al.* documented repositioning of inactive genes to chromocenters during mouse lymphocyte maturation and again relocation away from chromocenters for activated loci suggesting a role of heterochromatin on gene expression (Brown *et al.*, 1997). Heterochromatin association was scored by colocalization of genes to chromocenters visualizing only a small time frame of chromatin mobility. A later study (Harmon & Sedat, 2005) performed distance measurements from the gene locus to the chromocenters in *Drosophila* and thereby took the possibility of chromatin movement into account. Several other studies have shown a correlation between gene activity and distance to chromocenters (Brown *et al.*, 1997; Delaire *et al.*, 2004; Sabbattini *et al.*, 2001) as well as to the nuclear periphery (Meaburn & Misteli, 2008; Szczerbal *et al.*, 2009; Takizawa *et al.*, 2008). At the same time, other studies did not always observe relocation to chromocenters (Takizawa *et al.*, 2008) or to the nuclear periphery (Meaburn & Misteli, 2008). Additionally, genes tethered to the nuclear lamina or into chromocenters are not always silenced (Finlan *et al.*, 2008; Reddy *et al.*, 2008; Sabbattini *et al.*, 2001). Altogether, these studies do not allow drawing a general conclusion about the relation of heterochromatin proximity and gene expression.

A general challenge of these measurements is the changing morphology of cells and their nuclei. Hematopoietic cells are usually spherical whereas adherent cells are flat and ellipsoid. This shape difference strongly influences distances to the periphery. In addition, the structure of chromocenters is also remodeled in many cell types e.g. neuronal progenitor cells (Meshorer *et al.*, 2006), mouse muscle cells (Brero *et al.*, 2005; Mayer *et al.*, 2005), human embryonic stem cells (Park *et al.*, 2004) and seems to be a common feature of differentiation. This reorganization during differentiation hints to a functional role of heterochromatin.

To elucidate the importance of heterochromatin architecture on gene expression in an unbiased way, we need to perform distance measurements uninfluenced by nuclear morphology. To overcome limitations due to morphological differences, we used the single cell based normalization described in chapter 2. As first model, we used the mouse myogenesis as established cellular differentiation system. During myogenesis nuclei change their morphology from large and flat (myoblasts) to small and spherical (myotubes). This change is accompanied by clustering of many small chromocenters observed in myoblasts to fewer and larger chromocenters seen in myotubes (Brero *et al.*, 2005), which makes it an interesting target to study the effect of chromatin topology. Brero *et al*. could also demonstrate that nuclear architecture can be changed to the differentiated state by expression of the methyl CpG binding protein 2 (MeCP2) (Brero *et al.*, 2005) in a dose dependent manner. MeCP2 is thought to act as a transcriptional silencer (Nan *et al.*, 1997) and plays an important role in the neurological disease Rett syndrome (RTT, OMIM 312750). This gives us the opportunity to study not only relocation during differentiation but gives us an additional model system in which we can induce chromatin remodeling as observed during differentiation by transiently expressing only one protein.

RESULTS AND DISCUSSION

Rational and experimental design

To investigate the influence of chromatin architecture on genome expression, we used two different cellular systems. The first system consisted of adult muscle stem cell differentiation *in vitro* using undifferentiated mouse myoblasts compared with differentiated myotubes. The heterochromatin architecture undergoes large-scale remodeling with many small chromocenters coalescing into bigger chromocenters both during myoblast differentiation. The second system relied on the transient transfection of mouse myoblasts with GFP-tagged MeCP2 followed by FACS sorting of low and high expressing cells from the same transfection. Since Pmi28 myoblasts contain low or undetectable levels of endogenous MeCP2 (Brero *et al.*, 2005) and undergo the same large-scale chromatin remodeling in a dose dependent manner upon MeCP2 expression (Brero *et al.*, 2005) this was an ideal system to investigate gene positioning dependent on chromatin reorganization. Overexpression of MeCP2 enables the study of the effects of chromatin remodeling similar to those observed during differentiation but without the additional complication of a complex gene expression program being executed in the background. Both systems were analyzed for their expression profiles (Figure 17A).

To ensure maximal reproducibility and high quality of data the experiments were performed in quintuplicates. The results from the transcriptional profiling were analyzed for quality, and candidate genes for our study were chosen based on their statistical significance and expression level. Candidate genes were then visualized by 3D fluorescence *in situ* hybridization (FISH) (Cremer *et al.*, 2007) and their distance to the nearest heterochromatin was measured (Figure 17B). Heterochromatic regions studied consisted of chromocenters and the nuclear periphery which is enriched in heterochromatin. As association with both facultative and constitutive heterochromatin has been correlated with gene silencing, we measured the distances of each gene locus to both compartments (Figure 17B). The distances were normalized in a cell-by-cell procedure to eliminate artifacts from nuclear morphology and thresholding errors (See chapter 2). Correlations of raw and normalized distances with expression levels and other factors were then investigated.

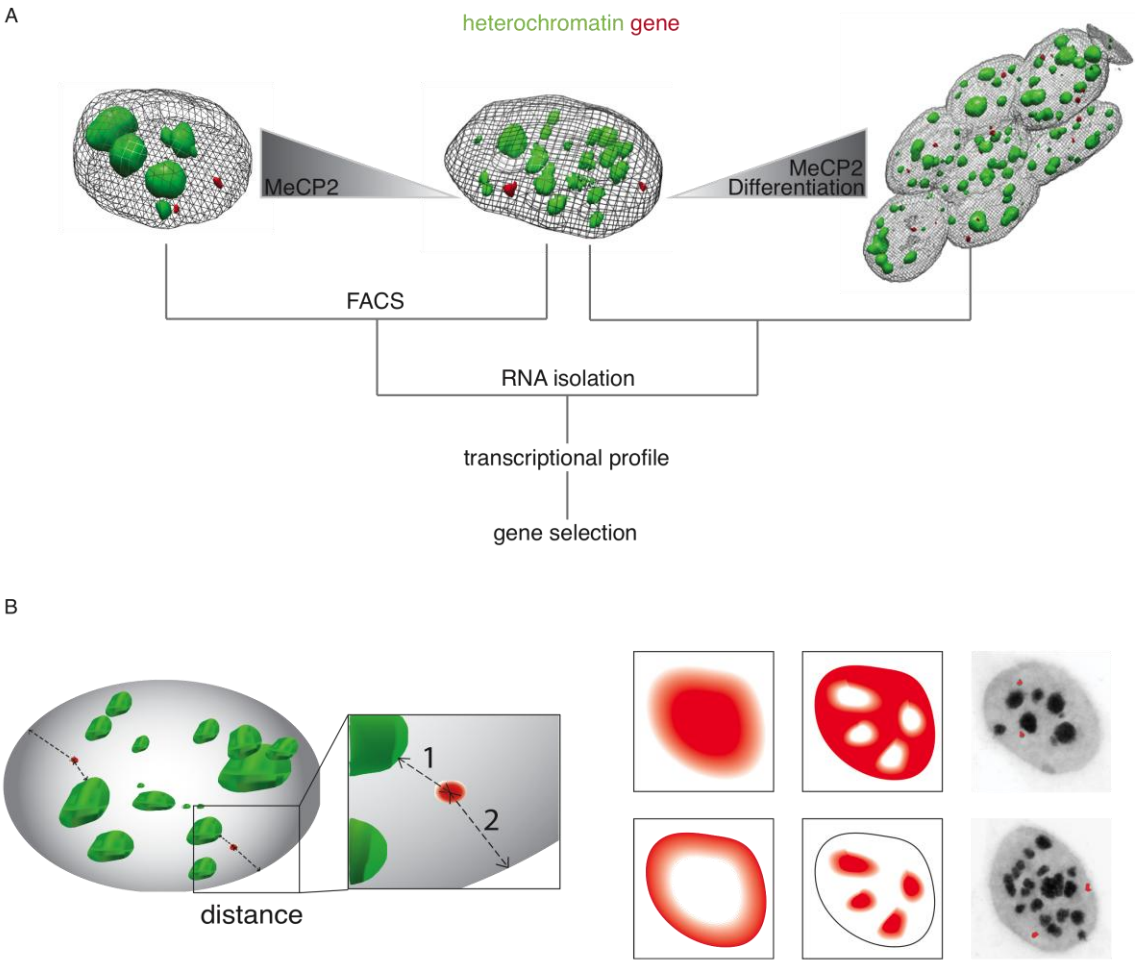


Figure 17. Experimental design and measurements. A Representation of differentiation and MECP2 system used and experiments leading to gene selection. **B** Gene distances to chromocenters (1) and the nuclear periphery (2) were measured to establish if genes would position themselves close or far to heterochromatic regions. The first column on the right represents gene position to the periphery, the second column relates gene position to chromocenters. The last column shows microscopic images representing those different gene positions**.**

Genome wide transcriptional profiling

MePC2-GFP expressing cells were sorted into high and low expressing populations using FACS. RNA was then prepared from all four conditions (myoblasts/myotubes, low/high MeCP2 expressing cells) and used for cDNA synthesis. The cDNA was hybridized to the Affymetrix mouse 430 2.0 microarray and for each set five independent samples were analyzed, maximizing quality assurance and reproducibility. Overall the different hybridizations of each conditions showed a high reproducibility between themselves and a very high data quality assured by: 3'/5' signal ratio of GAPDH and β-Actin, chip background and noise (RawQ): proportion and average expression value of detected genes. As expected the highest expression difference in low vs. high MeCP2 expressing cells was MeCP2 itself with a fold change of 11 (red diamond in Figure 18B). Additionally, we found a very high expression of known muscle markers and low expression of proliferation markers upon myogenesis (Figure S10), further validating the quality of the biological samples and experimental condition. No

array had to be rejected according to statistical analysis (Table S1). The normalized expression data was then used to determine suitable genes as determined by heir statistical significance (p-value (<0.05) by ANNOVA testing) and their fold change. To visualize the expression profiling result p-values and fold changes were plotted for myoblasts vs. myotubes and high vs. low MeCP2 expression in myoblasts. In this so called volcano plot (Figure 18) it is easy to detect genes with a high mathematical and biological significance reflected by the p-value and the fold change, respectively.

A striking feature of the MeCP2 volcano plot (Figure 18B) is the obvious imbalance between the numbers of up and down regulated genes. Since MeCP2 contains a transcriptional repressor domain, which interacts with histone deacetylases of the transcriptional corepressor SIN 3A complex (Jones *et al*., 1998) it was mostly regarded as a general repressor of transcription. However, more recent studies could show that MeC2 also acts as an activator of transcription (Chahrour *et al.*, 2008; Lasalle & Yasui, 2009). The role of MeCP2 as an activator is supported by our expression profiling data which showed more genes up than down regulated in cells expressing high MeCP2 levels. Additionally, it can be observed that the change in gene expression is generally low compared to those during differentiation. This observation is in accordance with previous expression data in MeCP2 deficient/mutated mouse and human brain (Colantuoni *et al.*, 2001; Tudor *et al.*, 2002), lymphocytes from patients (Delgado *et al.*, 2006; Ballestar *et al.*, 2005) which in general showed small changes. In contrast to the obvious imbalance of up and down regulated genes in the MeCP2 expressing cells, the expression profile of the myogenesis data shows almost equal numbers of up and down regulated genes. The diamonds marked in blue represent genes that were selected for our analysis of distance to heterochromatin. In our selection, we included biological (fold change) and mathematical (p-value) significant up and down regulated genes as well as a set of genes whose expression was unchanged as controls.

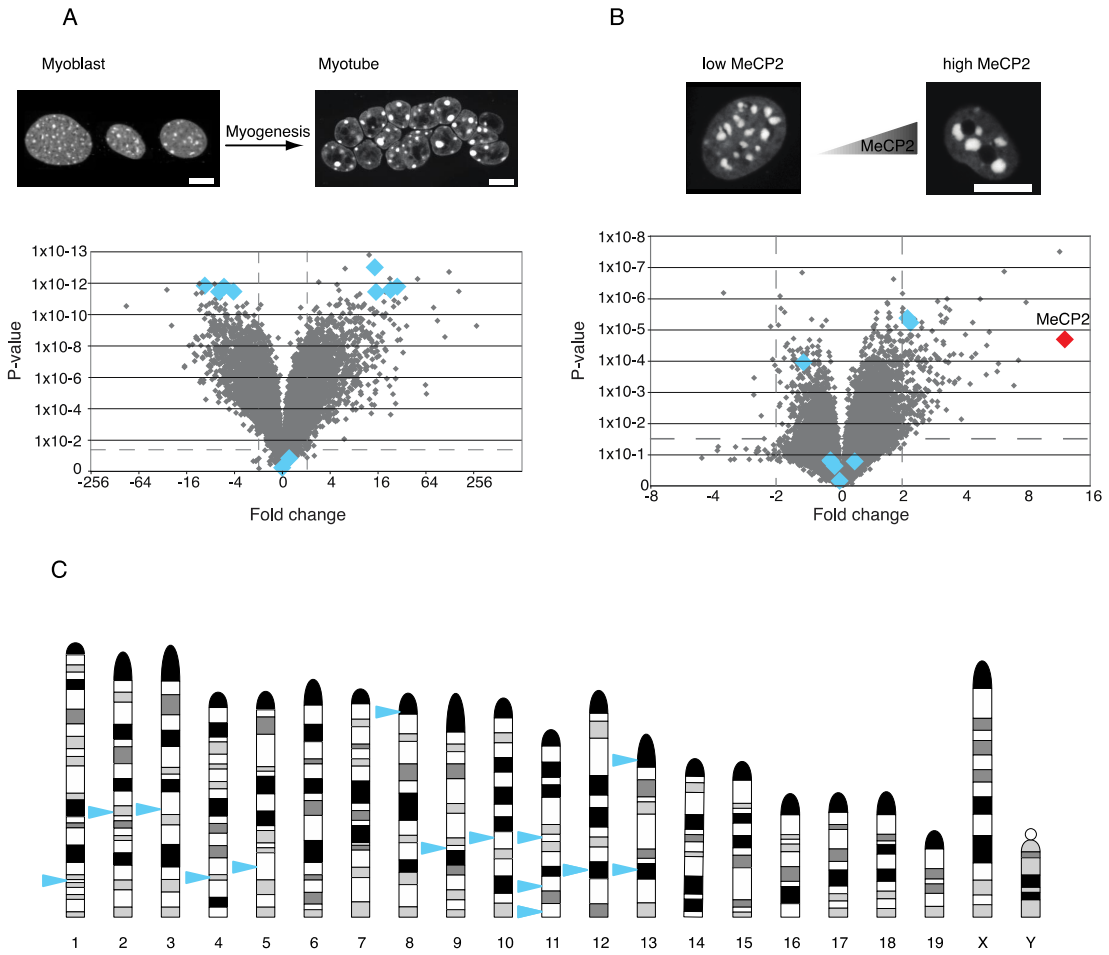


Figure 18. Volcano plots from gene expression data. The fold changes are represented by the x-axis and the p-value by the y-axis. The dashed line represents a threshold of p=0.05 and a fold change of 2. Blue diamonds represent genes chosen for our distance measurements. **A** Expression data from myogenesis. **B** Expression data from low and high MeCP2 expressing cells. The red diamond in the lower plot represents MeCP2. To better display the data points the axes are shown at different scales in A and B. Scale bar 5 µm. **C** Chromosomal location of all genes studied indicated on Giemsa banded mouse chromosomes.

In situ gene position analysis and normalization

To analyze gene position relative to heterochromatin we next generated biotin tagged DNA probes of each of the selected gene loci. These were made by nick translation using BAC clones covering the respective genomic loci (see methods section). Each labeled DNA probe was then hybridized *in situ* and detected by fluorescently labeled streptavidin. Heterochromatin (chromocenters as well as peripheral) was detected by counterstaining with the DNA dye DAPI. One of the most important aspects of the experiments was to maintain the 3D structure of the nucleus as well as possible during (Immuno) FISH. As standard FISH procedures tend to flatten nuclei, we adapted the FISH and Immuno FISH protocol developed and extensively tested by Cremer *et al.* (Cremer *et al.*, 2007) for our experiments. The success of this method can be appreciated by comparing 3D reconstructions of myoblast and myotube nuclei (Figure 19) after

the procedure. The nuclei are not flattened and the roundness of myotube nuclei can be appreciated.

To identify MeCP2-GFP expressing cells an immunofluorescence detection of MeCP2 was required since the GFP fluorescence is destroyed during the FISH procedure. For this combined (Immuno) FISH, the newly developed anti MeCP2 antibodies described in chapter 1 were used, followed by the FISH gene detection. Successful gene detection was evaluated by signal to noise ratio and the number of signals within the cell. The expected number of signals was two since Pmi28 cells are diploid. More than 2 signals were either a sign of divergence from the diploid state or reflected a cell in G2-phase; in both cases the cell was removed from the analysis. Each cell was completely imaged using a confocal microscope to allow full 3D reconstruction for distance measurements. In all measurements, the absolute and normalized distances were recorded.

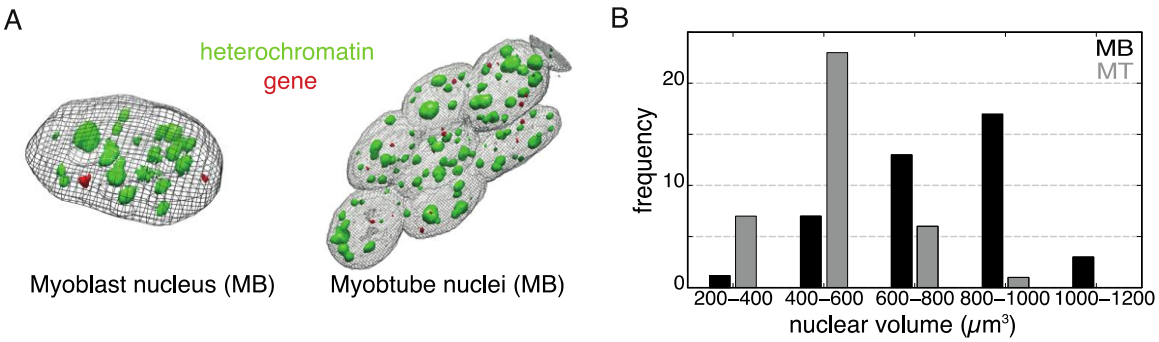


Figure 19. Volume differences between myoblasts and myotubes. A 3D reconstructions of myoblast and myotube nuclei. Each nuclei shows two gene signals in red detected by FISH. Heterochromatin is indicated in green. **B** The nuclear volume of >30 nuclei from myoblasts and myotubes was evaluated. In black myoblasts (MB) and in gray myotubes (MT), showing their volume difference.

Since chromatin architecture changes during differentiation and dependent on MeCP2 levels the question arose whether the observed reduced number of chromocenters (clustering) was accompanied by a chromatin compaction. To address this question, chromocenter volumes were measured in a cell cycle independent manner (excluding cells with doubled DNA volume) (Easwaran *et al.*, 2004). We transfected myoblasts with PCNA-Cherry (Proliferating cell nuclear antigen) as a marker for S-phase and pulse labeled the cells for 30min with BrdU followed by 4 hours chase. After fixation we could assess if cells were in S-phase at the moment of fixation using PCNA (Leonhardt *et al.*, 2000). G1 and G2 cells could be separated by BrdU incorporation 4 hours prior to fixation. Analyzing chromocenter volumes we could determine no differences between cells with low MeCP2 transfection vs. cells with high transfection. Therefore, a compaction of chromatin cannot be supported based on this result (Figure S11).

Normalization of the data is a crucial step when comparing datasets from different experimental conditions. However, it needs to be assured that normalization does not change the data to an unrealistic degree. To control for normalization artifacts, we examined to which extent raw and normalized measurements correlate with each other. A clear correlation between the raw and

the normalized data set with the corresponding Pearson correlation coefficient for the differentiation (Figure 20) and MeCP2 datasets (Figure S12) could be observed. The correlation coefficient lies in a range of 0.4-0.88. An outlier can be observed in Figure 20 in the upper left. Excluding this outlier, the correlation goes up to 0.82. A perfect correlation of 1 would mean identical results and would show that normalization is not necessary. A correlation coefficient of -1 would mean that the results are always contradicting each other, indicating a bad normalization process. Therefore, we can assume that the normalization procedure is necessary for our measurements but does not change our data to an unrealistic extent. Additionally, the correlation graphs show that the distance changes in µm (x-axes) observed during differentiation and depending on MeCP2 level lie in the same range (-0.4- 1.0 µm). Chromatin mobility is still discussed but is generally expected to be around 0.3-0.5 µm which is the range we observed in our study (Hubner & Spector, 2010).

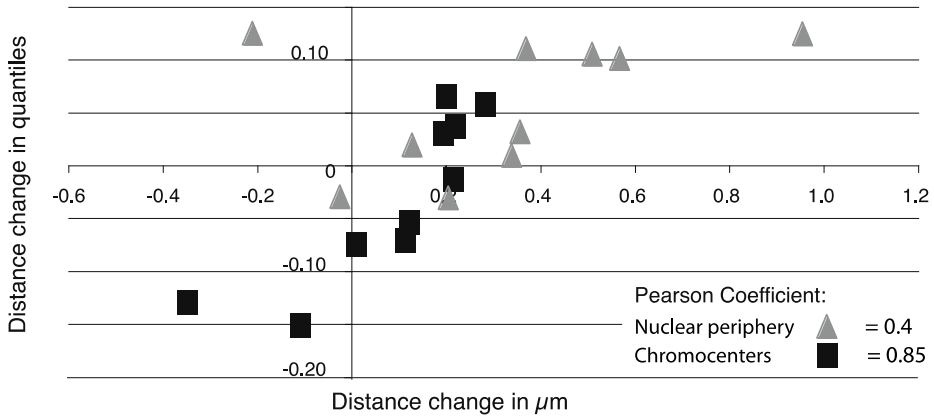


Figure 20. Raw vs. normalized data during differentiation. Distance changes in µm were plotted against normalized distances changes in quantiles. Triangles represent measurements to the nuclear periphery and squares to chromocenters. The Pearson correlation coefficient is indicated below the plots.

