iPS Cell Based Models of Silent Chromatin and of Gene Expression in Rett Syndrome Neurons

by

Ugljesa Djuric

A thesis submitted in conformity with the requirements for the degree of Doctorate of Philosophy

Department of Molecular Genetics
University of Toronto

© Copyright by Ugljesa Djuric
2014
Abstract

Induced pluripotent stem (iPS) cell technology is