Outcome of gene position relative to heterochromatin during differentiation

Gen – Heterochromatin distance and genomic context

Ten genes were selected for our distance measurements during differentiation (see Table 2). These ten genes include myogenesis specific genes (myocyte enhancer factor 2C (Mef2c), cytoskeletal calmodulin and titin-interacting RhoGEF (Obscn), tropomyosin 3 gamma (Tpm3) and myomesin 2 (Myom2)) as well as genes that are not specifically related to myogenesis (Ttk protein kinase (Ttk), cDNA sequence BC002230 (BC002230), baculoviral IAP repeat-containing 5 (Birc5), breast cancer 1 (Brca1) and solute carrier family 19 (thiamine transporter) member 2 (Slc19a2)). Those 10 genes were studied in undifferentiated myoblasts as well as in differentiated myotubes for their distance to the closest chromocenter surface as well as to the nuclear periphery. For each gene and condition, over 30 cells were analyzed. The average absolute and normalized distances are listed in Table 2. In addition, the change of distance during differentiation is noted.

Table 2 All distance measurements in myogenesis[1]

	Chromocenter						Periphery					
	Distance in µm			Normalized distance			Distance in µm			Normalized distance		
	MB	MT	Delta	MB	MT	Delta	MB	MT	Delta	MB	MT	Delta
Mef2c	0.94	1.16	0.22	0.55	0.58	0.04	1.10	1.23	0.13	0.64	0.66	0.02
Tpm3	1.08	1.21	0.12	0.67	0.62	-0.05	1.09	1.07	-0.02	0.66	0.63	-0.03
Myom2	0.97	0.98	0.01	0.58	0.50	-0.07	0.74	0.95	0.21	0.54	0.51	-0.03
Obscn	1.04	1.24	0.20	0.54	0.61	0.07	1.25	1.61	0.36	0.69	0.72	0.03
Slc19a2	1.23	1.35	0.12	0.72	0.65	-0.07	0.91	1.87	0.96	0.63	0.75	0.12
BC002230	1.38	1.04	-0.34	0.72	0.60	-0.13	0.96	1.33	0.37	0.63	0.74	0.11
Brca1	1.00	1.22	0.22	0.67	0.65	-0.01	1.10	1.67	0.57	0.74	0.84	0.10
Coro1c	1.19	1.38	0.20	0.63	0.66	0.03	1.35	1.14	-0.21	0.67	0.80	0.13
Ttk	1.11	1.01	-0.11	0.67	0.51	-0.15	1.07	1.59	0.51	0.58	0.68	0.11
Birc5	0.97	1.25	0.29	0.60	0.66	0.06	1.53	1.88	0.34	0.76	0.77	0.01

From the ten observed genes, only two had a negative distance change (distances in µm) for chromocenters and only one for the nuclear periphery. This reflects morphological differences (less, more clustered chromocenters and smaller, round nuclei in myotubes) observed in myoblasts vs. myotubes, which promote larger distances *per se* in myotubes. In general, all averaged absolute distances are clustered around 1 µm ranging from 0.74-1.88 µm. In contrast, the normalized distances show a different pattern especially in relation to the distances to chromocenters. As observed in the Pearson correlation coefficient, the distances in µm and the normalized distances do not always perfectly correlate (Figure 20). Differences between normalized and absolute values can be seen in the cases of Brca1, Tpm3, Myom2 and Slc19a2. The other six genes behave similar in both cases. Comparing the distances to the periphery, only Myom2 and Coro1c show a divergent pattern between absolute and normalized distance.

Next, we correlated distances within myoblasts and myotubes with the genomic context to assess the influence of gene neighborhood on gene position. Additionally, we analyzed distance changes during differentiation for correlation with gene expression levels and genomic context. The genomic properties considered in this study were gene density (number of genes in the neighborhood), number of CpG Islands, GC content (percent of GC in the neighborhood), the number of short interspersed nuclear elements (SINE), long interspersed elements (LINE) and the number of CpG Islands within the gene. The genomic properties were obtained from Ensembl Genome Browser (http://www.ensembl.org) and by the program RepeatMasker (http://www.repeat masker.org, version open-3.3.0). The genomic properties are collected in Table 3, separating the direct neighborhood (2-Mbp window) from the wider (5-Mbp window) neighborhood.

Table 3 Genes analyzed during differentiation and their genomic context

Mef2c	Tpm3	Myom2	Obscn	Slc19a2	BC002230	Brca1	Coro1c	Ttk	Birc5

[1] Distances are averaged over all measurements of one gene in one condition.

	Gene expression	27.66	22.77	14.97	14.44	1.22	-1.01	-4.13	-5.44	-6.26	-9.48
	CpG Islands in Gene	0	1	2	2	1	0	1	2	1	1
2-Mbp window	CpG Island	10	80	18	66	23	30	67	38	20	54
	GC Content	37.81	45.85	44.08	46.64	41.60	45.19	48.06	50.55	40.10	51.18
	Gene activity	2.35	1.05	2.87	1.29	-1.13	-1.17	1.11	1,00	-1.25	-1.14
	Gene density	9	39	7	63	14	13	50	25	7	23
	SINE	3.22	15.88	3.54	11.45	6.35	10.27	21.70	12.04	5.19	11.87
	LINE	37.77	15.23	13.31	8.62	15.04	10.47	2.45	2.09	21.10	2.06
5-Mbp window	CpG Island	24	147	58	117	63	63	127	99	35	162
	GC Content	38.34	44.13	44.15	46.35	41.84	43.72	47.08	49.80	39.05	50.7
	Gene density	32	186	47	157	82	60	221	105	29	153
	Gene activity	-1.01	1	1.34	1.11	-1.06	1.06	1.01	1.05	-1.17	-1.01
	SINE	3.38	13.38	4.97	12.46	7.14	8.86	17.57	12.27	4.45	14.78
	LINE	31.27	19.02	11.16	8.95	15.86	16.85	4.82	3.21	25.54	1.88

As can be seen in Table 3, four of the genes that we investigated are significantly down regulated during differentiation. This is expected for Birc5, Brca1, Ttk and Coro 1c since they are not directly relevant for myogenesis. The four genes that are significantly up regulated are all involved in myogenesis. Two genes (BC002230, Slc19a2) not related to myogenesis and without expression change were observed and taken as control. The surrounding gene density varies greatly in all observed cases, ranging from 7-63 genes in a 2-Mbp window and 29-221 genes in a 5-Mbp window. The average gene activity of genes in the neighborhood was calculated using the gene expression profiling data. The expression data however do not cover all genes in the region, therefore the gene activity may not necessarily reflect the gene activity of the complete neighborhood. An example of how such an activity profile looks like is given in Figure 21. From the 29 genes in the 5-Mbp neighborhood of the Ttk protein kinase, 15 were not identified by our expression profile (gray squares). In the distribution of gene activities in Figure 21 one can see that an equally strong up regulated gene lies right next to the gene of interest which might directly influence positional changes of Ttk.

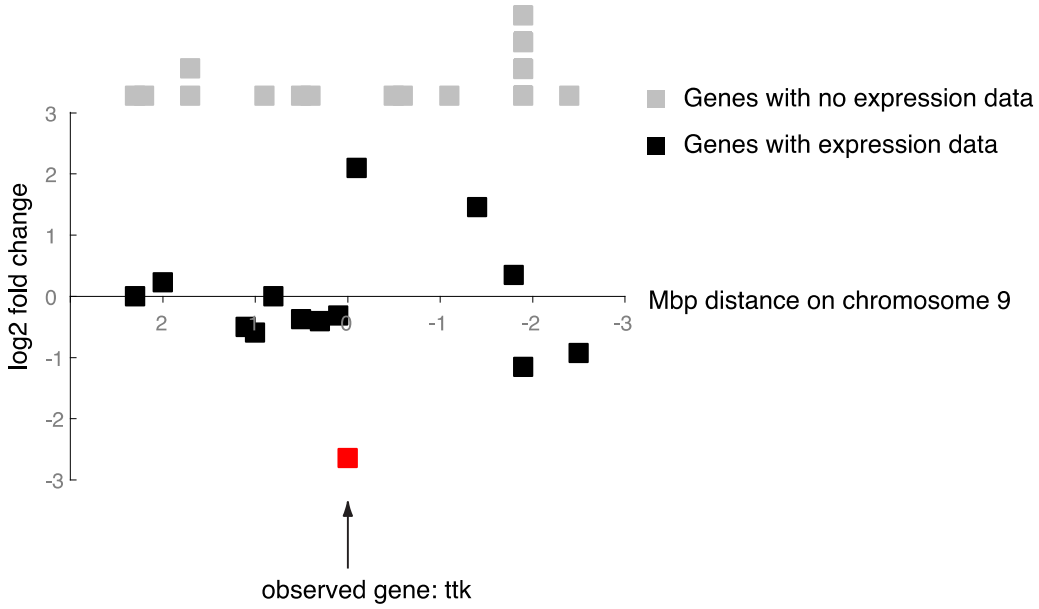


Figure 21. Gene activity in 5-Mbp neighborhood of the Ttk protein kinase gene. The red square indicates the observed Ttk protein kinase gene. Gray squares indicate genes not observed in the expression profiling. The x-axis shows 2.5 Mbp up and downstream from the gene center.

CpG islands are defined as regions of 500 bp minimum length, 50% or higher GC content and a ratio of 0.60 or higher observed CpG / expected CpG (Takai & Jones, 2002). They show a distribution ranging from 10 – 80 between the genes in a 2-Mbp neighborhood. The neighborhood around the gene Mef2c is very GC poor with only 37% of GC, whereas Birc5 has the highest percentage (51%). The average density lays around 45% which is expected. SINE and LINES represent percentage of the sequence occupied by them and have a range between 2.06 – 37.7%.

Correlation between gene – heterochromatin distances and genomic context

As mentioned before, data on gene-heterochromatin association and gene expression is controversial. In a first step, we wanted to assess whether there is a correlation between gene position in myoblasts and myotubes with the respective genomic context rather than looking at the gene movement during differentiation. The rationale behind this idea is, that it might be possible that genes prefer certain subnuclear localizations in only one state and are randomly localized in the other, which would not be properly reflected by gene position change. To study the correlation between the different parameters, we calculated the Pearson correlation coefficients between the distances listed in Table 2 and the different parameters describing the genomic context of the genes listed in Table 3. The correlations to normalized distances in a 2-Mbp neighborhood are depicted in Figure 22. Since the 2-Mbp and 5-Mbp neighborhood did not display large differences, we neglected them in our graphic correlations as well as the absolute distances. All correlations can be found in Table S2.

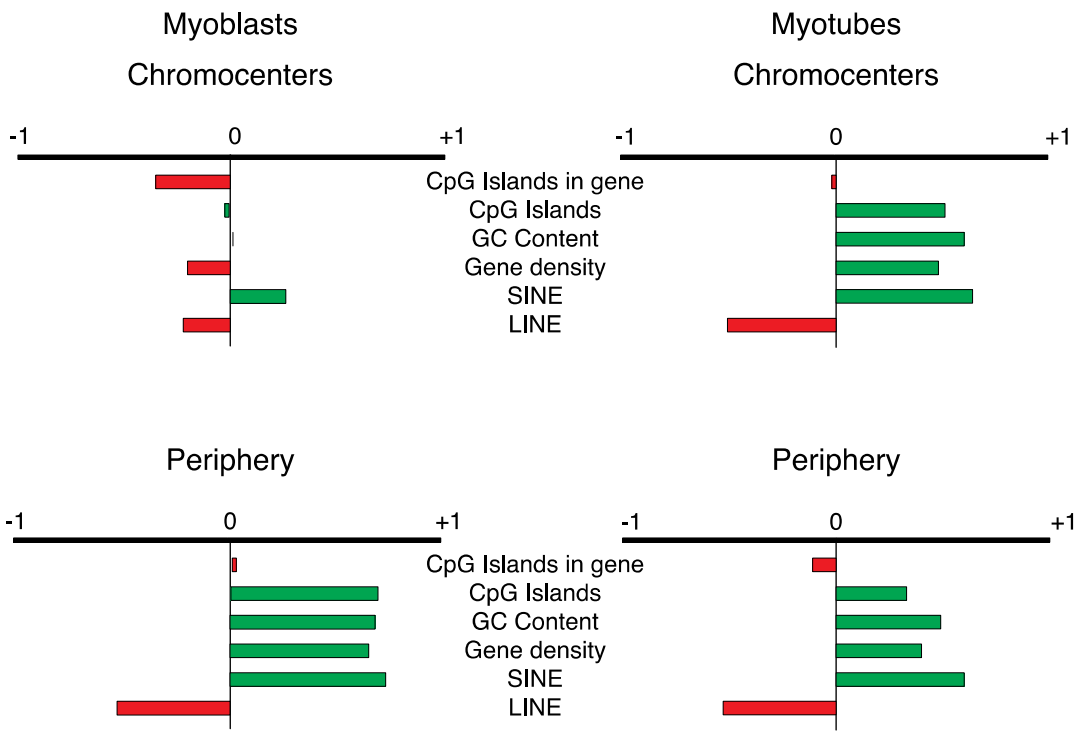


Figure 22. Correlation between gene - heterochromatin distances and genomic context. Graphic representation of Pearson correlation coefficients in differentiation between normalized gene distance and genomic context in a 2-Mbp neighborhood (except for CpG Islands in gene). One represents a perfect correlation, zero no correlation and minus one an anti-correlation. (N>30 cells)

Comparing distances and not changes in distances, no correlation with gene expression can be drawn since the latter reflect a change in expression and not absolute expression levels. Our data show no correlation of genomic properties with the gene distances to chromocenter in myoblasts, neither absolute nor normalized, to chromocenters in myoblasts. By contrast, normalized distances to the nuclear periphery correlate well with gene density, CpG islands, GC content and SINEs and show a high anti-correlation with LINEs whereas the correlation of the same parameters to absolute distances are much weaker.

In myotubes, we can detect medium to strong correlations (0.39-73) between gene positions and genomic context. We can observe positive correlations of absolute and normalized distances to gene density, CpG Islands, GC content and SINE numbers and a negative correlation for LINE number. No correlation can be observed in µm distances in myotubes. The positive correlation indicates that higher distances to heterochromatin are accompanied by a higher density of genes, CpGs, and SINE and a higher GC content.

In human cells, it has been shown that gene density might play a role in radial gene positioning within the nucleus (Kupper *et al.*, 2007; Murmann *et al.*, 2005). This positioning effect might be due to gene density dependent sorting of chromosome territories, which has been observed in myoblasts and myotubes (Mayer *et al.*, 2005). Our positive correlation data of distance to periphery and gene density contradicts observations made in a study in mouse ES cells and macrophages (Hepperger *et al.*, 2008). However, the study by Hepperger *et al.* does not

measure distances but radial positioning which does not take morphology into consideration. Since our normalization is independent of nuclear size and morphology we are able to observe correlations that are otherwise missed, as can bee seen by comparing the absolute distances to the periphery in myotubes and the normalized distances (Table S2). The more genes are in close vicinity the higher the chance that there will be genes that are actively expressed and prefer a certain positioning. A higher gene density might also simply decrease the flexibility of free movement for a gene. The effect is seen in the immediate (2-Mbp) and also wider neighborhood (5-Mbp), arguing for a larger influence of gene density.

CpG islands represent parts of the genome with a CpG occurrence that is significantly higher than in the rest of the genome (Gardiner-Garden & Frommer, 1987). CpG islands also have a high GC content and are usually unmethylated. In the vertebrate genome around 70% of all gene promoters are associated with CpG islands (Saxonov *et al.*, 2006). Therefore, high numbers of CpG islands are an indicator for active gene transcription and might correlate with higher distances to potentially repressive compartments such as chromocenters and the nuclear periphery. Since CpG islands and GC content are connected we can observe the same effect. Interestingly, this correlation is weaker or nonexistent when analyzing absolute distances with GC content. This again shows the influence of a cell-by-cell based normalization.

The last set of genomic properties we considered was the density of SINEs and LINEs. Our data show that SINEs and LINES have opposing influences on gene localization. For SINEs, we can observe positive or no correlation, with higher correlation for distances to the periphery than to chromocenters. LINES however exhibit negative correlation to gene positioning. Already in the sixties some scientists speculated that noncoding RNAs might regulate gene transcription (Britten & Davidson, 1969; Jacob & Monod, 1961). More recent studies have found more and more evidence for a functional role of noncoding RNA, which arise from heterochromatin (Chueh *et al.*, 2009). SINEs represent non coding RNAs and do not encode a functional reverse transcriptase and are often found within satellite DNA and not in coding regions (Singer, 1982). In contrast, LINEs code for a functional transcriptase and are thought to be actively transcribed, potentially playing a role in gene silencing (Chueh *et al.*, 2009). In the human genome, regions of increased gene expression have been described (RIDGEs) (Caron *et al.*, 2001) which are characterized by high gene density, high GC content, high SINE numbers and low LINE numbers. Regions with low gene expression are characterized by the opposite features and called anti-RIDGEs. Correlation of gene position to GC content, SINE and LINE could reflect whether the gene lies within a RIDGE or an anti-RIDGE and therefore reflect the transcriptional status of their surroundings. Our data show that a high SINE content correlates with high distances from silencing compartments whereas high LINE content correlates with short distances to silencing compartments. Three of the genes in our study are in regions of considerably higher SINEs than LINEs (Birc5, Brca1 and Coro1c) (Table 2) and might therefore lie in RIDGEs. In contrast, four genes have higher percentages of LINES (Ttk, Mef2c, Myom2 and Slc9a2) and might lie in anti-RIDGES. The genes BC002230, Obscn and Tpm3 have a

balanced ratio of SINE and LINE. Our studies therefore support a positive correlation of RIDGEs with gene distance to facultative as well as constitutive heterochromatin.

In summary, our data show that gene position is strongly influenced by the genomic context except when looking at distances to chromocenters in myoblasts. Since stem cells have a generally more open and active chromatin, this could explain how adult stem cells like myoblasts do not show influence of gene position by the genomic context.

Correlation between gene position changes and genomic context

The next question considered in this study, is whether gene position is changed upon activation or silencing and whether this positional change is correlated to gene expression or to the genomic context. To elucidate this question, we correlated the distance changes (Table 2, Delta) to the genomic properties including gene expression (Table 3). All correlations obtained are listed in Table S3. A schematic representation of the Pearson correlation coefficient for the 2-Mbp neighborhood can be found in Figure 23.

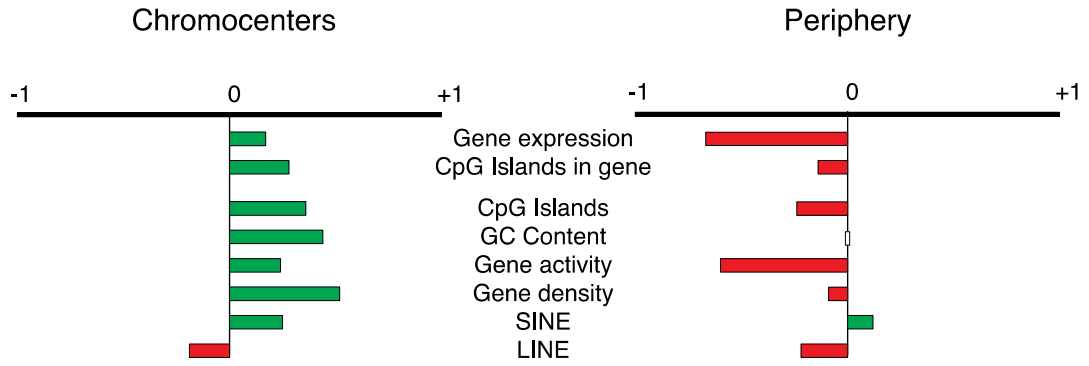


Figure 23. Graphic correlations between gene position changes and genomic context during differentiation. Graphic representation of Pearson correlation coefficients between normalized gene position change during differentiation and gene expression of the gene and genomic context in a 2-Mbp neighborhood (except for CpG Islands in gene). One represents a perfect correlation, zero no correlation and minus one an anti-correlation. (N>30 cells)

Gene movement relative to chromocenters shows no correlation with gene expression during differentiation (Pearson Coefficient = 0.17). By contrast, distances (absolute and normalized) to the periphery exhibit a negative correlation with gene expression (-0.34 and -0.67 respectively). Therefore, genes tend to be relocated closer to the periphery upon upregulation and further away upon down regulation. These data do not support the model of the nuclear periphery acting as a silencing compartment. Since the distances to the nuclear periphery range around 1 µm (Table 2) and chromatin mobility is thought to be around 0.5 µm, they might be out of the range of influence of the nuclear periphery. These observations might be falsified by taking genes into account that do not play any role in differentiation since their position might not be controlled so tightly. If we concentrate on genes strongly up regulated during myogenesis such as Obscn, Mef2c, Tpm3 and Myom2, there is no correlation between gene movement and its

expression status (Pearson coefficient: 0.07 to Periphery, 0.11 to chromocenters). Hence, we assume that other factors predominantly determine the position of genes and close proximity to heterochromatin does not automatically translate into gene silencing.

To ascertain whether the other factors could relate to genomic context of the genes, we next analyzed genomic context and gene movement (Figure 23 and Table S3). At a first glance, the correlations to chromocenter distances are positive or around zero whereas correlations to the periphery are mostly negative or nearly zero. Therefore, the genomic properties considered here all promote higher distances to chromocenters and lower distances to the nuclear periphery. Exceptions are LINEs which display more negative correlations as was observed for the relative position within the cells. Gene density might influence the ability to move since many genes with opposing expression profiles might be close to each other and negate each other in behavior. Therefore, the gene activity of the neighborhood is important. For movement towards chromocenters we see a low positive correlation in the gene activity and a higher negative correlation for the periphery. A cautionary note here is that these numbers are not exact since not all genes in the neighborhood appear in the expression profile (Figure 21). This means that an average up regulation of the neighborhood during differentiation leads to shorter distances to the periphery but to larger distances to chromocenters. Following the hypothesis of heterochromatin as a silencing compartment, this would only hold true for chromocenters and not for the periphery.

In summery our data does neither support nor contradict the idea of chromocenters as silencing compartments. Overall, the genomic context has a stronger effect on position changes to chromocenters. However, our data contradicts the hypothesis of the periphery acting as a silencing compartment as a strong negative correlation can be seen.

Outcome of gene position depending on heterochromatin (re)organization

Therefore, we could use this system to assess the effect of changes in chromatin architecture on gene position and expression without triggering the whole myogenic gene condition. We analyzed eight genes: Four of these were in common with the myogenesis analysis highlighted in bold in the Table 4 (Birc5, Brca1, Ttk and Myom2) and four genes that were not included in the differentiation study ((collagen type I alpha 2 (Col6a2), prolactin family 2 subfamily c member 2 (Prl2c2), brain derived neurotrophic factor (Bdnf), cell division cycle 20 homolog (cdc20)). We performed the same measurements as for the myogenic differentiation system and the results are summarized in Table 4.

Table 4 All distance measurements in MeCP2 expressing cells

	Chromocenter		Periphery	
	Distance in µm	Normalized distance	Distance in µm	Normalized distance

	Low MeCP2	High MeCP2	Delta	Low MeCP2	High MeCP2	Delta	Low MeCP2	High MeCP2	Delta	Low MeCP2	High MeCP2	Delta
Bdnf	0.99	1.47	0.48	0.68	0.72	0.03	0.74	0.76	0.02	0.54	0.54	0.00
Prl2c2	0.69	0.68	-0.01	0.50	0.47	-0.03	1.21	1.38	0.16	0.67	0.71	0.04
Myom2	0.71	0.86	0.15	0.52	0.57	0.05	0.99	1.00	0.00	0.58	0.57	0.00
cdc20	1.04	1.36	0.32	0.69	0.72	0.03	0.76	0.83	0.07	0.53	0.56	0.02
Brca1	0.86	0.93	0.07	0.54	0.56	0.02	1.70	1.72	0.02	0.81	0.79	-0.02
Birc5	0.94	0.89	-0.04	0.60	0.50	-0.09	1.49	1.69	0.20	0.72	0.79	0.06
Ttk	0.85	1.05	0.19	0.59	0.58	0.00	0.82	0.92	0.10	0.56	0.60	0.04
Col6a2	0.78	0.94	0.17	0.59	0.52	-0.07	1.04	1.30	0.26	0.67	0.72	0.05

An interesting observation is that the average distances in µm to chromocenters are generally shorter than in differentiation (Table 2) but the normalized distances are similar. As in the differentiation experiments distances in µm and normalized distances show a strong correlation but are not identical (Figure S12). Comparing the distances of the common genes studied in differentiation and MeCP2 ectopic expression collected in Table 2 and Table 4, neither the position within the cell nor the position changes are identical. Genes do not automatically reposition in a similar way just because chromatin is remodeled identically. This observation suggests a real regulatory function of repositioning.

The expression profile and genomic properties were collected as described for differentiation and are listed in Table 5.

Table 5 Genes analyzed in MeCP2 expressing cells and their genomic context

		Bdnf	Prl2c2	**Myom2**	cdc20	**Brca1**	**Birc5**	**Ttk**	Col6a2
	Gene Expression	2.13	2.06	1.15	1.08	-1.02	-1.13	-1.13	-1.52
	CpG Islands in Gene	4	0	2	1	1	1	1	2
2-Mbp window	CpG Island	12	25	18	49	67	54	20	36
	GC Content	38.98	42.02	44.08	47.15	48.06	51.18	40.10	47.42
	Gene activity	1.51	1.34	1.05	1.06	1.04	-1.07	-1.42	-1.13
	Gene density	9	24	7	34	50	23	7	21
	SINE	5.13	9.02	3.54	14.82	21.70	11.87	5.19	9.93
	LINE	25.65	21.75	13.31	8.07	2.45	2.06	21.10	8.53
5-Mbp window	CpG Islands	27	38	58	128	127	162	35	76
	GC Content	38.09	39.16	44.15	46.14	47.08	50.7	39.05	45.73
	Gene activity	1.51	1.28	-1.05	1.12	1.04	-1.01	1,00	-1.01
	Gene density	77	57	47	143	221	153	29	131
	SINE	3.52	5.12	4.97	16.78	17.57	14.78	4.45	8.62
	LINE	32.35	28.27	11.16	7.40	4.82	1.88	25.54	14.70

All four genes in common with the differentiation study (bold in Table 5) do not show a significant gene expression change due to MeCP2 expression, which enables us to differentiate between correlation with the genomic architecture and their expression profiles. Since the fold changes due to MeCP2 are very small, even changes of 1.5 fold can be considered as

significant. Taking a change of 1.5 fold is standard practice in other MeCP2 expression profiles (Ballestar *et al.*, 2005; Delgado *et al.*, 2006; Tudor *et al.*, 2002). Therefore, the genes Col6a2, Prl2c2 and Bdnf are significantly changed upon MeCP2 expression. Bdnf is a known target gene of MeCP2 (Martinowich *et al.*, 2003) with an important role in neurodevelopment (Gonzales & LaSalle, 2010).

As observed in our differentiation data (Table 3), the percentage of DNA covered by SINEs and LINEs shows an inverse pattern: high percentages of SINEs are followed by lower numbers of LINEs and *vice versa*. The only exception is Col6a2 which has nearly identical percentages of SINEs and LINEs. In the 5-Mbp window this difference is more pronounced and also detectable for Col6a2.

Correlation between gene – heterochromatin distances and genomic context relative to MeCP2 level

To detect positional preferences within the nucleus, we correlated the gene position with the genomic context as performed for differentiation. The correlations to normalized distances in a 2-Mbp neighborhood are depicted in Figure 24. All results are collected in Table S4.

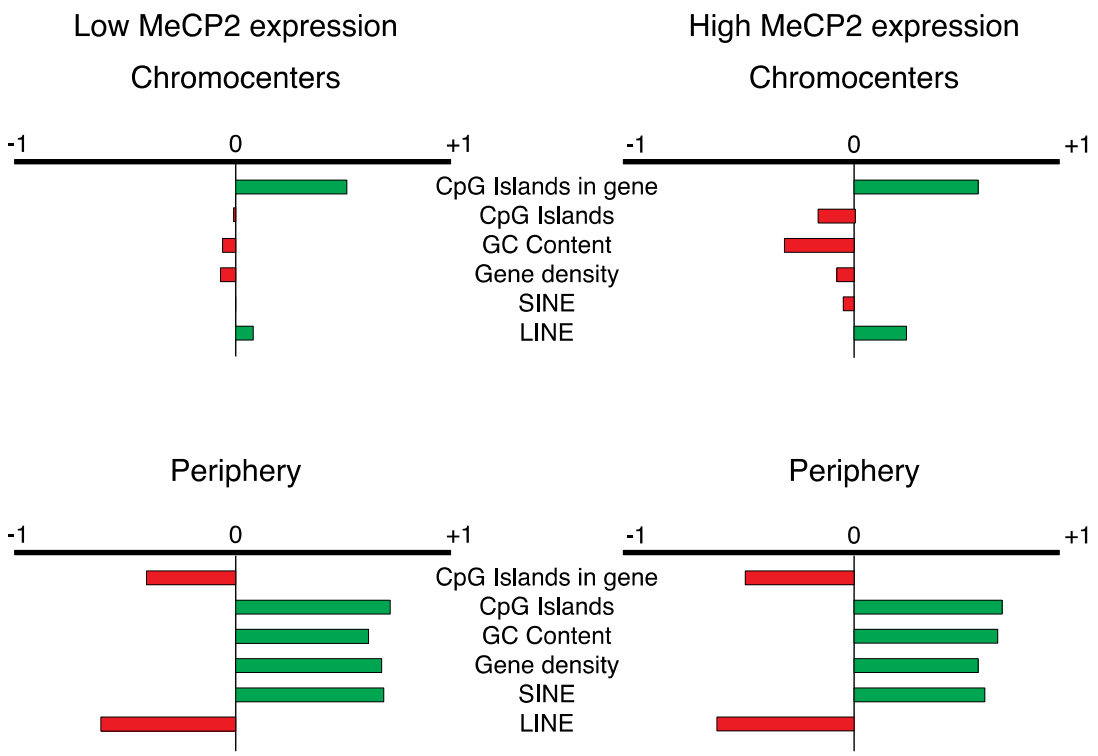


Figure 24. Correlation between gene - heterochromatin distances and genomic context relative to MeCP2 levels. Graphic representation of Pearson correlation coefficients relative to MeCP2 levels between normalized gene distance and genomic context in a 2-Mbp neighborhood (except for CpG Islands in gene). One represents a perfect correlation, zero no correlation and minus one an anti-correlation. (N>30 cells)

The results obtained for MeCP2 expressing cells differ from the observation in the myogenic differentiation system in several parts. In the distances to chromocenters (in µm as well as normalized) no trend is visible. The correlations of the genomic context in a 2-Mbp are very low both in low and high MeCP2 expressing cells. The only clear positive correlation is observed between CpG islands within the gene with gene distance to chromocenters (high number of CpG, large distance to chromocenters). CpG islands are devoid of methylated CpG to which MeCP2 is known to bind. Therefore genes with CpG islands have potentially less binding sites for MeCP2 and are not bound to chromocenters at which MeCP2 is concentrated. Since the periphery is not enriched in MeCP2 such a correlation is not observed.

Similar to the myogenesis study many and very high correlations are observed for gene positioning to the periphery. As in the differentiation data set, opposite behavior is observed for SINEs and LINEs; SINEs being positively correlated with gene position and LINEs negatively correlated. Other genomic features (CpG islands in a 2-Mbp neighborhood, GC content and gene density) are positively correlated meaning genes with higher gene density, CG content and many CpG islands are found further from the nuclear periphery. Besides LINEs, all correlations are positive no matter if in the direct neighborhood or within the 5-Mbp neighborhood (Table S3).

In general, our data show that low and high MeCP2 expressing cells exhibit the same gene positioning correlations as observed for myoblasts in the myogenic differentiation system. The

only exception being CpG islands within the gene which are directly correlated to ectopic MeCP2 expression. This underlines the dominant influence of genomic context and overall expression on gene positioning and its independence of chromatin rearrangements.

Correlation between gene position change and genomic context relative to MeCP2 level

To determine whether gene position changes are influenced more strongly by genomic architecture or by other factors we correlated the change in gene position dependent on MeCP2 level with the genomic context. Interestingly, the average distances in µm to chromocenters are much shorter than in differentiation (Table 2) but normalization shows that this is due to morphology. As mentioned before, distances in µm and normalized distances show a good correlation but are not identical (Figure S12). Comparing the distances and distance changes of the genes common between the differentiation and ectopic MeCP2 expression analyses collected in Table 2 and Table 4, neither the position within the cell nucleus nor the position changes are identical. This is a strong indication that the genomic architecture itself is not the main regulatory force.

The expression profile and genomic properties were collected as described for differentiation and are listed in Table 5. The correlations to normalized distances in a 2-Mbp neighborhood is depicted in Figure 25. All results are collected in Table S5.

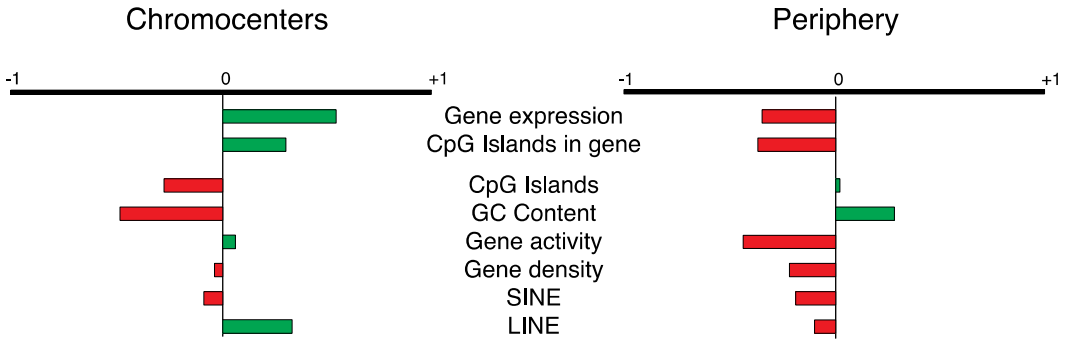


Figure 25. Correlation between gene position change and genomic context relative to MeCP2 levels. Graphic representation of Pearson correlation coefficients between normalized gene position change in relative to MeCP2 level and gene expression of the gene and genomic context in a 2-Mbp neighborhood (except for CpG Islands in gene). One represents a perfect correlation, zero no correlation and minus one an anti-correlation. (N>30 cells)

A positive correlation could be observed for gene expression changes and gene movement towards chromocenters whereas a concomitant negative correlation is observed for gene movement to the periphery. This observation means that genes up regulated upon MeCP2 expression increase their distance to chromocenters but decrease their distance to the periphery and *vice versa* for genes that are down regulated. This observation fits the general hypothesis of chromocenters being silencing compartments. A difference can be observed in the correlations to LINEs. In the differentiation experiments, we observed a negative correlation of gene movement towards chromocenters with LINEs but in MeCP2 expressing cells a positive correlation is observed. A moderate to stronger negative correlation can also be observed in

CpG islands and GC content. Since the GC content is higher in CpG islands it is logical that similar correlations are observed. A light positive correlation of gene movement relative to chromocenters is observed with the number of CpG islands within the gene itself underlining the difference between the neighborhood (which had a negative correlation) and the single gene (positive correlation). This means that genes with high content of CpG islands show larger movements away from chromocenters due to MeCP2 expression, than genes with low numbers of CpG islands. MeCP2 is enriched at chromocenters and is known to bind to methylated CpGs. However, CpG islands contain mostly unmethylated CpGs and hence are no binding sites for MeCP2 and are further away from chromocenters.

Genomic context influences gene position

In our study, we were able to elucidate the influence of expression and genomic context on gene position and gene movement during differentiation. We could show that in myogenic differentiation no correlation between gene expression and movement further to chromocenters exists disagreeing with the popular hypothesis of gene silencing due to heterochromatin association. In addition, we could show that genes up regulated during differentiation move closer to the nuclear periphery. This observation contradicts published data proposing that up regulated genes increase their distance to the periphery and down regulated genes move closer to the nuclear periphery (Lee *et al.*, 2006). Even though much data has been published about periphery acting as a silencing compartment, an early theory stated the reverse possibility. G. Blobel stated that genes might be assigned to specific nuclear pore complexes in order to gate all transcripts from this gene through a certain pore (Blobel, 1985). This means active genes would become associated with the nuclear pore complex and would be perceived as nuclear periphery associated in our study. We could also show that gene movement is much more dependent on the genomic context around the gene (Figure 23) which might override the effect of expression of the particular gene.

To dissect the influence of (hetero)chromatin organization separately from cellular differentiation we used cells transfected with MeCP2 mimicking large-scale chromatin remodeling with lower gene expression changes and no nuclear size and shape changes. Using this reduced system we could show that a positive correlation between gene expression and gene movement relative to chromocenters can be observed i.e., higher gene expression; farther from chromocenters. However, the maximum Pearson correlation coefficient is only around 0.5 meaning that we cannot observe clear correlations but mere trends. Expressed genes move farther from chromocenters and closer to the nuclear periphery depending on MeCP2 levels and concomitant chromocenter rearrangement. This is consistent with chromocenters being silencing compartments but does not support the hypothesis of the nuclear periphery as a general repressive environment.

MATERIALS AND METHODS

Cell culture

Pmi 28 mouse myoblasts were cultured using standard conditions described previously (Brero *et al.*, 2005). For myotubes, $8x10^5$ cells were seeded on a 100 mm Ø dish and one day later FCS (fetal calf serum) working medium was replaced by 5% horse serum in the medium. Cells were then differentiated until the formation of large polynucleated myotubes could be observed (4-7 days). Cells plated on glass coverslips for FISH studies were washed with PBS and subsequently fixed with 4% paraformaldehyde

Transfection and FACSorting

Pmi 28 myoblasts were transfected with a mammalian expression construct coding for YFP tagged rat MeCP2 (Brero *et al.*, 2005) using Transfectin (Bio Rad, München, Germany) or Amaxa according to the manufacturer's advice.

The cells were washed with PBS-EDTA twice, trypsinized and collected in medium. After a centrifugation step at 900 rpm cells were diluted in sterile PBS for FACSorting. Using the FACS Aria I (Becton Dickinson, Franklin Lakes, NJ USA) we separated cells with high fluorescent intensity gating 8.3% of all cells (fluorescent intensity mean 322) and cells with low fluorescence intensity gating 25% of all cells (fluorescent intensity mean 247). Intensity sorted cells were frozen as pellets for RNA extraction.

RNA Preparation and cDNA synthesis

Cell pellets with $6.5x10^5$ to $1.7x10^6$ cells were used for RNA extraction. Total RNA was extracted using Trizol reagent (Invitrogen, Paisley PA4 9RF, UK) and purified using RNeasy Mini kit (Qiagen, Valencia, CA 91355 USA) in accordance with the manufacturer's protocol. Double-stranded cDNA was synthesized from total RNA using two different kits according to RNA concentration. If the total RNA yield was between 1-20 µg/µl, the cDNA Synthesis kit from Roche (Roche, Mannheim, Germany) was used, taking 10 µg total RNA. For yield between 10-100 ng/µl the two-cycle kit from Invitrogen (Invitrogen, Paisley PA4 9RF, UK) was used, taking 50 ng/µl total RNA.

Microarray analysis

The resulting cDNA was hybridized to the Affymetrix mouse 430 2.0 microarray, which carries 39,000 transcripts per chip. Five independent experiments were performed for all conditions (myoblast, myotube, low and high MeCP2)

The quality of the hybridization and overall microarray performance was determined by visual inspection of the raw scanned data for artifacts, scratches and bubbles. The GeneChip® Operating Software report file was then used to determine if the following statistics were within

acceptable limits: 3'/5' signal ratio of GAPDH and β-Actin, assay background and noise, and proportion and average expression value of detected genes.

Arrays were normalized for condition project individually by the log scale robust multi-array analysis (RMA). Its estimates are based upon a robust average of log2 (B(PM)), where B(PM) are background corrected perfect match intensities. Compared with the GCOS method the RMA method has better precision, provides more consistent estimates of fold change and provides higher specificity and sensitivity when using fold change analysis to detect differential expression (Irizary *et al.*, 2003).

First a Nalimov test was performed to exclude outliers from the statistical analysis (threshold: p=0.0001). No arrays were excluded by this test. The average and standard deviation of the antilog RMA values were calculated and from there the fold change was obtained. An unpaired student t-test was performed. Only genes with changes of very high statistical significance ($p \leq 4x10^{-06}$) were chosen for further analysis. Bacterial artificial chromosomes (BACs) were obtained from BAC-PAC resource center (Oakland, CA, USA, http://bacpac.chori.org).

Baculoviral IAP repeat-containing 5	RP23-220P14
Breast cancer 1	RP23-222H10
Ttk protein kinase	RP24-211B11
RIKEN cDNA 6720454P05	RP24-117A2
Obscurin	RP23-113H6
Myocyte enhancer factor 2C	RP23-205E14
Tropomyosin 3, gamma	RP23-163L22
Procollagen, type VI, alpha 2	RP23-27P21
Proliferin 2	RP23-155I17
Coronin, actin binding protein 1C	RP24-156M14
Brain derived neurotrophic factor	RP24-310A6
Myomesin 2	RP24-244I21
Solute carrier family 19 (thiamine transporter), member 2	RP24-158B1
Cdc20 like	RP23-118J14

These DNA probes were labeled by standard nick translation with Biotin-dUTP (Amersham, Buckinghamshire, UK).

DNA probes (Immuno) Fluorescence in situ hybridization

From each BAC DNA 10 µl of labeled DNA was precipitated by adding 2 µl of fish sperm DNA, 15 µl sodium acetate and 150 µl of 100% pure ice cold ethanol. The precipitation mix was incubated at -75 °C for 50 minutes. The probe was centrifuged at 4 °C for 45 minutes at 13.000 rpm. The pellet was then washed with 70% pure ice-cold ethanol and centrifuged for 30 minutes. Next the pellet was air dried to make sure no ethanol remnants are retained. The cleaned labeled DNA was dissolved in 6 µl hybridization solution, which consisted of 50%

formamide, 2xSSC, 10% dextran sulfate, pH 7.0. The probes were denatured at 80 °C for 5 minutes.

In the case of Immuno FISH cells fixed with 4% Paraformaldehyde were permeabilized with 0.25% Triton X-100 for 10 minutes and incubated in blocking solution 4% BSA/PBS (bovine serum albumin) for 30 minutes. Our self-made MeCP2 antibodies described in chapter 1 were used as first antibody. The anti MeCP2 rabbit polyclonal antibody was diluted 1:500 in the blocking solution the monoclonal antibodies were used undiluted and incubated for 1 h. Anti rabbit Alexa 488 or anti rat Alexa were used as secondary antibodies respectively. After a 15 minutes posfixation with 1% PFA the protocol was continued as for FISH.

Fixed cells were permeabilized with 0.5% Triton X-100 in PBS for 20 minutes, 0.1 M HCl for 20 minutes and with 0.5% Triton X-100 in PBS for 20 minutes.

Probe and cells were brought together in hybridization chambers to decrease evaporation of the probe over night. Probe and cells are then denatured simultaneously at 75 °C for 5 minutes and hybridized over night at 37 °C. After incubation unhybridized probe was washed off using 50% formamide in SSC at 45 °C three times followed by two washes with 2xSSC. The detection of the probe was performed by Strep-Cy5 detection and an enhancement of the signal by anti Streptavidin-biotin detection followed by again Strep-Cy5. A DAPI counterstain was then performed and the cells were mounted using vectashield.

Microscopy and image analysis

Confocal optical Z stacks of images (voxel size: 80 x 80 x 200 nm) were obtained using a Leica SP5 laser scanning microscope, equipped with 63x/1.4NA oil objective. The fluorophores were excited with 405 nm DPSS (for DAPI detection), 488 nm (for Alexa 488 detection) and 633 nm (for Cy5 detection) laser lines. Care was taken in selecting the imaging conditions to avoid under and over exposed pixels, while keeping the imaging conditions constant. Distance measurements and analysis were performed as previously described (chapter 2)

Databases and genomic context analysis

Suitable BACs as well as neighboring genes were identified in the “cytoview” display of the Ensembl Genome Browser (http://www.ensembl.org). For the 2- and 5-Mbp windows distances of 1 or 2.5 Mbp were calculated upstream and downstream from the center of each gene. The CpG islands and gene in these regions were noted and the sequences downloaded. Sequences were submitted to RepeatMasker (http://www.repeat masker.org, version open-3.3.0)

Statistical analysis

Data analyses of all measurements were performed using excel.

REFERENCES

Baccarini, P. (1908) Sulle cinesi vegeative del "Cynomorium coccineum L.". *Nuovo Girornale botanico italiano Nuova serie*, **15**, 189-203.

Baharvand, H. and Matthaei, K.I. (2003) The ultrastructure of mouse embryonic stem cells. *Reprod Biomed Online*, **7**, 330-335.

Ballestar, E., Ropero, S., Alaminos, M., Armstrong, J., Setien, F., Agrelo, R., Fraga, M.F., Herranz, M., Avila, S., Pineda, M. *et al.* (2005) The impact of MECP2 mutations in the expression patterns of Rett syndrome patients. *Hum Genet*, **116**, 91-104.

Blobel, G. (1985) Gene gating: a hypothesis. *Proc Natl Acad Sci U S A*, **82**, 8527-8529.

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.

Britten, R.J. and Davidson, E.H. (1969) Gene regulation for higher cells: a theory. *Science*, **165**, 349-357.

Brown, K.E., Guest, S.S., Smale, S.T., Hahm, K., Merkenschlager, M. and Fisher, A.G. (1997) Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell*, **91**, 845-854.

Caron, H., van Schaik, B., van der Mee, M., Baas, F., Riggins, G., van Sluis, P., Hermus, M.C., van Asperen, R., Boon, K., Voute, P.A. *et al.* (2001) The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science*, **291**, 1289-1292.

Chahrour, M., Jung, S.Y., Shaw, C., Zhou, X., Wong, S.T., Qin, J. and Zoghbi, H.Y. (2008) MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science*, **320**, 1224-1229.

Chueh, A.C., Northrop, E.L., Brettingham-Moore, K.H., Choo, K.H. and Wong, L.H. (2009) LINE retrotransposon RNA is an essential structural and functional epigenetic component of a core neocentromeric chromatin. *PLoS Genet*, **5**, e1000354.

Colantuoni, C., Jeon, O.H., Hyder, K., Chenchik, A., Khimani, A.H., Narayanan, V., Hoffman, E.P., Kaufmann, W.E., Naidu, S. and Pevsner, J. (2001) Gene expression profiling in postmortem Rett Syndrome brain: differential gene expression and patient classification. *Neurobiol Dis*, **8**, 847-865.

Cremer, M., Muller, S., Kohler, D., Brero, A. and Solovei, I. (2007) Cell Preparation and Multicolor FISH in 3D Preserved Cultured Mammalian Cells. *CSH Protoc*, **2007**, pdb prot4723.

Delaire, S., Huang, Y.H., Chan, S.W. and Robey, E.A. (2004) Dynamic repositioning of CD4 and CD8 genes during T cell development. *J Exp Med*, **200**, 1427-1435.

Delgado, I.J., Kim, D.S., Thatcher, K.N., LaSalle, J.M. and Van den Veyver, I.B. (2006) Expression profiling of clonal lymphocyte cell cultures from Rett syndrome patients. *BMC Med Genet*, **7**, 61.

Duraisingh, M.T., Voss, T.S., Marty, A.J., Duffy, M.F., Good, R.T., Thompson, J.K., Freitas-Junior, L.H., Scherf, A., Crabb, B.S. and Cowman, A.F. (2005) Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in Plasmodium falciparum. *Cell*, **121**, 13-24.

Easwaran, H.P., Schermelleh, L., Leonhardt, H. and Cardoso, M.C. (2004) Replication-independent chromatin loading of Dnmt1 during G2 and M phases. *EMBO Rep*, **5**, 1181-1186.

Finlan, L.E., Sproul, D., Thomson, I., Boyle, S., Kerr, E., Perry, P., Ylstra, B., Chubb, J.R. and Bickmore, W.A. (2008) Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet*, **4**, e1000039.

Gardiner-Garden, M. and Frommer, M. (1987) CpG islands in vertebrate genomes. *J Mol Biol*, **196**, 261-282.

Gonzales, M.L. and LaSalle, J.M. (2010) The role of MeCP2 in brain development and neurodevelopmental disorders. *Curr Psychiatry Rep*, **12**, 127-134.

Harmon, B. and Sedat, J. (2005) Cell-by-cell dissection of gene expression and chromosomal interactions reveals consequences of nuclear reorganization. *PLoS Biology*, **3**, e67.

Hepperger, C., Mannes, A., Merz, J., Peters, J. and Dietzel, S. (2008) Three-dimensional positioning of genes in mouse cell nuclei. *Chromosoma*, **117**, 535-551.

Hubner, M.R. and Spector, D.L. (2010) Chromatin dynamics. *Annu Rev Biophys*, **39**, 471-489.

Jacob, F. and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *Journal of molecular biology*, **3**, 318-356.

Kupper, K., Kolbl, A., Biener, D., Dittrich, S., von Hase, J., Thormeyer, T., Fiegler, H., Carter, N.P., Speicher, M.R., Cremer, T. *et al.* (2007) Radial chromatin positioning is shaped by local gene density, not by gene expression. *Chromosoma*, **116**, 285-306.

Lasalle, J.M. and Yasui, D.H. (2009) Evolving role of MeCP2 in Rett syndrome and autism. *Epigenomics*, **1**, 119-130.

Lee, H., Quinn, J.C., Prasanth, K.V., Swiss, V.A., Economides, K.D., Camacho, M.M., Spector, D.L. and Abate-Shen, C. (2006) PIAS1 confers DNA-binding specificity on the Msx1 homeoprotein. *Genes Dev*, **20**, 784-794.

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. *The Journal of Cell Biology*, **149**, 271-280.

Martinowich, K., Hattori, D., Wu, H., Fouse, S., He, F., Hu, Y., Fan, G. and Sun, Y.E. (2003) DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. *Science*, **302**, 890-893.

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.

Meaburn, K.J. and Misteli, T. (2008) Locus-specific and activity-independent gene repositioning during early tumorigenesis. *J Cell Biol*, **180**, 39-50.

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.

Muller, H. (1930) Types of visible variations induced by X-rays in Drosophila. *Journal of Genetics*, **22**, 299-334.

Murmann, A.E., Gao, J., Encinosa, M., Gautier, M., Peter, M.E., Eils, R., Lichter, P. and Rowley, J.D. (2005) Local gene density predicts the spatial position of genetic loci in the interphase nucleus. *Exp Cell Res*, **311**, 14-26.

Nan, X., Campoy, F.J. and Bird, A. (1997) MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell*, **88**, 471-481.

Park, S.H., Kook, M.C., Kim, E.Y., Park, S. and Lim, J.H. (2004) Ultrastructure of human embryonic stem cells and spontaneous and retinoic acid-induced differentiating cells. *Ultrastruct Pathol*, **28**, 229-238.

Reddy, K.L., Zullo, J.M., Bertolino, E. and Singh, H. (2008) Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature*, **452**, 243-247.

Sabbattini, P., Lundgren, M., Georgiou, A., Chow, C., Warnes, G. and Dillon, N. (2001) Binding of Ikaros to the lambda5 promoter silences transcription through a mechanism that does not require heterochromatin formation. *Embo J*, **20**, 2812-2822.

Saxonov, S., Berg, P. and Brutlag, D.L. (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A*, **103**, 1412-1417.

Simmer, F., Buscaino, A., Kos-Braun, I.C., Kagansky, A., Boukaba, A., Urano, T., Kerr, A.R. and Allshire, R.C. (2010) Hairpin RNA induces secondary small interfering RNA synthesis and silencing in trans in fission yeast. *EMBO Rep*, **11**, 112-118.

Singer, M.F. (1982) SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. *Cell*, **28**, 433-434.

Szczerbal, I., Foster, H.A. and Bridger, J.M. (2009) The spatial repositioning of adipogenesis genes is correlated with their expression status in a porcine mesenchymal stem cell adipogenesis model system. *Chromosoma*, **118**, 647-663.

Takai, D. and Jones, P.A. (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A*, **99**, 3740-3745.

Takizawa, T., Meaburn, K.J. and Misteli, T. (2008) The meaning of gene positioning. *Cell*, **135**, 9-13.

Tudor, M., Akbarian, S., Chen, R.Z. and Jaenisch, R. (2002) Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. *Proc Natl Acad Sci U S A*, **99**, 15536-15541.

Zeuschner, D., Mildner, K., Zaehres, H. and Scholer, H.R. (2010) Induced pluripotent stem cells at nanoscale. *Stem Cells Dev*, **19**, 615-620.

SUPPLEMENTARY INFORMATION

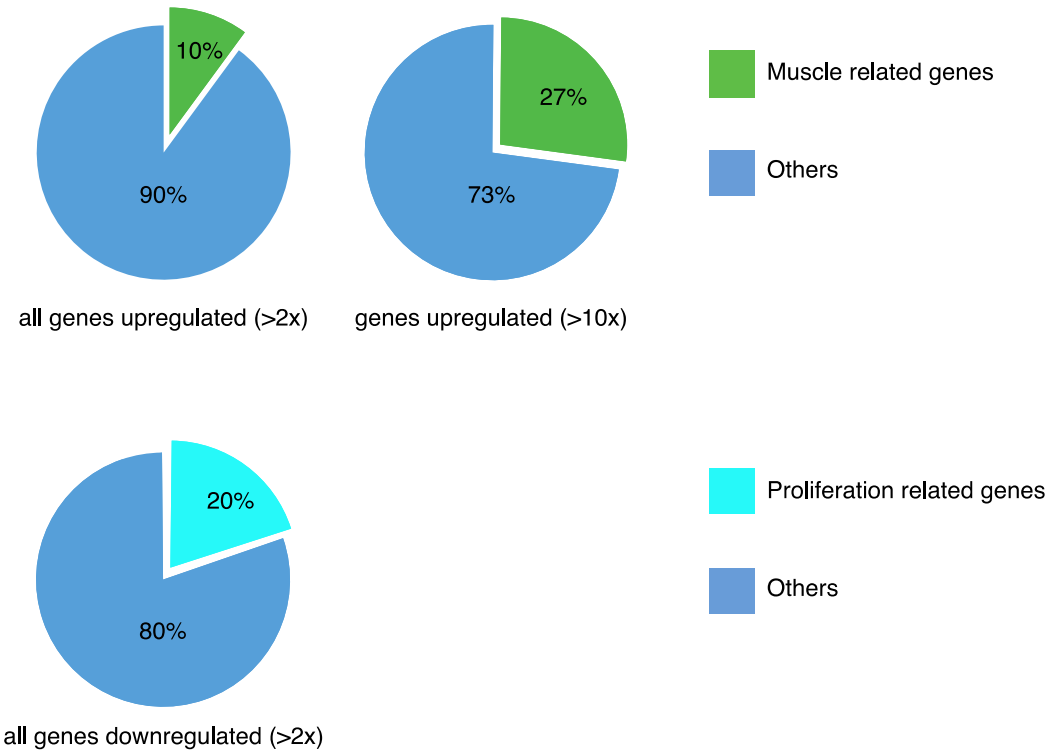


Figure S10. Gene ontology analysis of transcriptional profiling. All genes detected in the transcriptional profiling with statistical significant and at least two fold expression (up and down) were analyzed for their gene ontology using DAVID. Muscle related gene a highly up regulated whereas proliferation related genes are down regulated.

Supplement Figure 1

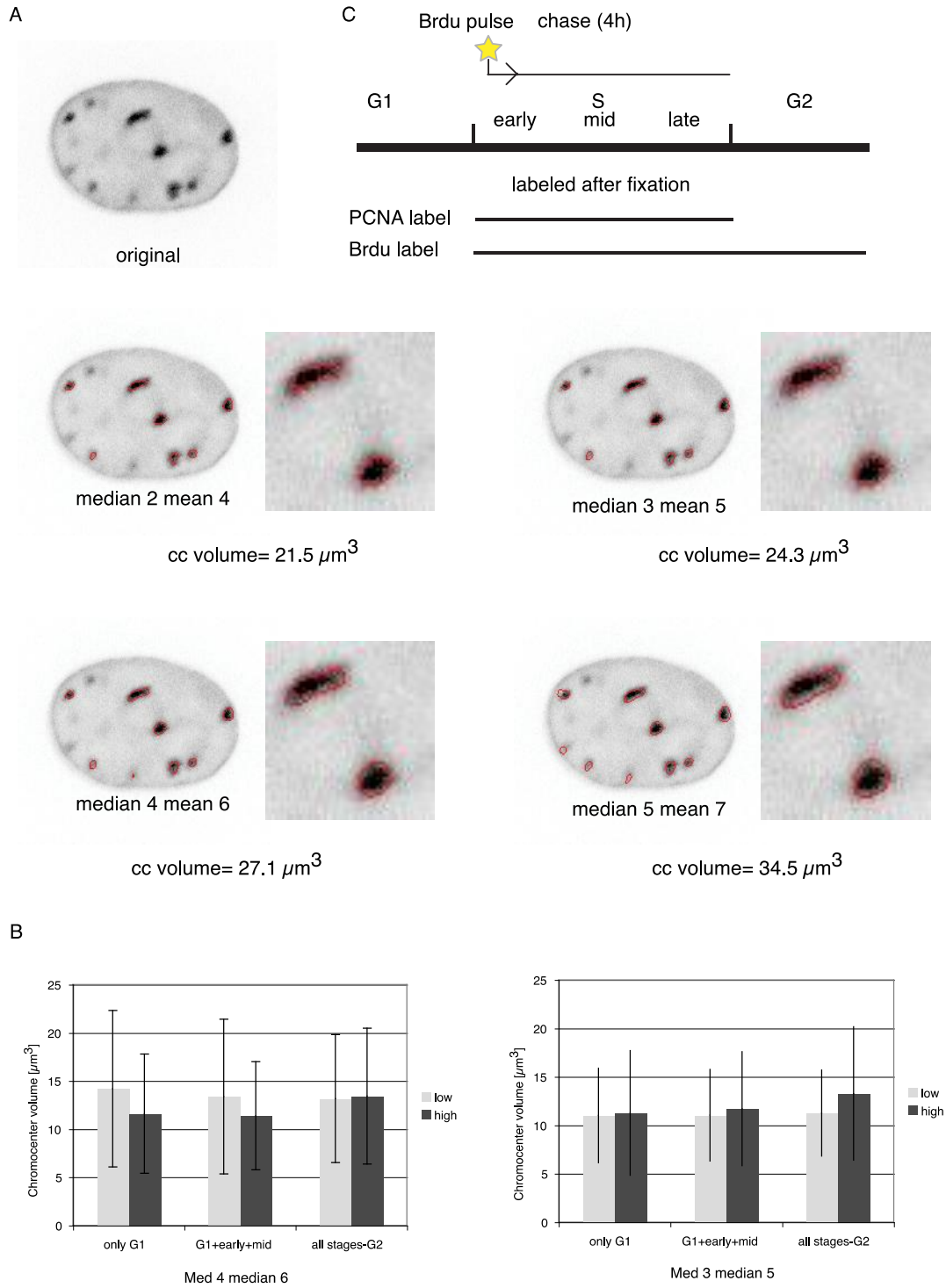


Figure S11. Thresholding effect on chromocenter volume. Using the software described in chapter 2 we were able to establish that volume differences are larger within between different settings then between low and high MeCP2 expressing cells. **A**. Illustrates the strong effect of thresholding on volume measurements. **B.** The plot illustrates the average volume of chromocenters taking different cell cycle stages and two thresholding parameters into account. **C** The scheme illustrates how the experiment allowed distinguishing between different stages of the cell cycle.

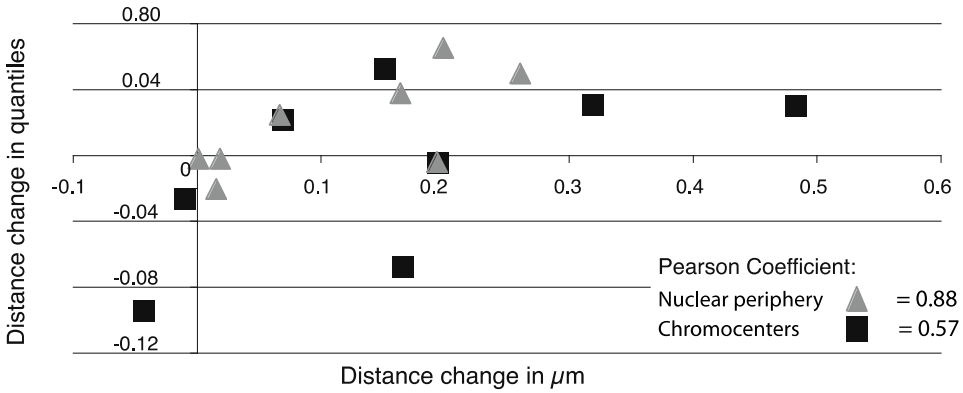


Figure S12. Raw vs. normalized data in MeCP2 expressing cells. Distances measured in µm (*x*-axis) plotted against the normalized distances in quantiles (*y*-axis). Distances relating to the periphery are depicted by gray triangles and distances to chromocenters by black squares. The Pearson correlation coefficient is indicated below.

Table S1 Quality parameters

File	Date	Chip	Alias	Outlier*	RawQ	SF	TGT	NF	Backg	Noise	No_P
CAR_01_MB1_190805_430_2.CHP	11:44AM 08/23/2005	Mouse 430_2	MB1	36	2,19	0,775	200	1	61,9	4,6	47,00 %
CAR_02_MB2_190805_430_2.CHP	11:44AM 08/23/2005	Mouse 430_2	MB2	54	2,35	0,805	200	1	70,31	4,39	46,30 %
CAR_03_MB3_190805_430_2.CHP	11:44AM 08/23/2005	Mouse 430_2	MB3	44	2,22	0,907	200	1	67,71	4,2	45,50 %
CAR_04_MB4_190805_430_2.CHP	11:44AM 08/23/2005	Mouse 430_2	MB4	39	2,31	0,852	200	1	69,13	4,13	46,40 %
CAR_05_MB6_190805_430_2.CHP	11:44AM 08/23/2005	Mouse 430_2	MB6	53	2,31	0,752	200	1	69,29	4,58	46,70 %
CAR_06_MT1_190805_430_2.CHP	11:44AM 08/23/2005	Mouse 430_2	MT1	56	2,34	0,884	200	1	68,61	4,35	48,00 %
CAR_07_MT_190805_430_2.CHP	11:44AM 08/23/2005	Mouse 430_2	MT0	34	2,3	0,775	200	1	69,04	4,68	47,90 %
CAR_08_MT3_190805_430_2.CHP	11:44AM 08/23/2005	Mouse 430_2	MT3	52	2,1	0,94	200	1	65,56	4,02	46,80 %
CAR_09_MT4_190805_430_2.CHP	11:44AM 08/23/2005	Mouse 430_2	MT4	40	2,34	0,72	200	1	65,25	4,59	49,30 %
CAR_10_MT2_190805_430_2.CHP	11:44AM 08/23/2005	Mouse 430_2	MT2	109	2,31	0,881	200	1	66,71	4,95	46,90 %
CAR_11_R5_1_240805_430_2.CHP	01:28PM 08/25/2005	Mouse 430_2	R5_1	29**	3,85	0,802	200	1	128,29	12	39,40 %
CAR_12_R5_3_240805_430_2.CHP	01:30PM 08/25/2005	Mouse 430_2	R5_3	104**	4,36	0,903	200	1	157,98	13,15	37,90 %
CAR_13_R5_5_240805_430_2.CHP	01:30PM 08/25/2005	Mouse 430_2	R5_5	32	2,46	0,891	200	1	74,9	4,72	42,90 %
CAR_14_R5_1807_240805_430_2.CHP	01:31PM 08/25/2005	Mouse 430_2	R5_1807	15	2,67	1,065	200	1	85,87	5,87	41,80 %
CAR_15_R5_10_240805_430_2.CHP	01:32PM 08/25/2005	Mouse 430_2	R5_10	32	2,34	0,994	200	1	73,42	4,78	42,70 %
CAR_16_R4p1_240805_430_2.CHP	01:33PM 08/25/2005	Mouse 430_2	R4_1	35	2,41	1,01	200	1	74,51	4,83	42,70 %
CAR_17_R4p2_240805_430_2.CHP	01:34PM 08/25/2005	Mouse 430_2	R4_2	114	2,24	0,936	200	1	69,63	4,55	45,30 %
CAR_18_R4p3_240805_430_2.CHP	01:36PM 08/25/2005	Mouse 430_2	R4_3	35	2,24	1,039	200	1	69,85	4,28	42,60 %
CAR_19_R4p4_240805_430_2.CHP	01:36PM 08/25/2005	Mouse 430_2	R4_4	53	2,28	1,052	200	1	70,69	4,1	42,70 %
CAR_20_R4p1807_240805_430_2.CHP	01:38PM 08/25/2005	Mouse 430_2	R4_1807	46	2,38	1,1	200	1	73,25	4,73	43,40 %
Minimum				15	2,1	0,72			61,9		0,379
Maximum				114	4,36	1,1			157,98		0,493

File	No_A	No_M	Avg_P	Avg_A	Avg_M	Avg_All	GAPDH	18S	β_ACTIN	BIOB	BIOC
CAR_01_MB1_190805_430_2.CHP	51,30%	1,60%	590,5	28	76,2	293,4	0,86	0,32	1,16	0,83	1,11
CAR_02_MB2_190805_430_2.CHP	51,90%	1,70%	619,3	25,9	73,4	301,7	0,84	0,28	1,16	0,71	1,03
CAR_03_MB3_190805_430_2.CHP	53,00%	1,60%	636,3	27,7	82,7	305,4	0,82	0,33	1,15	0,92	1,1
CAR_04_MB4_190805_430_2.CHP	51,90%	1,70%	618,9	26,8	77,4	302,3	0,81	0,38	1,14	0,75	1,05
CAR_05_MB6_190805_430_2.CHP	51,70%	1,60%	604,1	24,5	67,5	295,8	0,82	0,42	1,13	0,76	1,08
CAR_06_MT1_190805_430_2.CHP	50,30%	1,70%	574,6	28,5	73,5	291,1	0,81	0,53	1,17	0,8	0,96
CAR_07_MT_190805_430_2.CHP	50,30%	1,70%	577,1	25,6	68,7	290,8	0,81	0,58	1,17	0,78	1,01
CAR_08_MT3_190805_430_2.CHP	51,60%	1,60%	604,1	26,5	70,7	297,6	0,8	0,52	1,19	0,76	1,03
CAR_09_MT4_190805_430_2.CHP	48,90%	1,80%	537,8	28,8	89,9	280,7	0,79	0,62	1,13	1,18	1,02
CAR_10_MT2_190805_430_2.CHP	51,30%	1,70%	579,9	30,3	83,1	289,1	0,79	0,62	1,26	0,79	1,07
CAR_11_R5_1_240805_430_2.CHP	58,90%	1,70%	676,5	38,3	121,7	291,3	1,82	0,35	1,61	0,81	0,75
CAR_12_R5_3_240805_430_2.CHP	60,40%	1,70%	696,8	44,9	130,4	293,7	1,95	0,27	1,91	0,69	0,8
CAR_13_R5_5_240805_430_2.CHP	55,50%	1,70%	678,6	26,6	82,4	307,2	1,92	0,3	1,61	0,93	0,87
CAR_14_R5_1807_240805_430_2.CHP	56,60%	1,60%	694,5	31,2	87,8	309,4	2,06	0,23	1,66	0,94	0,88
CAR_15_R5_10_240805_430_2.CHP	55,60%	1,70%	680,8	27,2	85,8	307,4	1,71	0,22	1,67	0,75	1,15
CAR_16_R4p1_240805_430_2.CHP	55,60%	1,70%	687,7	26,8	82,4	310	1,77	0,27	1,83	0,81	1,25
CAR_17_R4p2_240805_430_2.CHP	53,10%	1,60%	661,1	22,1	62,7	312,3	1,93	0,3	1,9	0,96	1,16
CAR_18_R4p3_240805_430_2.CHP	55,80%	1,60%	693,8	26,4	76,1	311,5	1,96	0,25	1,71	0,75	1,04
CAR_19_R4p4_240805_430_2.CHP	55,70%	1,60%	691,5	25,8	90,2	310,9	2,25	0,26	1,76	0,91	1,14
CAR_20_R4p1807_240805_430_2.CHP	55,00%	1,70%	687,8	25,1	73,9	313,2	2,15	0,25	1,6	1,07	1,16
Minimum			537,8				0,79		1,13		
Maximum			696,8				2,25		1,91		

File	BIOD	CRE	DAP	LYS	PHE	THR	TRP	Alpha1	Alpha2	Tau
CAR_01_MB1_190805_430_2.CHP	2,28	1,32	0,25	1,9	30,66	2,23	1,95	0,05	0,065	0,015
CAR_02_MB2_190805_430_2.CHP	2,05	1,31	0,6	0,5	11,01	0,2	0,67	0,05	0,065	0,015
CAR_03_MB3_190805_430_2.CHP	2,57	1,39	0,26	2	6,46	0,9	0,79	0,05	0,065	0,015
CAR_04_MB4_190805_430_2.CHP	2,4	1,3	2,31	1,04	10,33	1,58	0,09	0,05	0,065	0,015
CAR_05_MB6_190805_430_2.CHP	2,11	1,32	0,63	0,61	27,86	1,38	0,18	0,05	0,065	0,015
CAR_06_MT1_190805_430_2.CHP	2,29	1,23	0,35	2,46	33,86	0,49	1,64	0,05	0,065	0,015
CAR_07_MT_190805_430_2.CHP	1,95	1,16	1,06	0,3	19,17	0,56	0,49	0,05	0,065	0,015
CAR_08_MT3_190805_430_2.CHP	2,34	1,25	2,3	11,72	25,17	1,15	0,39	0,05	0,065	0,015
CAR_09_MT4_190805_430_2.CHP	1,89	1,28	0,13	0,84	18,21	0,78	0,56	0,05	0,065	0,015
CAR_10_MT2_190805_430_2.CHP	2,37	1,18	1,17	9,14	30,6	0,51	1,05	0,05	0,065	0,015
CAR_11_R5_1_240805_430_2.CHP	2,38	1,16	16,68	14	6,93	7,19	0,19	0,05	0,065	0,015
CAR_12_R5_3_240805_430_2.CHP	2,37	1,21	14,75	49,1	15,92	17,66	0,31	0,05	0,065	0,015
CAR_13_R5_5_240805_430_2.CHP	2,42	1,24	21,04	13,27	13,45	7,87	0,67	0,05	0,065	0,015
CAR_14_R5_1807_240805_430_2.CHP	2,78	1,29	24,49	9,26	9,92	6,76	1,55	0,05	0,065	0,015
CAR_15_R5_10_240805_430_2.CHP	2,87	1,18	19,31	12,92	7,44	5,85	0,1	0,05	0,065	0,015
CAR_16_R4p1_240805_430_2.CHP	2,82	1,21	22,2	10,09	6,73	18,36	0,79	0,05	0,065	0,015
CAR_17_R4p2_240805_430_2.CHP	2,63	1,15	18,8	6,34	8,56	3,89	0,06	0,05	0,065	0,015
CAR_18_R4p3_240805_430_2.CHP	2,75	1,21	25,14	14,2	6,87	10,17	0,07	0,05	0,065	0,015
CAR_19_R4p4_240805_430_2.CHP	2,57	1,25	32,03	30,61	6,81	7,43	0,12	0,05	0,065	0,015
CAR_20_R4p1807_240805_430_2.CHP	2,68	1,28	26,04	12,09	12,94	8,28	0,13	0,05	0,065	0,015

Table S2 Correlation between gene - heterochromatin distance and genomic context

		Chromocenter				Periphery			
		Distance in µm		Normalized distance		Distance in µm		Normalized distance	
		MB	MT	MB	MT	MB	MT	MB	MT
	CpG Islands in Gene	-0.23	0.25	-0.35	-0.02	0.11	-0.13	-0.02	-0.11
2-Mbp window	CpG Islands	-0.15	0.37	0.02	0.51	0.45	0.14	0.69	0.33
2-Mbp window	GC Content	0.01	0.44	0.00	0.60	0.61	0.11	0.68	0.49
2-Mbp window	Gene density	-0.2	0.42	-0.2	0.48	0.39	0.21	0.65	0.4
2-Mbp window	SINE	0	0.39	0.26	0.64	0.39	0.19	0.73	0.6
2-Mbp window	LINE	-0.19	-0.36	-0.22	-0.51	-0.38	-0.26	-0.53	-0.53
5-Mbp window	CpG Island	-0.21	0.47	-0.02	0.63	0.61	0.23	0.80	0.39
5-Mbp window	GC Content	-0.06	0.49	-0.1	0.62	0.62	0.14	0.69	0.48
5-Mbp window	Gene density	-0.23	0.49	0.03	0.65	0.44	0.27	0.8	0.48
5-Mbp window	SINE	-0.08	0.51	0.11	0.73	0.57	0.29	0.86	0.63
5-Mbp window	LINE	0.05	-0.45	0.04	-0.55	-0.45	-0.24	-0.59	-0.48

Table S3 Correlation between gene position changes and genomic context during differentiation

		Chromocenter		Periphery	
		Δ Distance in µm	Δ Normalized distance	Δ Distance in µm	Δ Normalized distance
	Gene expression	0.15	0.17	-0.34	-0.67
	CpG Islands in Gene	0.35	0.28	-0.21	-0.14
2-Mbp window	CpG Islands	0.38	0.36	-0.16	-0.24
2-Mbp window	GC Content	0.31	0.44	-0.31	0.00
2-Mbp window	Gene activity	0.21	0.24	-0.31	-0.60
2-Mbp window	Gene density	0.45	0.52	-0.05	-0.09
2-Mbp window	SINE	0.28	0.25	-0.08	0.12
2-Mbp window	LINE	-0.12	-0.19	0.00	-0.22
5-Mbp window	CpG Islands	0.49	0.49	-0.18	-0.27
5-Mbp window	GC Content	0.40	0.54	-0.28	-0.02
5-Mbp window	Gene activity	-0.02	0.18	-0.34	-0.47
5-Mbp window	Gene density	0.52	0.45	-0.03	-0.14
5-Mbp window	SINE	0.43	0.45	-0.10	0.03
5-Mbp window	LINE	-0.36	-0.44	0.06	-0.08

Table S4 Correlation between distances and genomic context relative to MeCP2 levels

	Chromocenter	Periphery

		Distance in µm		Normalized distance		Distance in µm		Normalized distance	
		Low MeCP2	High MeCP2	Low MeCP2	High MeCP2	Low MeCP2	High MeCP2	Low MeCP2	High MeCP2
	CpG Islands in Gene	0.32	0.65	0.51	0.57	-0.45	-0.50	-0.41	-0.50
2-Mbp window	CpG Islands	0.30	-0.15	-0.01	-0.17	0.72	0.70	0.71	0.68
	GC Content	0.15	-0.28	-0.06	-0.32	0.63	0.68	0.61	0.66
	Gene density	0.2	-0.11	-0.07	-0.08	0.65	0.6	0.67	0.57
	SINE	0.3	-0.06	0.00	-0.05	0.66	0.61	0.68	0.6
	LINE	-0.16	0.23	0.08	0.24	-0.65	-0.66	-0.62	-0.63
5-Mbp window	CpG Islands	0.42	-0.06	0.14	-0.08	0.58	0.58	0.52	0.55
	GC Content	0.2	-0.2	-0.01	-0.23	0.59	0.61	0.54	0.58
	Gene density	0.38	0.04	0.13	-0.01	0.66	0.63	0.69	0.61
	SINE	0.47	0.05	0.18	0.05	0.55	0.52	0.51	0.49
	LINE	-0.18	0.2	0.07	0.16	-0.56	-0.53	-0.48	-0.47

Table S5 Correlation between gene position change and genomic context relative to MeCP2 level

		Chromocenter		Periphery	
		Δ Distance in µm	Δ Normalized distance	Δ Distance in µm	Δ Normalized distance
	Gene expression	0.41	0.54	-0.44	-0.35
	CpG Islands in Gene	0.77	0.30	-0.32	-0.37
2-Mbp window	CpG Islands	-0.45	-0.28	0.14	0.02
	GC Content	-0.54	-0.49	0.38	0.28
	Gene activity	0.32	0.06	-0.33	-0.44
	Gene density	-0.32	-0.04	-0.01	-0.22
	SINE	-0.30	-0.09	0.00	-0.19
	LINE	0.48	0.33	-0.21	-0.10
5-Mbp window	CpG Islands	-0.40	-0.34	0.17	0.18
	GC Content	-0.46	-0.40	0.26	0.21
	Gene activity	0.52	0.24	-0.22	-0.24
	Gene density	-0.22	-0.20	0.07	-0.15
	SINE	-0.27	-0.15	0.03	-0.02
	LINE	0.44	0.19	-0.06	-0.05

SUPPLEMENTARY MATERIALS AND METHODS

Pmi28 myoblasts were co-transfected with a mammalian expression construct coding for red fluorescent tagged PCNA (cherry-PCNA) and green fluorescent tagged rat MeCP2 (MeCP2-GFP) (Brero *et al.*, 2005). After transfection (12-24 h) cells were pulse labeled BrdU (20mM) was added to the cells and incubated for 30 minutes. After 30 minutes the media was changed and a 4h chase followed. After fixation BrdU was detected using a mouse monoclonal anti BrdU antibody (Becton Dickinson, Franklin Lakes, NJ USA) as first and anti mouse IgG Alexa Fluor 647 (Invitrogen, Paisley PA4 9RF, UK) as secondary antibody. Using PCNA as a marker we were able to differentiate between early mid and late S-phase (Leonhardt *et al.*, 2000) and with BrdU in addition we could differentiate G1 and G2 from each other (Easwaran *et al.*, 2004). Using this method could separate all different cell cycle stages to eliminate volume changes due to DNA synthesis. All nuclei and chromocenters were segmented as described by in chapter 2.

General Discussion

The aim of this thesis was to test whether proximity to heterochromatin correlates with gene repression.

This hypothesis first arose from observations that genes were silenced when translocated into heterochromatin of the same DNA molecule (Muller, 1930). These studies were later extended to whether there was a general repressive effect of heterochromatin on genes in its proximity independent of being on the same chromosome. Various studies have concluded that heterochromatin is capable of silencing genes in their proximity (Chahrour *et al.*, 2008; Delaire *et al.*, 2004; Sabbattini *et al.*, 2001; Szczerbal *et al.*, 2009). This was the accepted hypothesis until more recent studies presented data that did not support an effect on gene expression (Meaburn & Misteli, 2008; Takizawa *et al.*, 2008). Concomitant to a yet controversial role in gene expression, heterochromatin is highly remodeled in different cell types upon differentiation (Brero *et al.*, 2005; Meshorer & Misteli, 2006). This chromatin reorganization observed during myogenesis can be induced in mouse myoblasts by transient expression of the protein MeCP2 whose concentration increases during myogenesis. Since chromatin reorganization can be induced by the sole expression of one protein this creates the possibility to use MeCP2 expressing cells as a model system to study the influence of chromatin reorganization *per se* on gene expression. Additionally, it raises the question whether chromatin reorganization is enough to lead to a similar gene expression pattern as in differentiated cells.

Both large-scale cell morphology (size and shape) changes as well as chromatin organization changes take place coupled with cellular differentiation and thus may largely influence the outcome of data analysis in previous studies and was not properly accounted for and normalized. To reconsider the role of heterochromatin proximity to other genes in an unbiased way several tools had to be established first.

In the first chapter, I describe the characterization of novel monoclonal anti MeCP2 antibodies. I needed a MeCP2 antibody, as many experiments described in chapter 3 required the simultaneous visualization of MeCP2 protein levels with the localization of specific gene loci by FISH. Although MeCP2 can be expressed as a GFP-tagged protein, the antibody was essential as the FISH procedure destroys the GFP fluorescence. The antibody has the additional advantages of being able to detect the endogenous protein in immune fluorescence (shown in mouse brain), avoiding overexpression issues associated with GFP-tagged proteins. It also works for western blotting (shown in mouse, rat and pig) which might be important for future experiments. To further demonstrate the utility of the new set of antibodies I conducted an X-inactivation study on the brain of a heterozygous MeCP2 knockout mouse. Since the MeCP2 gene lies on the X-chromosome one would expect a balanced ratio of cells expressing the wild type allele of MeCP2 and the knockout allele. However, I could show that two regions of the mouse brain show a strong skewing of X-inactivation favoring wild type MeCP2 expression over

the knockout (73%). This study is especially important for future clinical research as MeCP2 is mutated in the neurological disease Rett syndrome and these antibodies allow the detection of endogenous protein levels in the brain *in situ* as well as *in vitro*. The anti-MeCP2 antibody enabled my studies characterizing the influence of genome architecture on gene expression and position (second half of chapter 3).

The second instrument developed during my doctoral thesis is a computational tool described in chapter 2. After first successful FISH experiments, the lack of proper distance measurement programs became obvious. All publications in the field measured distances on the basis of dividing the nucleus in different concentric shells going from the center to the periphery and counting how often a gene signal appears in which shell. This system had several severe disadvantages for my studies. The first problem lies in the shell system itself which never gives exact distances but the residence probability of genes in this particular shell. Second, the shells are spherical while a nucleus is not. Additionally, some programs require manual thresholding beforehand since they do not perform signal detection themselves leading to a system that is easily biased.

Initially, we developed a program that was only designed to measure distances from an automatically detected gene signal to the nuclear periphery and nearest heterochromatic cluster (chromocenter). A preliminary analysis of gene position and expression using our own program showed a high percentage of increased distances to chromocenters during differentiation. After careful examination of the data, it became apparent that the change from many small to few larger chromocenters would statistically always lead to longer distances. Following the published hypothesis of correlation between gene expression and position, this would mean that most genes in the experiment and during myogenesis were up regulated, which when correlating it with expression data is not the case. We concluded that this observation is most probably an artifact of cell morphology and introduced a novel cell-by-cell normalization procedure to compensate for biological variability. The normalization is based on a reference distance distribution in each cell. This reference distribution is obtained by measuring the distance to chromocenters and the nuclear periphery from 10.000 random points within the nucleus. The distance of the gene to be studied is then set into relation with the reference distribution resulting in a so called quantile. A quantile reflects the fraction of reference distances that are smaller than the measured gene distance. A quantile of 1 means that all reference distances are smaller, reflecting a very high distance of the gene to heterochromatin, a quantile of zero reflects very close proximity to the structure of interest.

To evaluate and validate our normalization procedure, I analyzed the nuclear localization of the pluripotency genes Sox2, Oct4 and Nanog by measuring their distances to heterochromatin in mouse embryonic stem cells (ES), induced pluripotency stem cells (iPS) and mouse embryonic fibroblasts (MEF). The reasoning behind these measurements was that the three pluripotency genes are active in ES and iPS cells and therefore far away from repressive heterochromatin while they are silenced in MEF cells and located close to heterochromatin. I could show that iPS and ES cells are similar in their morphology and heterochromatin distribution, but differ greatly

from MEFs, which would make a direct comparison highly biased and makes normalization necessary. In my measurements in ES and iPS cells, all three genes have a preferred location far from the periphery, reflecting an active gene state. In MEFs, Nanog and Sox2 repositioned closer to the periphery reflecting their silenced state, while Oct4 did not relocate and remained in an internal position. Following the hypothesis of periphery acting as a silencing compartment, this would mean that Oct4 is not silenced in MEF cells. One possible explanation for the diverging gene positioning in the three genes is that an internal localization is necessary but not sufficient for active gene expression. Another interpretation would be that the genomic context of a gene does not allow repositioning. By analyzing the genomic context of the genes it became clear that Oct4 lies in a very gene dense region (75 genes/Mb), whereas Nanog resides in a region of intermediate gene density (33 genes/Mb) and Sox2 in a region of low gene density (six genes/Mb). The very high gene density around Oct4 might restrict its relocation within the nucleus, as at any given time neighboring genes may be active. The internal positioning in the active state may facilitate spatial interactions of the extended transcription network of these genes. Using only these three genes, I could show that the simple correlation of proximity to heterochromatin translating into gene silencing is most probably not true. However, the data suggested that an internal position might be required for active transcription.

The next step was to extend the initial study validating our normalization procedure, to a broader analysis with genes from different pathways.

To facilitate a larger scale analysis, I switched to the established differentiation system of myogenesis. Myogenesis has the advantage over the ES, iPS and MEF system that undifferentiated myoblasts can be easily distinguished from differentiated myotubes by morphology and do not require staining for differentiation markers. During myogenesis, large-scale chromatin reorganization can be observed suggesting the importance of heterochromatin for genome expression. A similar large-scale reorganization can be induced in a dose dependent manner by the methyl CpG binding protein 2 (MeCP2). As a first step, we used expression profiling to identify all genes that are up or down regulated during differentiation or show different levels of expression in myoblasts expressing low or high amounts of MeCP2. The expression profile from myogenesis showed a balanced amount of up and down regulated genes whereas the MeCP2 profile displayed more up then down regulated genes. This was a surprising finding as MeCP2 is usually referred to as a repressor (Nan *et al.*, 1997) but recent data could also show a activating effect of MeCP2 (Chahrour *et al.*, 2008). Fourteen genes were chosen from both profiles from which ten genes were studied in the myogenic differentiation consisting of four regulated, four down regulated and two unchanged genes. The genes were visualized by FISH and the distances to the periphery as well as to chromocenters were measured and normalized using our previously developed tools. In our preliminary study with the three pluripotency genes, I had shown that they all occupied positions in the interior of the nucleus matching their activated state but differed in their position in their silenced state. Therefore, I first analyzed whether the genes favored certain nuclear positions

within one cell type according to their genetic context. The presented data shows that gene position is strongly influenced by the genomic context especially in differentiated myotubes. I could observe positive correlations to gene density, CpG islands, GC content and SINEs. A positive correlation in this context means that a long distance to heterochromatin is accompanied by high gene density, many CpG islands etc. In contrast, LINEs showed a negative correlation to gene position being consistent with their proposed association with more silent gene regions. Therefore, high numbers of LINEs result in shorter distances to heterochromatin. My observations fit into the proposed active gene domains called RIDGEs (regions of increased gene expression) which are defined by high gene density, high GC content, high SINEs and low LINEs and their inactive gene counterparts called anti RIDGEs.

Distances to chromocenters in myoblasts are not influenced by any genomic property observed in this study. In contrast, distances to the periphery show correlations to their genomic context. Since stem cells have a generally more open and active chromatin, this could explain how adult stem cells like myoblasts do not show influence of gene position by the genomic context. Additionally, the data underlines the importance of the distance normalization tool described in chapter 2. The distance in µm of genes to the periphery in myotubes shows no correlation to their genomic context but normalized distances exhibit high correlations. Next I analyzed whether the change in gene position during myogenesis is influenced by gene expression and/or genomic context. Correlating these factors I could determine whether gene upregulation is accompanied by a positive distance change during differentiation (movement away from heterochromatin).

The correlation analysis between gene position changes and expression changes revealed no correlation for distances to chromocenters and a negative correlation for distances to the periphery. These data contradict the general model of the periphery acting as a silencing compartment as the results obtained here suggest the opposite. One possible explanation for this discrepancy is that previous studies neglected the effect of nuclear morphology and were biased by this factor. On the other hand more recent publications gave a more intricate view of the field. In a study conducted by Meaburn and Misteli, only seven out of eleven genes showed a correlation between position and transcription (Meaburn & Misteli, 2008). Other studies tethered genes directly to the lamina and did not obtain any evidence that these genes were silenced (Finlan *et al.*, 2008; Reddy *et al.*, 2008). We therefore hypothesize that the genomic architecture plays a role in genome expression but only on a low level and can be overruled by other factors.

To test this hypothesis, I used low and high MeCP2 expressing myoblasts mimicking the genomic architecture of myoblasts and myotubes, respectively. This analysis was made possible by our newly developed anti-MeCP2 antibodies which in contrast to the commercially available ones worked very well in immuno fluorescence. In the MeCP2 system, I analyzed eight genes, of which four were in common with our myogenesis study enabling us to directly compare the two systems with each other. In general our data show that low and high MeCP2 expressing cells exhibit the same gene positioning correlations as observed for myoblasts in the

myogenic differentiation system. The only exception being CpG islands within the gene which are directly correlated to ectopic MeCP2 expression. This underlines the dominant influence of genomic context and overall expression on gene positioning and its independence of chromatin rearrangements.

Coming back to the initial question I also correlated distance change with expression and the genomic context in MeCP2 expressing cells. Overall, the results obtained for MeCP2 expressing cells differ from the observation in the myogenic differentiation system. As described for gene relocation during myogenesis I see a negative correlation with peripheral gene association and gene activity. But contrary to the myogenic study, gene repositioning to chromocenters is positively correlated with gene activity matching the hypothesis that these can be silencing compartments. These data emphasize that heterochromatin can influence gene expression but also demonstrates that other factors activated during differentiation might overrule its influence.

In summary, I was able to show that heterochromatin-gene distance is strongly influenced by nuclear morphology. This bias can be corrected by our normalization procedure which can be applied in several different cell types (pluripotent embryonic stem cells, induced pluripotency stem cells, fibroblasts, adult muscle stem cells and differentiated myotubes) revealing that gene position is highly influenced by the overall genomic context. The hypothesis of the periphery acting as a silencing compartment is not supported by the data presented. However, gene position changes to chromocenters can be correlated with gene expression if only heterochromatin reorganization is induced.

One has to keep in mind that a big challenge in these studies is the fact that correlation does not necessarily mean causation. An observed correlation between gene position and gene expression might also be a secondary effect since it is not apparent whether a gene first becomes silenced and then is relocated or whether it gets relocated in order to be silenced or if there is no causality at all between those events. In fact, studies by Groudine published both observations. In one case gene relocation preceded the change in expression (Francastel *et al.*, 2001) and in another study gene activation preceded relocation away from the periphery (Ragoczy *et al.*, 2006). Even though heterochromatin is found in many species including mouse, human, *Drosophila* and yeast it is very different in composition and shape. It is a common feature of these species to have heterochromatin concentrated at the nuclear periphery but not to form chromocenters. Most prominent human nuclei do not form chromocenters but contain heterochromatin. Therefore arguing that chromocenters are a main feature of expression control would deprive humans of this important control mechanism. However, using human diploid cells I could show that upon MeCP2 expression a clustering of heterochromatin can also be observed in human cells (this data is part of a publication which can be found in the annex). These data would again argue for a role of heterochromatin even in cells not forming chromocenters.

Even though these data do not support the theory of a general repressive effect on gene expression by heterochromatin, they show new possibilities for the influence of chromatin

architecture on gene expression control. The observed positive effect of the periphery might support another controversial hypothesis, as will be discussed in the perspectives.

REFERENCES

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. *The Journal of cell biology*, **169**, 733-743.

Chahrour, M., Jung, S.Y., Shaw, C., Zhou, X., Wong, S.T., Qin, J. and Zoghbi, H.Y. (2008) MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science*, **320**, 1224-1229.

Delaire, S., Huang, Y.H., Chan, S.W. and Robey, E.A. (2004) Dynamic repositioning of CD4 and CD8 genes during T cell development. *The Journal of experimental medicine*, **200**, 1427-1435.

Finlan, L.E., Sproul, D., Thomson, I., Boyle, S., Kerr, E., Perry, P., Ylstra, B., Chubb, J.R. and Bickmore, W.A. (2008) Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet*, **4**, e1000039.

Francastel, C., Magis, W. and Groudine, M. (2001) Nuclear relocation of a transactivator subunit precedes target gene activation. *Proc Natl Acad Sci U S A*, **98**, 12120-12125.

Meaburn, K.J. and Misteli, T. (2008) Locus-specific and activity-independent gene repositioning during early tumorigenesis. *J Cell Biol*, **180**, 39-50.

Meshorer, E. and Misteli, T. (2006) Chromatin in pluripotent embryonic stem cells and differentiation. *Nat Rev Mol Cell Biol*, **7**, 540-546.

Muller, H. (1930) Types of visible variations induced by X-rays in Drosophila. *Journal of Genetics*, **22**, 299-334.

Nan, X., Campoy, F.J. and Bird, A. (1997) MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell*, **88**, 471-481.

Ragoczy, T., Bender, M.A., Telling, A., Byron, R. and Groudine, M. (2006) The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation. *Genes & development*, **20**, 1447-1457.

Reddy, K.L., Zullo, J.M., Bertolino, E. and Singh, H. (2008) Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature*, **452**, 243-247.

Sabbattini, P., Lundgren, M., Georgiou, A., Chow, C., Warnes, G. and Dillon, N. (2001) Binding of Ikaros to the lambda5 promoter silences transcription through a mechanism that does not require heterochromatin formation. *EMBO J*, **20**, 2812-2822.

Szczerbal, I., Foster, H.A. and Bridger, J.M. (2009) The spatial repositioning of adipogenesis genes is correlated with their expression status in a porcine mesenchymal stem cell adipogenesis model system. *Chromosoma*, **118**, 647-663.

Takizawa, T., Meaburn, K.J. and Misteli, T. (2008) The meaning of gene positioning. *Cell*, **135**, 9-13.

Perspectives

The results presented in this thesis lead to new questions in the field of chromatin topology and its influence on gene position and expression. The general hypothesis up to now has been that the nuclear periphery represents a repressive compartment within the nucleus, which our data do not confirm. A much older hypothesis by G. Blobel stated that active genes might be associated with nuclear pores gating the mRNA out of the nucleus. This hypothesis would explain our observation of active genes being closer to the periphery. Whether genes are associated with the peripheral heterochromatin or with the nuclear pores is not apparent from our data. To elucidate this question one would need to mark the nuclear pore complex and the nuclear lamina and establish e.g., via super resolution light microscopy to which a gene is associated.

A potential mechanism by which chromocenters can exert silencing effects could be by acting as trapping device for silencing factors. This would facilitate the control on silencing factor distribution and prevent uncontrolled gene repression. One such a repression factor might be MeCP2 itself. To test this hypothesis, preliminary experiments tracking the movement of single molecules are ongoing. Mouse myoblasts were transiently transfected with MeCP2-GFP and differentiated into myotubes. We could show that single MeCP2-GFP molecules exhibit longer residence time in myotube chromocenters than in the remaining nucleus and in myoblast chromocenters. An important difference between myoblasts and myotubes is their DNA methylation profile. Myoblasts as adult stem cells have very little methylation whereas myotubes have a higher DNA methylation level. One possible speculation is that MeCP2 trapping is due to DNA methylation and, therefore, more predominant in myotubes.

To further dissect the role of chromocenters it is crucial to elucidate their full structure and composition which even after over 100 years after their discovery is still unknown. Therefore it would be very important to elucidate all proteins building up chromocenters and also determine if there are tissue specific differences. We have started this analysis by purifying chromocenters from mouse liver. Using electron microscopy, we were able to show they retain their structure throughout the purification procedure and that we can obtain relatively high purity. The samples were characterized by western blotting and revealed very low impurities for known proteins of other nuclear compartments. The next goal is to produce enough material for mass spectroscopy analysis and further ultrastructural characterization resulting in complete picture of chromocenter composition.

Annex

Human Molecular Genetics, 2011, Vol. 20, No. 21 ***4187–4195***
doi:10.1093/hmg/ddr346
Advance Access published on August 10, 2011

MeCP2 Rett mutations affect large scale chromatin organization

Noopur Agarwal[1,†,‡], Annette Becker[2,†], K. Laurence Jost[1,2], Sebastian Haase[1,¶], Basant K. Thakur[3], Alessandro Brero[1], Tanja Hardt[1,§], Shinichi Kudo[4], Heinrich Leonhardt[5], and M. Cristina Cardoso[1,2,*]

[1]Max Delbrück Center for Molecular Medicine, Berlin 13125, Germany, [2]Department of Biology, Technische Universität Darmstadt, Darmstadt 64287, Germany, [3]Department of Pediatric, Hematology and Oncology, Medical School Hannover, Hannover 30625, Germany, [4]Hokkaido Institute of Public Health, Sapporo 060-0819, Japan and [5]Department of Biology, Ludwig Maximilians University Munich, Planegg-Martinsried 82152, Germany

Rett syndrome is a neurological, X chromosomal-linked disorder associated with mutations in the *MECP2* gene. MeCP2 protein has been proposed to play a role in transcriptional regulation as well as in chromatin architecture. Since MeCP2 mutant cells exhibit surprisingly mild changes in gene expression, we have now explored the possibility that Rett mutations may affect the ability of MeCP2 to bind and organize chromatin. We found that all but one of the 21 missense MeCP2 mutants analyzed accumulated at heterochromatin and about half of them were significantly affected. Furthermore, two-thirds of all mutants showed a significantly decreased ability to cluster heterochromatin. Three mutants containing different proline substitutions (P101H, P101R and P152R) were severely affected only in heterochromatin clustering and located far away from the DNA interface in the MeCP2 methyl-binding domain structure. MeCP2 mutants affected in heterochromatin accumulation further exhibited the shortest residence time on heterochromatin, followed by intermediate binding kinetics for clustering impaired mutants. We propose that different interactions of MeCP2 with methyl cytosines, DNA and likely other heterochromatin proteins are required for MeCP2 function and their dysfunction lead to Rett syndrome.

INTRODUCTION

Rett syndrome (RTT, MIM 312750) is a post-natal neurological disorder, with an incidence of ~1/10 000 female births. The females develop normally until 6–18 months of age, but after that, the growth is drastically slowed down, followed by the development of stereotypical hand movements, autistic behavior, loss of speech and motoric skills, respiratory disorders etc. Mutations in the chromosome Xq28 region corresponding to the *MECP2* gene have been shown to be linked to the disease (1).

MeCP2 recognizes methylated cytosines (5 mC) via a highly conserved methyl-cytosine-binding domain (MBD) and is concentrated in the densely methylated pericentric heterochromatin (2). It has also been shown to contribute to transcriptional regulation via its transcriptional repression domain (TRD) interaction with histone deacetylases (3,4). In addition to its role in regulating gene expression, we have recently shown that the MBD of MeCP2 has the ability to reorganize and cluster heterochromatin *in vivo* (5,6). MeCP2 has accordingly been described to compact nucleosomal arrays *in vitro* (7).

Several types of *MECP2* gene mutations including deletions and also duplications have been found in Rett patients (8,9). Whereas the most common mutations, the missense mutations, are mostly clustered within the MBD (amino acids 78–162), the majority of nonsense mutations occur after the MBD predominantly within the TRD (amino acids 207–310; Fig. 1A). In view of the random nature of X chromosome inactivation in females and, hence, the chimeric pattern of cells expressing

*To whom correspondence should be addressed at: Department of Biology, Technische Universität Darmstadt, Schnittspahnstr. 10, D-64283 Darmstadt, Germany. Tel: +49 6151 162377; Fax: +49 6151 162375; Email: cardoso@bio.tu-darmstadt.de
†These authors contributed equally to this work.
‡Present address: BRIC-Biotech Research & Innovation Center, 2200 Copenhagen, Denmark.
¶Present address: Department of Physics, Freie Universität Berlin, 14195 Berlin, Germany.
§Present address: Medical Proteomics Center, Ruhr University, 44801 Bochum, Germany.

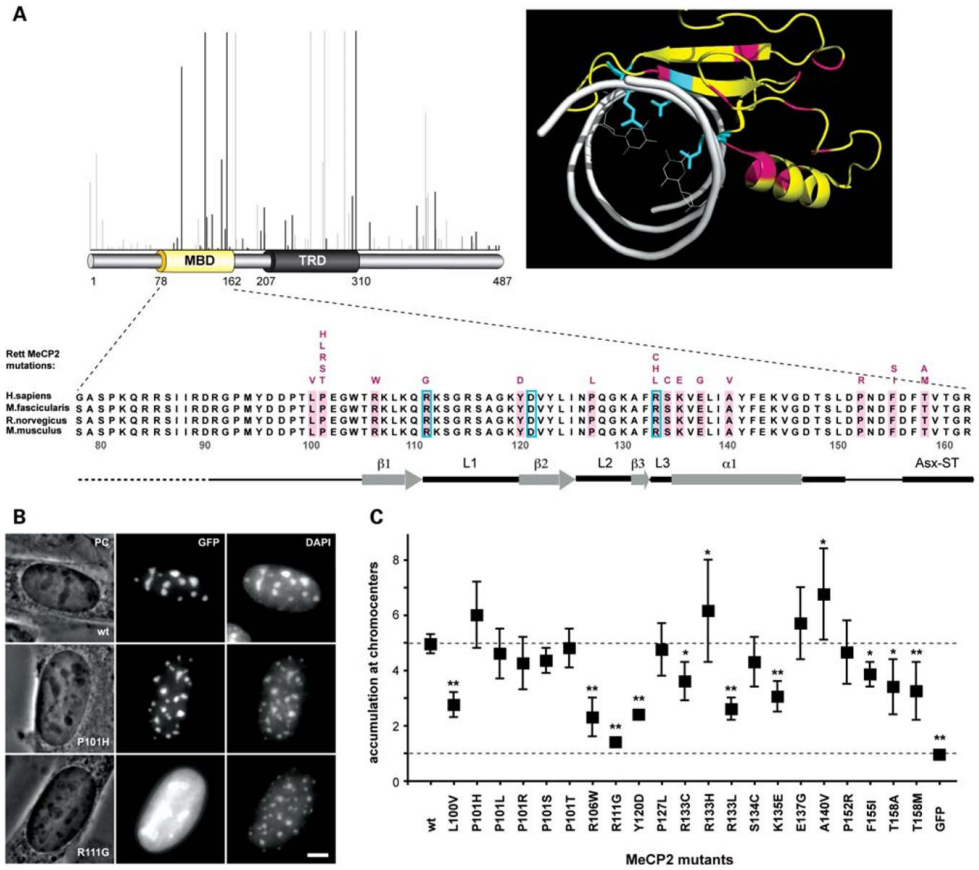


Figure 1. Mutant MeCP2 proteins accumulate at chromocenters *in vivo* to very different extent. (**A**) Top left panel shows the mutation spectrum in Rett patients (IRSA http://mecp2.chw.edu.au/cgi-bin/mecp2/search/printGraph.cgi#MS; last accessed: 15.08.2011), with missense mutations shown in black and others in grey color. Location of individual mutations is indicated in a schematic representation of the MeCP2 protein (numbers are amino acids coordinates). MBD stands for the methyl CpG-binding domain and TRD for the transcription repression domain. Top right panel shows the X-ray structure of the MBD of MeCP2 (displayed in yellow) interacting with its target 5 mC within the DNA double helix (shown in white; PDB accession code 3C2I) (20). Structural data were displayed and annotated using PyMOL software (http://pymol.sourceforge.net/; last accessed: 15.08.2011). The residues that directly interact with the two 5 mC are shown in cyan and the Rett mutations included in our study in pink. Bottom panel: the Rett mutations analyzed are listed (in pink) above the corresponding wild-type amino acid within the sequence of MeCP2 MBD. The location of the defined α-helix (α1), β-strands (β1 and β2) and loops/Asx-ST motif (bold black line) is illustrated below the protein sequence (19,20). (**B**) Representative maximum intensity projections generated from Z-stacks of mouse myoblasts expressing wild-type MeCP2 and mutants thereof. DNA was counterstained with DAPI. PC, phase contrast. Scale bar: 5 μm. (**C**) The plot shows the fold accumulation at chromocenters of 21 Rett mutants, wild-type MeCP2 and GFP in mouse myoblasts. Asterisks represent statistically significant difference in regard to the wild-type: $^{*}P < 0.05$; $^{**}P < 0.001$. All mutants accumulated significantly different ($P \leq 0.05$) with respect to GFP alone (not shown). The experiment was repeated at least two times with 10 cells per mutant evaluated each time.

wild-type or mutant MeCP2, genotype–phenotype correlations are rather complex (10).

Surprisingly, MeCP2 null mice showed unexpectedly mild gene expression changes, strengthening the relevance of other MeCP2 chromatin functions (11). Similar results were obtained from samples of Rett patients (12,13). In this regard, MeCP2 mutations have recently been shown to affect DNA binding, its capacity to induce compaction of nucleosomal arrays (7,14) as well as its chromatin binding kinetics (15,16). However, the mechanism and regulation of MeCP2-induced higher-order heterochromatin organization is still unclear.

In this study, we have systematically characterized 21 missense Rett MeCP2 mutants in terms of their ability to bind and aggregate pericentric heterochromatin and found several which severely affected one or both of these MeCP2 properties.

RESULTS AND DISCUSSION

Based on our finding that the MBD of MeCP2 clusters heterochromatin and most Rett missense mutations affect this domain, we set out to investigate whether they were impaired in binding to and/or clustering heterochromatin.

We selected Pmi28 mouse myoblasts as our cellular assay system. This cell line was used before to characterize the dose-dependent effect of wild-type MeCP2 on the spatial organization of chromocenters and it expresses very low to undetectable level of endogenous wild-type MeCP2 (5). Moreover, it showed a stable and nearly normal karyotype, with minimal variations in chromocenter number caused by variable numerical chromosome aberrations.

We used mammalian expression constructs containing the mutant human MeCP2e2 isoform cDNAs fused at the C-terminus of the enhanced green fluorescent protein (GFP) coding sequence (17). All missense mutations within the MBD are highlighted in pink in Figure 1A, and the mutated forms of the wild-type residues are indicated above the sequence alignment. Intranuclear localization of the fusion proteins and the induction of chromocenter clustering in transfected cells were assessed by epifluorescence microscopy using AT-selective DNA dyes [Hoechst 33258, DAPI (4′-6′-diamidino-2-phenylindol) or TOPRO-3 iodide (4-[3-(3-methyl-2(3H)-benzothiazolylidene)-1-propenyl]-1-[3-(trimethylammonio)propyl]-, diiodide)] to independently visualize pericentric heterochromatin. As MeCP2 strongly accumulates in these heterochromatic regions, we visualize them as a proxy for how (Rett) mutations in the *MECP2* gene can change its ability to associate with chromatin and to reorganize it.

We first tested these MeCP2 Rett mutants for their protein accumulation at the chromocenters by taking a ratio of average mean intensity of protein bound at chromocenters versus nucleoplasm. The results indicate that all the mutant proteins showed enrichment at chromocenters (ratio was >1), but to very different extents (Fig. 1B and C). R111G mutant protein accumulated to the lowest extent. This mutant has been shown before to exhibit complete loss of function of MeCP2 and to no longer repress Sp1-mediated transcriptional activation of methylated and unmethylated promoters (17). We found that it mislocalizes to the nucleoli (phase-dense regions in Fig. 1B) instead of pericentric heterochromatin, which was further confirmed by staining with the nucleolar marker B23 (data not shown). Concomitantly to the lack of heterochromatin association, this mutant depicted an increased nucleoplasmic pool apparent from the high diffuse signal throughout the nucleus (Fig. 1B). F155S exhibited a similar subcellular distribution and a deficit in the heterochromatin accumulation (Supplementary Material, Fig. S1). Except for P101H, R133H, E137G and A140V, all the other analyzed mutant proteins accumulated at chromocenters less than the wild-type, with more than half significantly affected in their accumulation ability when compared with wild-type (Fig. 1C).

Since several mutants associated less efficiently with heterochromatin, we further addressed whether they would be impaired in their ability to cluster heterochromatin *in vivo*. To assess the degree of heterochromatin clustering in a quantitative manner, we scored the number of chromocenters in cells expressing either GFP-tagged wild-type or mutant MeCP2. By this assay, we address the ability of these proteins to reorganize heterochromatin architecture. In Figure 2, the clustering potential of the proteins is displayed as cumulative frequency curves, which represent the percentage of nuclei with up to a certain number of chromocenters. Cells expressing the Rett mutants P101H, P101R and P152R showed a highly significant increase in chromocenter numbers compared with wild-type MeCP2-expressing cells (Fig. 2). R111G and F155S mutants had the most dramatic effect with completely abolished chromocenter clustering (Fig. 2C and Supplementary Material, Fig. S1). Additionally, 10 more mutants exhibited significantly decreased clustering abilities in comparison with wild-type MeCP2 (Fig. 2B). In contrast, the other mutants behaved similarly to the wild-type. Among them is the A140V exchange that has been reported in association with very mild clinical symptoms (18). Altogether, two-thirds of the Rett MeCP2 missense MBD mutants were significantly affected in clustering potential compared with wild-type MeCP2.

We further tested whether the clustering ability of selected mutants was also conserved in human cells. We performed immunostaining in combination with fluorescence *in situ* hybridization using three DNA probes simultaneously to detect the major pericentric heterochromatin regions that are present in chromosomes 1, 9 and 16 (Supplementary Material, Fig. S2A and B) of human cells expressing either wild-type or mutant MeCP2. The outcome of this analysis essentially confirms the results obtained in mouse cells.

Next, we tested whether the clustering of chromocenters generally reflected the amount of protein that accumulated at these regions. Hence, we plotted the median of chromocenter number versus the average accumulation at chromocenters (Fig. 3A). Mutants falling onto an arbitrary line connecting the negative GFP alone control and the positive wild-type MeCP2 control show an inverse correlation between binding to chromocenters and corresponding numbers of chromocenters, i.e. binding less is accompanied by less clustering. The majority of the mutants were both affected in their binding to chromocenters and in clustering chromocenters (Fig. 3A, pink and blue). This could be a consequence of impaired binding affecting their ability to cluster or, alternatively, these mutations could independently affect both properties. Interestingly, some mutants (P101R, P101H and P152R) were mildly to non-affected in heterochromatin binding but severely affected in clustering of chromocenters (Fig. 3A, green). When we applied the same color code to label the corresponding residues in the MBD structure, these residues mapped to the outer part of the MBD (Fig. 3B). Both residues are located adjacent to structured motifs [P101 to β-sheet 1 (β1) and P152 to α-helix 1 (α1)] and within loops (19,20).

As it has been previously shown that some Rett mutations affected chromatin binding kinetics of MeCP2 *in vivo* (15,16), we performed *in situ* detergent extraction as well as fluorescence photobleaching (FRAP) recovery experiments on mutants affected in either chromatin clustering, binding or both. We chose the A140V mutant that was affected neither in binding nor in clustering of chromatin, P101H which was affected only in clustering and R133L affected in both functions (Fig. 4A). The R133L mutation resulted in higher extractability and a much faster FRAP recovery,

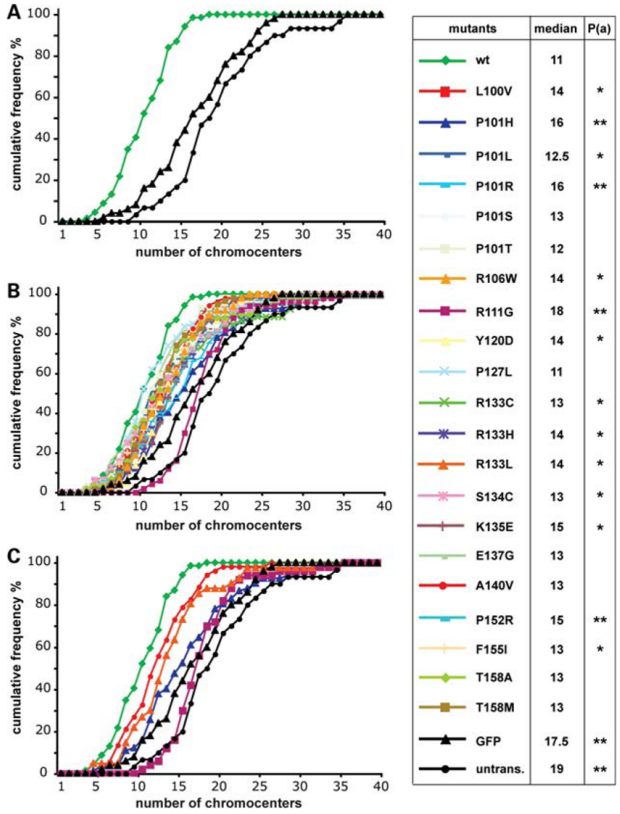


Figure 2. Rett mutant proteins are affected in their ability to cluster chromocenters. Pmi28 mouse myoblasts were transfected with an expression vector coding for GFP or GFP-fused MeCP2 construct as indicated. Z-stacks of images were recorded of nuclei with similar expression levels of the GFP-tagged protein and constant image acquisition parameters. (**A**) The plot shows the percentage of cumulative frequencies of chromocenter numbers in cells expressing GFP-tagged wild-type MeCP2 in comparison to untransfected and GFP-expressing cells. (**B**) Cumulative frequencies of chromocenter numbers in cells expressing each of the 21 GFP-tagged MeCP2 mutants. (**C**) Depicts Rett mutants with extreme phenotypes together with the controls (wild-type MeCP2, GFP and untransfected cells). The table lists the median number of chromocenters for each mutant and depicts the *P*-value with asterisks representing statistically significant difference in regard to the wild-type: $^{*}P < 0.05$; $^{**}P < 0.001$. The experiment was repeated two times with at least 25 cells evaluated per mutant each time.

probably reflecting disruption of binding to 5 mC (Fig. 4B and C). A minimal level of 5 mC seems to be required as a priming event for efficient accumulation of MeCP2 at heterochromatin, as shown by the lack of chromocenter localization of a GFP-tagged MBD fusion in Dnmt1/3a/3b triple knock-out cells (21). Our live-cell kinetic data indeed indicated that although the R133L mutant MeCP2 was still able to accumulate at heterochromatin to a lower extent, it interacted only very transiently and with low affinity. Very similar FRAP kinetics were recently reported for a different mutation of this residue, R133C (15). Both substitutions had comparable heterochromatin clustering potential (Figs 2 and 3), although the R133L exhibited a somehow lower ability to accumulate at heterochromatin (Figs 1 and 3). The *in vivo* accumulation of this mutant at chromatin may be either the result of retaining low-affinity recognition of 5 mC and/or binding to DNA or other heterochromatin-associated proteins. On the other extreme, the A140V mutant protein performed in both assays (*in situ* extraction and FRAP kinetics) as the wild-type (Fig. 4B and C). Importantly, the P101H mutant, which accumulated as the wild-type at heterochromatin but was drastically impaired in clustering chromocenters, had intermediate FRAP kinetics and was also easier to extract from heterochromatin. The FRAP kinetics follow the same trend for the different mutants independently of whether the region photobleached included only chromocenters (Fig. 4C; which measures mostly the contribution of heterochromatin-bound MeCP2) or half of the nucleus (Fig. 4C; with a

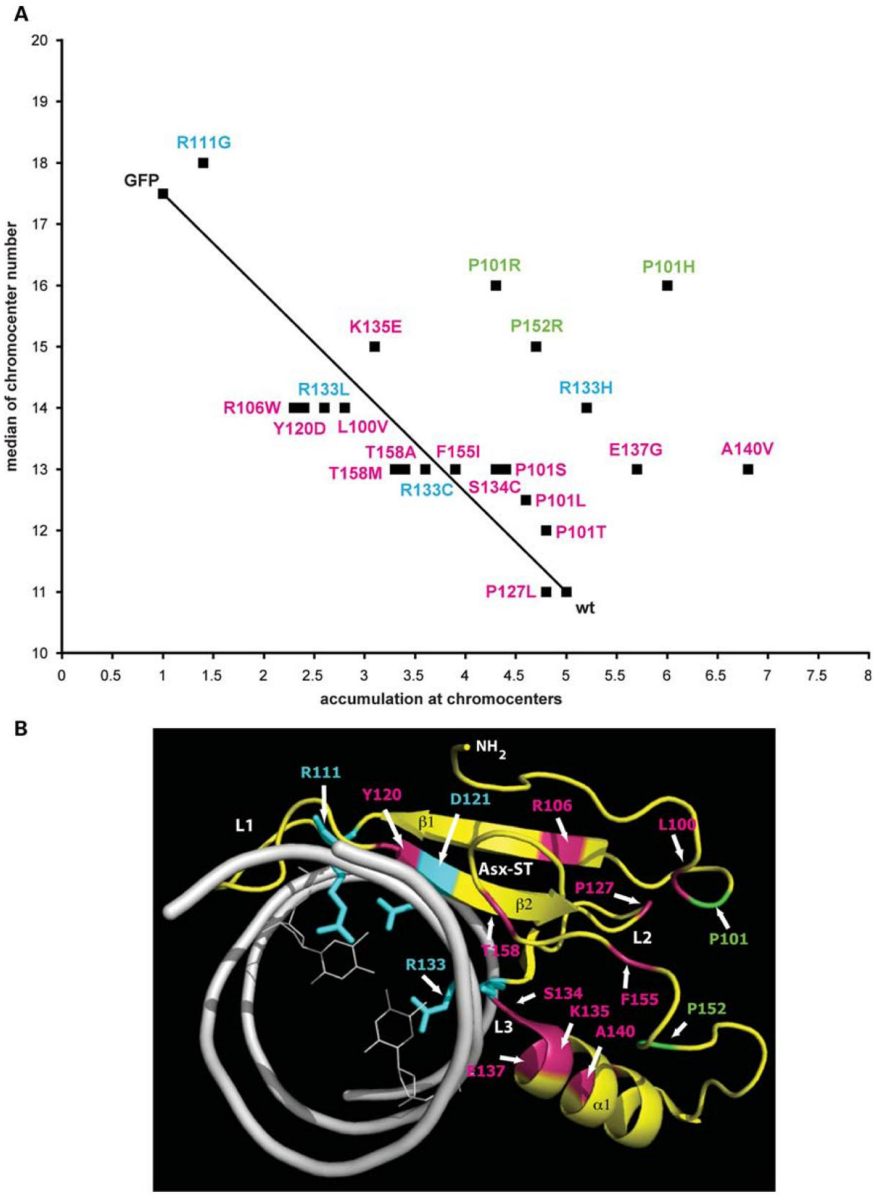


Figure 3. Correlation analysis of chromocenter clustering and accumulation at chromatin. (**A**) Accumulation at chromocenters (Fig. 1) and median of chromocenter number (Fig. 2) were plotted on the *x*- and *y*-axes, respectively. The line connecting the GFP alone and GFP-MeCP2 delineates the inverse relationship between accumulation at chromocenters and chromocenter number (clustering). Mutations of residues directly interacting with 5 mC are shown in blue, and those not directly interacting are illustrated in pink. The green color highlights the drastic examples of chromatin clustering impaired mutants. (**B**) Structure of the MBD (in yellow) of MeCP2 in complex with DNA (in white) is displayed as in Figure 1A, and the residues are color coded as in (A). β-Strands (β1 and β2), α-helix (α1), loops (L1, L2, L3) and the Asx-ST motif are marked. The NH_2-terminal part of the protein structure is indicated.

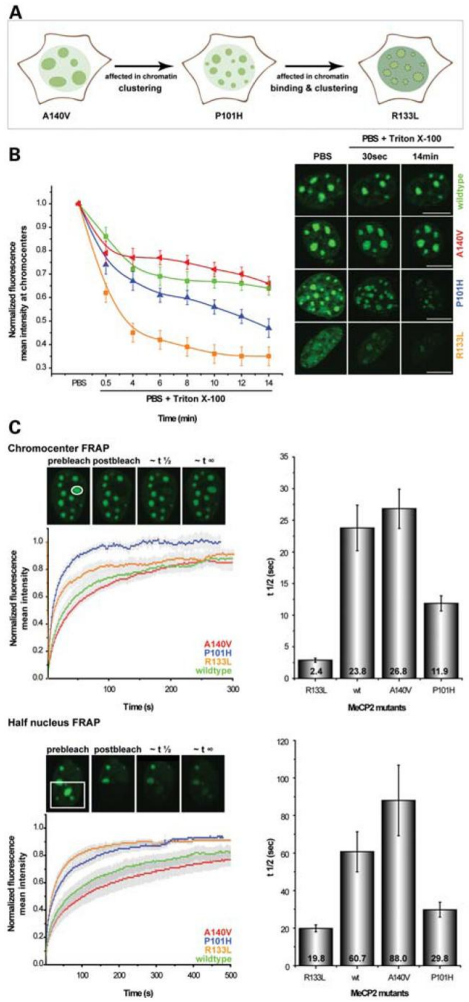


Figure 4. MeCP2 Rett mutant proteins show different kinetics *in vivo*. (**A**) Schematic representation of cells expressing selected Rett mutants defective in binding or clustering of heterochromatin. (**B**) *In situ* extraction kinetics for GFP-tagged proteins was performed by permeabilizing the cells on the microscope stage with Triton X-100 and measuring the decrease in protein at chromocenters over time. The experiment was repeated twice and 7–10 cells were analyzed each time, for each mutant. The line graph shows the extraction kinetics of the mutants over time. Error bars represent the standard error of the mean, and representative mid section images are shown on the right. Scale bar represents 10 μm. (**C**) FRAP curves of GFP-tagged wild-type and mutant MeCP2 together with representative images before and after photobleaching. For FRAP analysis, either a whole chromocenter or half of the nucleus (marked in white) was photobleached. The experiments were repeated two times with 10–20 cells photobleached per construct each time. Results were averaged and the mean curve as well as the standard error of the mean was calculated. Half times of recovery shown on the bar histograms were calculated from the mean curves and the error bars represent the standard error of the mean.

higher contribution of the nucleoplasmic MeCP2 fraction) or was measured in human cells (Supplementary Material, Fig. S2C).

Since the P101 is located far away from the 5 mC interacting pocket, these data suggest that it is primordially involved in connecting chromatin fibers most likely through interactions with other chromatin proteins. From all the MBD residues, the P101 seems particularly sensitive to any substitution (Fig. 1A). Whereas mutations to L, S and T have a mild heterochromatin clustering effect, substitution to positively charged residues (R or H) results in a drastic effect. This residue is located in the NH_2-terminal part of the MBD and likely induces a sharp turn before the two opposing β-sheets (β1 and β2, Fig. 1). Interference with this rigid proline-induced conformation may be more significant on replacement with the not very flexible histidine and less with the other more malleable amino acids. The same could apply for the P152 substitution to the basic residue R.

In summary, our analysis of the *in vivo* chromocenter clustering ability of the different mutations clearly indicated that all mutants where this property was significantly disrupted mapped to the same outer surface of the MBD structure and, significantly, these mutants were not concomitantly affected in chromatin binding *per se*. From the chromocenter accumulation analysis, we conclude that such mutants should be able to bind well to DNA/chromatin *in vivo*, but could be affected in interactions to other heterochromatin proteins and, thus, could not induce chromocenter clustering resulting in faster FRAP kinetics. When compared with other chromatin binders, the FRAP kinetics of wild-type MeCP2 are much faster than the core histone components (22) but close to the kinetics of the linker histone H1 (23,24). Indeed, both MeCP2 and histone H1 compete for binding to nucleosomes *in vitro* (25,26) and bind to the linker DNA (14,27,28). Moreover, MeCP2 is able to condense chromatin *in vitro* to the same level as histone H1 and under physiological salt concentrations (7). These *in vitro* data suggest that MeCP2 can 'cross-link' chromatin fibers together as is the case with the linker histone H1. On the other hand, other stereotypical heterochromatin proteins, such as HP1α, are highly mobile and have a much faster exchange on chromatin (Supplementary Material, Fig. S3) (29–32). We have previously shown that MeCP2 but not HP1α clusters chromatin *in vivo* (5,6) and our present data suggest that mutations occurring in Rett patients are defective in the chromatin architecture function of MeCP2.

We therefore propose that proteins involved in the formation and stabilization of higher-order chromatin structures ('chromatin linkers') bind to chromatin through multiple modes of interactions including, in the case of MeCP2, interactions with 5 mC, DNA and other chromatin proteins and have a longer residence time on chromatin. The latter might also conversely facilitate the establishment of multiple higher-order contacts in a self-reinforcing loop. We suggest that the cooperation of all these binding modes, which individually may have low affinity, promotes ultimately stable association of MeCP2 at heterochromatin, measured in the FRAP and *in situ* extraction experiments. This stable binding could facilitate connections within and between chromatin fibers and lead to a dynamic yet stable organization of heterochromatin

domains with a modulating effect on the level of transcriptional noise. Less stable MeCP2 heterochromatin binding and/or smaller heterochromatin domains within the nucleus could conceivably play a role in Rett syndrome etiology.

MATERIALS AND METHODS

Expression plasmids

Expression vectors encoding GFP-tagged fusions of human wild-type or mutant MeCP2 cDNA cloned into the pEGFP-C1 vector were described before (17) as was GFP-HP1α (32).

Cell culture, transfection and staining

Pmi28 mouse myoblasts were cultured as described (33). Cells were plated on glass coverslips or multiwell dishes (ibidi μ-dishes 8 well; Ibidi GmbH, Munich, Germany) prior to transfection for fixed cell or live cell experiments, respectively. Cells were transfected using TransFectin™ (Bio-Rad, Hercules, CA, USA) following the manufacturer's protocol. Cultures were fixed and DNA stained as described (6). In short, cultures were rinsed in phosphate-buffered saline (PBS) and fixed in 3.7% formaldehyde in PBS. Nuclear DNA was counterstained using TOPRO-3 (Invitrogen, Carlsbad, CA, USA), Hoechst 33258 or DAPI, and samples were mounted in vectashield antifading medium (Vector Laboratories, Burlingame, CA, USA) or moviol.

Microscopy and image analysis

For chromocenter counting, cells were fixed and examined on a Zeiss Axiovert 200 epifluorescence microscope. Image stacks (0.5 μm Z-interval) were acquired with a 63× Plan-Apochromatic NA 1.4 or 40× Plan-Neofluar NA 1.3 oil immersion phase contrast objectives and a PCO Sensicam QE cooled CCD camera. The image stacks were analyzed using a semi-automated approach. For this, we developed a custom application using the priithon platform. Image stacks were treated as three-dimensional volumes and segmented displaying an optical section view and a maximum intensity Z-projection. Nuclei and chromocenters were automatically identified by intensity-based thresholding and implementation of the water algorithm (34). Identified nuclei and chromocenters were outlined and numbered and the performance of the algorithm was controlled by visual inspection using optical section views and maximum intensity Z-projections. Parameters were adjusted to account for different sample brightness and chromocenter density. All intermediate images, parameters and counting results were automatically saved. Cumulative frequencies of chromocenter numbers were tested for statistical significance utilizing the Kolmogorov–Smirnov test.

To assess the chromocenter binding ability, we collected confocal Z stacks (voxel size: 0.05 × 0.05 × 0.3 μm) of formaldehyde-fixed cells expressing similar levels of the GFP fusion protein on either Zeiss LSM510Meta or Leica SP5 confocal microscopes, using 63×/1.4 NA oil objective and 405 nm diode pumped solid state (for Hoechst 33258, DAPI), 488 nm Argon (for GFP) and 633 nm He-Ne (TOPRO-3) laser excitation. Care was taken in selecting the imaging conditions to avoid under- and overexposed pixels, while keeping the imaging conditions constant. Heterochromatic foci were identified by counterstaining with TOPRO-3, Hoechst 33258 or DAPI. Image analysis was performed using ImageJ version 1.38× (http://rsb.info.nih.gov/ij; last accessed: 15.08.2011). The average mean intensity at the chromocenters versus the nucleoplasm was assessed by selecting four regions of equal size in the two compartments, calculating the mean fluorescent intensity in each compartment and then taking the ratio between both. The formula used to calculate the accumulation of MeCP2 and mutants at chromocenters for each construct was:

$$\text{Accumulation at chromocenter} = \frac{\text{average mean intensity at chromocenters}}{\text{average mean intensity in nucleoplasm}}$$

In the case of chromocenter binding assays, statistical significance was checked through the *t*-test.

In situ extraction of GFP-tagged wild-type MeCP2 and MeCP2-bearing mutants was done by transfecting the cells plated on ibidi dishes with the respective construct and extracting them directly on the microscope while imaging. Cells were first washed with PBS containing 0.5 mM $MgCl_2$, 0.5 mM $CaCl_2$ and imaged. Then, the solution was changed to PBS containing 0.5 mM $MgCl_2$, 0.5 mM $CaCl_2$ and 0.5% Triton X-100. Confocal Z-series were recorded over time on a Zeiss LSM510Meta confocal microscope, using 63×/1.4 NA oil objective. The microscope was equipped with a microscope cage incubation chamber (Oko-lab, Ottaviano, Italy) and the temperature was maintained at 37°C. GFP was excited with the 488 nm argon laser line. Confocal Z-stacks were acquired with a frame size of 1024 × 1024 pixels (voxel size: 0.20 × 0.20 and 1.0 μm), at 2 min time intervals for 14 min. Quantitative evaluation was performed using ImageJ. The mean fluorescence intensities at the chromocenters for each cell and time point were calculated for PBS and PBS–Triton X-100. First, using ImageJ 'adjust threshold' plug-in, the chromocenters were identified and then 'create selection' plug-in was used to assess the mean fluorescence intensity only at chromocenters. This procedure was repeated for each cell and time point. The whole data set for each cell was then normalized to the mean fluorescence intensities of the chromocenters before extraction with Triton X-100. The results were evaluated using Microsoft Excel and plotted using Microsoft Excel or Origin 7.5 software (Origin Lab Corp.).

Fluorescence recovery after photobleaching

Live cell imaging and FRAP experiments were performed on an on a LSM510Meta confocal microscope (Zeiss) using a 63×/1.4NA Plan-Apochromat oil immersion objective. The microscope was maintained at 37°C with the help of an Oko-lab cage incubation chamber. Confocal image series were recorded with a frame size of 512 × 512 pixels (pixel size: 60 nm) and at 2 s time intervals; 488 nm argon laser

line (25 mW) was used at 100% transmission to bleach and at 0.05% transmission to record GFP-tagged proteins over time, with the pinhole opened to 3 Airy units. Either a whole chromocenter or half of the nucleus was photobleached, and 5–10 pre-bleach and 250–400 post-bleach frames were recorded for each time-series. Quantitative evaluation was performed using ImageJ and Microsoft Excel. Briefly, the time-series was first corrected for translational movements using 'stackreg' plug-in from ImageJ and the analysis of the FRAP data was performed exactly as described (35).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We are indebted to Ingrid Grunewald and Anne Lehmkuhl for excellent technical support. We are further grateful to Akos Dobay, Andrea Rottach, Garwin Pichler and Tino Dittrich for helpful discussion and advice and to Tom Misteli for the GFP-HP1α expression construct.

Conflict of Interest statement. None declared.

FUNDING

T.H. was supported by the European Union (ESF program). This work was funded by grants of the Deutsche Forschungsgemeinschaft and by the E-Rare EuroRETT network (BMBF) to M.C.C.

REFERENCES

1. Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U. and Zoghbi, H.Y. (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.*, **23**, 185–188.
2. Nan, X., Tate, P., Li, E. and Bird, A. (1996) DNA methylation specifies chromosomal localization of MeCP2. *Mol. Cell. Biol.*, **16**, 414–421.
3. Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J. and Wolffe, A.P. (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.*, **19**, 187–191.
4. Nan, X., Cross, S. and Bird, A. (1998) Gene silencing by methyl-CpG-binding proteins. *Novartis Found. Symp.*, **214**, 6–16.; discussion 16–21, 46–50.
5. 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.
6. Agarwal, N., Hardt, T., Brero, A., Nowak, D., Rothbauer, U., Becker, A., Leonhardt, H. and Cardoso, M.C. (2007) MeCP2 interacts with HP1 and modulates its heterochromatin association during myogenic differentiation. *Nucleic Acids Res.*, **35**, 5402–5408.
7. Georgel, P.T., Horowitz-Scherer, R.A., Adkins, N., Woodcock, C.L., Wade, P.A. and Hansen, J.C. (2003) Chromatin compaction by human MeCP2. Assembly of novel secondary chromatin structures in the absence of DNA methylation. *J. Biol. Chem.*, **278**, 32181–32188.
8. Archer, H.L., Whatley, S.D., Evans, J.C., Ravine, D., Huppke, P., Kerr, A., Bunyan, D., Kerr, B., Sweeney, E., Davies, S.J. *et al.* (2006) Gross rearrangements of the MECP2 gene are found in both classical and atypical Rett syndrome patients. *J. Med. Genet.*, **43**, 451–456.
9. Pan, H., Li, M.R., Nelson, P., Bao, X.H., Wu, X.R. and Yu, S. (2006) Large deletions of the MECP2 gene in Chinese patients with classical Rett syndrome. *Clin. Genet.*, **70**, 418–419.
10. Chahrour, M. and Zoghbi, H.Y. (2007) The story of Rett syndrome: from clinic to neurobiology. *Neuron*, **56**, 422–437.
11. Tudor, M., Akbarian, S., Chen, R.Z. and Jaenisch, R. (2002) Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. *Proc. Natl Acad. Sci. USA*, **99**, 15536–15541.
12. Nectoux, J., Fichou, Y., Rosas-Vargas, H., Cagnard, N., Bahi-Buisson, N., Nusbaum, P., Letourneur, F., Chelly, J. and Bienvenu, T. (2010) Cell cloning-based transcriptome analysis in Rett patients: relevance to the pathogenesis of Rett syndrome of new human MeCP2 target genes. *J. Cell. Mol. Med.*, **14**, 1962–1974.
13. Urdinguio, R.G., Lopez-Serra, L., Lopez-Nieva, P., Alaminos, M., Diaz-Uriarte, R., Fernandez, A.F. and Esteller, M. (2008) Mecp2-null mice provide new neuronal targets for Rett syndrome. *PLoS One*, **3**, e3669.
14. Nikitina, T., Ghosh, R.P., Horowitz-Scherer, R.A., Hansen, J.C., Grigoryev, S.A. and Woodcock, C.L. (2007) MeCP2-chromatin interactions include the formation of chromatosome-like structures and are altered in mutations causing Rett syndrome. *J. Biol. Chem.*, **282**, 28237–28245.
15. Kumar, A., Kamboj, S., Malone, B.M., Kudo, S., Twiss, J.L., Czymmek, K.J., LaSalle, J.M. and Schanen, N.C. (2008) Analysis of protein domains and Rett syndrome mutations indicate that multiple regions influence chromatin-binding dynamics of the chromatin-associated protein MECP2 *in vivo*. *J. Cell Sci.*, **121**, 1128–1137.
16. Marchi, M., Guarda, A., Bergo, A., Landsberger, N., Kilstrup-Nielsen, C., Ratto, G.M. and Costa, M. (2007) Spatio-temporal dynamics and localization of MeCP2 and pathological mutants in living cells. *Epigenetics*, **2**, 187–197.
17. Kudo, S., Nomura, Y., Segawa, M., Fujita, N., Nakao, M., Schanen, C. and Tamura, M. (2003) Heterogeneity in residual function of MeCP2 carrying missense mutations in the methyl CpG binding domain. *J. Med. Genet.*, **40**, 487–493.
18. Orrico, A., Lam, C., Galli, L., Dotti, M.T., Hayek, G., Tong, S.F., Poon, P.M., Zappella, M., Federico, A. and Sorrentino, V. (2000) MECP2 mutation in male patients with non-specific X-linked mental retardation. *FEBS Lett.*, **481**, 285–288.
19. Ohki, I., Shimotake, N., Fujita, N., Nakao, M. and Shirakawa, M. (1999) Solution structure of the methyl-CpG-binding domain of the methylation-dependent transcriptional repressor MBD1. *EMBO J.*, **18**, 6653–6661.
20. 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.
21. 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.
22. Kimura, H. (2005) Histone dynamics in living cells revealed by photobleaching. *DNA Repair (Amst.)*, **4**, 939–950.
23. Misteli, T., Gunjan, A., Hock, R., Bustin, M. and Brown, D.T. (2000) Dynamic binding of histone H1 to chromatin in living cells. *Nature*, **408**, 877–881.
24. Lever, M.A., Th'ng, J.P., Sun, X. and Hendzel, M.J. (2000) Rapid exchange of histone H1.1 on chromatin in living human cells. *Nature*, **408**, 873–876.
25. Ghosh, R.P., Horowitz-Scherer, R.A., Nikitina, T., Shlyakhtenko, L.S. and Woodcock, C.L. (2010) MeCP2 binds cooperatively to its substrate and competes with histone H1 for chromatin binding sites. *Mol. Cell. Biol.*, **30**, 4656–4670.
26. Nan, X., Campoy, F.J. and Bird, A. (1997) MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell*, **88**, 471–481.
27. Ishibashi, T., Thambirajah, A.A. and Ausio, J. (2008) MeCP2 preferentially binds to methylated linker DNA in the absence of the terminal tail of histone H3 and independently of histone acetylation. *FEBS Lett.*, **582**, 1157–1162.

28. Chandler, S.P., Guschin, D., Landsberger, N. and Wolffe, A.P. (1999) The methyl-CpG binding transcriptional repressor MeCP2 stably associates with nucleosomal DNA. *Biochemistry*, **38**, 7008–7018.
29. Schmiedeberg, L., Weisshart, K., Diekmann, S., Meyer Zu Hoerste, G. and Hemmerich, P. (2004) High and low mobility populations of HP1 in heterochromatin of mammalian cells. *Mol. Biol. Cell.*, **15**, 2819–2833.
30. Phair, R.D., Scaffidi, P., Elbi, C., Vecerova, J., Dey, A., Ozato, K., Brown, D.T., Hager, G., Bustin, M. and Misteli, T. (2004) Global nature of dynamic protein–chromatin interactions *in vivo*: three-dimensional genome scanning and dynamic interaction networks of chromatin proteins. *Mol. Cell. Biol.*, **24**, 6393–6402.
31. Festenstein, R., Pagakis, S.N., Hiragami, K., Lyon, D., Verreault, A., Sekkali, B. and Kioussis, D. (2003) Modulation of heterochromatin protein 1 dynamics in primary mammalian cells. *Science*, **299**, 719–721.
32. Cheutin, T., McNairn, A.J., Jenuwein, T., Gilbert, D.M., Singh, P.B. and Misteli, T. (2003) Maintenance of stable heterochromatin domains by dynamic HP1 binding. *Science*, **299**, 721–725.
33. Kaufmann, U., Kirsch, J., Irintchev, A., Wernig, A. and Starzinski-Powitz, A. (1999) The M-cadherin catenin complex interacts with microtubules in skeletal muscle cells: implications for the fusion of myoblasts. *J. Cell Sci.*, **112 (Pt 1)**, 55–68.
34. Harmon, B. and Sedat, J. (2005) Cell-by-cell dissection of gene expression and chromosomal interactions reveals consequences of nuclear reorganization. *PLoS Biol.*, **3**, e67.
35. Schermelleh, L., Haemmer, A., Spada, F., Rosing, N., Meilinger, D., Rothbauer, U., Cardoso, M.C. and Leonhardt, H. (2007) Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Res.*, **35**, 4301–4312.

Supplemental figures:

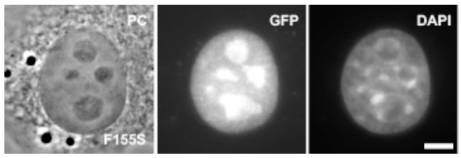


Figure S1: F155S mutant mislocalizes to the nucleolus and is deficient in heterochromatin binding and clustering. Cells were transfected, formaldehyde fixed and DNA was counterstained with DAPI. Representative epifluorescence images of a mouse cell expressing the GFP-tagged MeCP2 mutant are shown. PC: phase contrast. Scale bar: 5 µm.

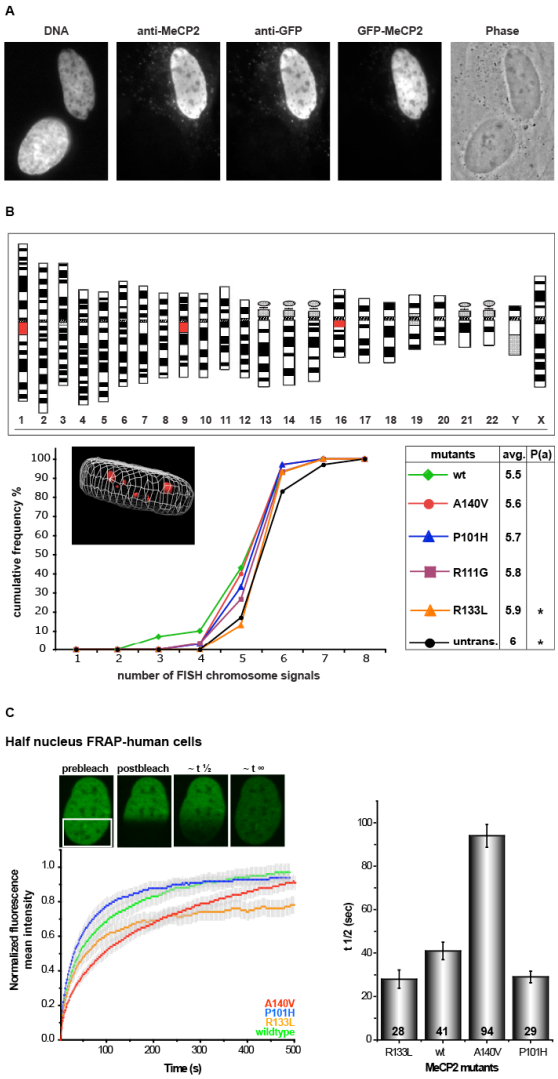


Figure S2: MeCP2 induces heterochromatin clustering in human diploid cells. Human foreskin diploid fibroblasts (Bj-hTERT) were transfected with a plasmid encoding for GFP-tagged human MeCP2, fixed after 12h and immunostained with anti-MeCP2 and anti-GFP. The image shows one exemplary field including one transfected cell identified by direct GFP fluorescence as well as anti-GFP and anti-MeCP2 antibody staining. The second cell was not transfected and hence shows no GFP or anti-GFP signals. The lack of any signal with the MeCP2-specific antibody in the untransfected cells indicates that Bj-hTERT cells, similar to Pmi28 cells, do not contain detectable levels of endogenous MeCP2. (B) Top panel: ideogram of G-banded human chromosomes (www.pathology.washington.edu/galleries/cytogallery/main.php?file=human%20karyotypes). Chromosomes 1, 9 and 16 were selected for our analysis as they contain the largest pericentric heterochromatin regions (marked in red). Bottom panel: Cells were transfected with constructs coding for GFP-tagged wild type and mutant human MeCP2 and clustering of these heterochromatic regions was analyzed by simultaneous hybridization with three DNA probes from the pericentric heterochromatin DNA of the three indicated chromosomes. Cells expressing the GFP-tagged MeCP2 protein were identified by immunostaining with anti-MeCP2 antibody and DNA was counterstained with DAPI. Confocal Z stacks of images from the GFP-MeCP2 signal, overall DNA signal and DNA FISH probes were then acquired. The three dimensional rendering of one such cell is shown where the contour of the nucleus is depicted by the white grid and the FISH signals of the three pericentric heterochromatin regions in red. The cumulative frequency of the

FISH signals counted is shown by the graph. The table lists the average number of chromosome signals and presents the p value (t-test) through asterisks representing statistically significant difference in regard to the wild type: * for P<0.05. Experiments were repeated twice with 30 cells evaluated each time per construct. (C) FRAP and the corresponding kinetic data analysis on the MeCP2 and mutants expressed in human Bj-hTERT cells. Half times of recovery were calculated from the mean curves and are shown in the bar chart. The error bars represent the standard error of the mean. The experiment was repeated twice with 15-20 cells evaluated each time per construct.

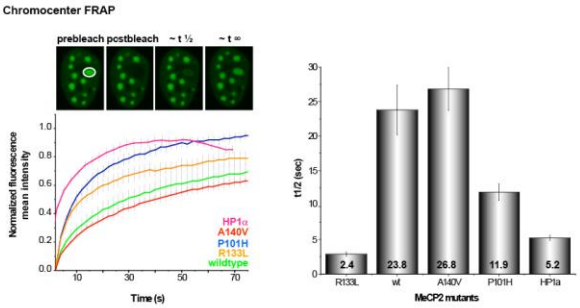


Figure S3: Comparison of heterochromatin binding kinetics of MeCP2 Rett mutants and HP1. FRAP curves of GFP-tagged wild type, mutant MeCP2 and HP1α expressed in Pmi28 cells. Experiments were repeated two times with 10-20 cells photobleached each time. Results were averaged and the mean curve and standard error of the mean were calculated. Half times were determined from the mean curves and are shown in the bar chart together with standard error of the mean.

Supplemental methods:

Human cell culture and transfection

The human foreskin fibroblast (Bj-hTERT) cell line (ATCC BJ-5ta) was derived by transfection of human foreskin fibroblasts with the pGRN145 hTERT expression plasmid and selection of stable immortalized cell clones (1). It is a diploid human cell line with a modal chromosome number of 46 that occurred in 90% of the cells counted and karyotypically normal X and Y sex chromosomes.

Human Bj-hTERT fibroblasts were cultured in DMEM medium containing 10% FCS, glutamine and gentamicin. Cells were transfected using the Amaxa nucleofactor (Amaxa AG, Cologne, Germany) or TransFectin™ (BioRad, Hercules, CA) following the manufacturer's protocols.

ImmunoFISH

For fluorescence *in situ* hybridization, the following DNA probes were used: repetitive specific human DNA probe pUC 1.77 (2) for chromosome 1, alphoid DNA probe pMR9A for the centromeric region 9q12 of chromosome 9 and alphoid DNA probe pHUR-195 for the centromeric region 16q11.2 of chromosome 16. These DNA probes were labeled by standard nick translation with Cy5-dUTP (Amersham, Buckinghamshire, UK). The labeled DNA was further purified by ethanol-precipitation and the pellet resuspended in hybridization solution (70% formamide, 2xSSC, 10% dextran sulfate, pH 7.0). The probes were denatured at 80 °C for 5 minutes.

For immunoFISH cells were fixed with 4% paraformaldehyde in PBS for 10 minutes and

permeabilized with 0.25% Triton X-100 in PBS for another 10 minutes. Primary (rabbit polyclonal anti-MeCP2) and secondary (anti-rabbit IgG Alexa Fluor 568; Molecular probes, CA, USA) antibodies were diluted in PBS with 0.2% fish skin gelatin and incubated sequentially for one hour each at room temperature. After immunostaining, the cells are post-fixed with 4% paraformaldehyde for 60 minutes followed by post-permeabilization with 0.5% Triton X-100 in PBS for 10 minutes, 0.1 M HCl for 10 minutes and 20% glycerol for 4 minutes. Probes were added to the cells and sealed with rubber cement to decrease evaporation of the probe over night. They were then denatured simultaneously at 75 °C for 5 minutes and hybridized over night at 37 °C. Non-hybridized probe was washed off using 50% formamide in SSC at 45 °C three times followed by two washes with 2xSSC. DNA was counterstained with DAPI and the cells were mounted using vectashield.

MeCP2 expressing cells were identified by the positive staining with anti-MeCP2 antibody and complete Z stacks of images (voxel size: 80 x 80 x 200 nm) of the DAPI (excited at 405 nm) and Cy5 (excited at 633 nm) signals for whole DNA and chromosomes 1, 9 and 16 pericentric heterochromatin regions, respectively, were acquired on a Leica SP5 laser scanning microscope using a 63x/1.4NA oil objective.

FISH signals were counted manually through these stacks. 3D rendering was done using UCSF chimera (www.cgl.ucsf.edu/chimera).

Supplemental references:

1 Bodnar, A.G., Ouellette, M., Frolkis, M., Holt, S.E., Chiu, C.P., Morin, G.B., Harley, C.B., Shay, J.W., Lichtsteiner, S. and Wright, W.E. (1998) Extension of life-span by introduction of telomerase into normal human cells. *Science*, **279**, 349-352.

2 Cooke, H.J. and Hindley, J. (1979) Cloning of human satellite III DNA: different components are on different chromosomes. *Nucleic Acids Res*, **6**, 3177-3197.

Abbreviations

5hmC	5-hydroxymethylcytosine
BAC	bacterial artificial chromosome
BC002230	cDNA sequence BC002230
BDNF	brain derived neurotrophic factor
BGT	basic global thresholding
Birc5	baculoviral IAP repeat-containing 5
Brca1	breast cancer 1
BSA	bovine serum albumin
Cdc20	cell division cycle 20 homolog
CenpB	centromere protein B
ChIP	chromatin immunoprecipitation
CoA	Co factor A
Col6a2	collagen type I alpha 2
CT	chromosome territories
DAPI	40,6-diamidino-2-phenylindole
Dnmt	DNA methyltransferase
ELISA	enzyme linked immunosorbent assay
EM	Electron microscopy
ES cells	Embryonic stem cells
FCS	fetal calf serum
FI	fluorescent intensity
FISH	fluorescence *in situ* hybridization
H3K9me3	histone 3 lysine 9 trimethylation
HAT	histone acetyl transferase
HDAC	histone deacetylase
HEK	human embryonic kidney cells
HP1	heterochromatin protein 1
IF	immunofluorescence
IP	immunoprecipitation
iPS	induced pluripotent stem cells
LADs	lamina-associated domains
LINE	long interspersed elements
MB	myoblast
MBDs	methyl-CpG binding proteins
Mbp	mega base pairs
MeCP2	Methyl CpG binding protein 2
Mef2c	myocyte enhancer factor 2C
MT	myotube
Myom2	myomesin 2
Obscn	cytoskeletal calmodulin and titin-interacting RhoGEF
Pc	polycomb protein
PcG	polycomb group proteins
PCNA	proliferating cell nuclear antigen
PEV	Position effect variegation
Prl2c2	prolactin family 2 subfamily c member 2
RIDGE	regions of increased gene expression
RT	room temperature
RTT	Rett syndrome
SINE	short interspersed elements
Sir	silent information regulator
Slc19a2	solute carrier family 19 thiamine transporter member 2
TH	tyrosine hydroxylase
TPEV	telomeric position effect variegation
Tpm3	tropomyosin 3 gamma
TRD	transcriptional repression domain
TSA	trichostatin A
Ttk	Ttk protein kinase
WB	western blot

Acknowledgements

First things first, I would like to thank my Doktormutter M. Cristina Cardoso. For not only giving me the opportunity to carry out my PhD Thesis in her lab but for all her time, discussions, her patience with me and many more things. I will never forget your honesty when you told me after my first month that you would leave Berlin and gave me the opportunity to change groups, I never regretted my decision and I am glad I stayed in your group.

I would like to thank Prof. Gerhard Thiel for agreeing without hesitation to be my second examiner.

For a fantastic working atmosphere I would like to thank the Darmstadt Lab including Alex, Anne, Änne, Annette, Bianca, Britta, Corella, Franzi (Frau Diplom-Biologin Rönicke), Henry, Janine, Jenny, Lena, Malini, Manu, Marius (Nu-ku-lar!), Siyka and Wei. A special thanks goes to the Lunch-Club members for saving me from culinary monotony and for all the fun lunch breaks.

A big thanks also goes to the members of the former Berlin lab. Thank you to Danny, Gilla, Jeff, Marion, Maria, Noopur, Petra and Robert for being great friends and colleagues and for making it hard to leave Berlin. Sebastian Haase I would like to thank for his help in Berlin and Darmstadt and his willingness to pick up the phone at any time.

I would like to thank all the people I worked with outside of the Lab being it in Munich or Berlin: Andrea Rottach, Daniel Smeets, Prof. Hanspeter Herzel, Jörn Schmiedel, Prof. Marion Cremer, Nadine Schrode and Patricia Wolf. A special mention to Prof. Heinrich Leonhardt who has always been helpful with ideas, discussions and connected me with other groups.

I would like to thank Gise, Vadim and Vale for being more than colleagues and becoming my friends.

Ein besonderer Dank geht an meine Familie für ihre Unterstützung, Verständnis und vieles mehr. Ein extra Danke geht an mein persönliches Editoren-Team meine Schwester Christiane und mein Schwager Markus für ihre ehrliche Meinung und Hilfe (und ihr Durchhaltevermögen).

Last but not least möchte ich Martin für vieles aber hauptsächlich dafür danken, mich vor dem allgemeinen Wahnsinn bewahrt zu haben.

Declaration – Ehrenwörtliche Erklärung

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit selbständig angefertigt habe. Sämtliche aus fremden Quellen direkt oder indirekt übernommene Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und noch nicht veröffentlicht.

Darmstadt, den

K. Laurence Jost

Curriculum vitæ

Name:	Katharina Laurence Jost
Address:	Technische Universität Darmstadt
	Schnittspahnstrasse 10
	64287 Darmstadt, Germany
Date of birth:	28.09.1981
Place of birth:	Frankfurt/Main

Education:

2006 – present	PhD thesis
	Max Delbrück Center for Molecular Medicine, Berlin, Germany
	Technische Universität Darmstadt, Darmstadt, Germany
2005 - 2006	Diplom Biology (final grade: 1.1)
	MPI für terrestrische Mikrobiologie
	Marburg, Germany
2003 - 2004	Biology student
	University of Massachusetts Dartmouth, USA
2001 – 2005	Biology student
	Philipps-Universität Marburg, Germany
2001	Abitur (final grade: 2.4)
	Gymnasium Oberursel, Germany

Practicals and teaching:

2009 – present	Supervision of basic practical courses of cell biology, TUD
	Supervision of students during research and diploma work, TUD
March-April 2005	Work in the group of Lotte Soogard-Andersen, at the MPI for terrestrial microbiology, Marburg, Germany
June-July 2004	Work in the group of Elisa Izaurralde at the European Laboratory for molecular Biology (EMBL), Heidelberg, Germany
January 2004	Work in the group of Bal Ram Singh at the Department of Biochemistry, University of Massachusetts Dartmouth, USA

List of Publications

K. L. Jost, A. Rottach, P. Wolf, B. Bertulat, A. Becker, M. Milden, H. Leonhardt, E. Kremmer, M. C. Cardoso "*Generation and characterization of rat and mouse monoclonal antibodies specific for MeCP2*" PLoS ONE (2011) 6(11): e26499. doi:10.1371/journal.pone.0026499

K. L. Jost, S. Haase, D. Smeets, N. Schrode, J. M. Schmiedel, B. Bertulat, H. Herzel, M. Cremer & M. C. Cardoso „*3D-Image analysis platform monitoring relocation of pluripotency genes during reprogramming.*" *Nucl. Acids Res. (2011) 39 (17): e113. doi: 10.1093/nar/gkr486.*

N. Agarwal*, A. Becker*, **K. L. Jost**, A. Brero, T. Hardt, S. Kudo, H. Leonhardt, M. C. Cardoso, "*MeCP2 Rett mutations affect large scale chromatin organization*", *Hum. Mol. Genet. (2011) doi: 10.1093/hmg/ddr346*.

K. L. Jost, B. Bertulat, T. Hardt, A. Brero, C. Goesele, H. Schulz, N. Hubner, M. C. Cardoso "*Role of heterochromatin in gene silencing*" Manuscript in Preparation.

K. L. Jost, F. Rönicke, M. Cardoso
"P*rotein composition and structure of pericentric heterochromatin in mouse*"
Manuscript in Preparation.

Conference contributions

Poster: **3rd International Congress on Stem Cells and Tissue Formation**
Titel: *„Relevance of Heterochromatin Topology in Genome Silencing during Mouse Adult Stem Cell Differentiation.“*
Dresden, Germany; July 11 – 14, 2010.

Talk: **2nd European Working Group on Rett Syndrome**,
Titel: “*Role of MeCP2 Induced Heterochromatin Reorganization in Genome Silencing During Differentiation”* Stresa, Italien, 17 - 19 September 2009.

Poster: **3rd MC-GARD Meeting on Higher Order Genome Architecture.**
Titel: “*Relevance of Heterochromatin Organization in Genome Silencing during Differentiation*“
Edinburgh, Scotland, 1-5 April, 2009.

Poster: **XX International Congress of Genetics**
Titel: “*Role of MeCP2 Induced Heterochromatin Reorganization in Genome Silencing during Differentiation”*
Berlin, Germany; July 12 - 17, 2008.

Index of electronic supplementary material

1. Dissertation PDF: Relevance of heterochromatin topology in genome silencing during (retro) differentiation
2. Publications PDFs:
   2.1. 3D-Image analysis platform monitoring relocation of pluripotency genes during reprogramming.
   2.2. MeCP2 Rett mutations affect large scale chromatin organization.
3. Supplementary movies
   3.1. Randomizer software step-by-step tutorial
   3.2. Random points simulation in 3D.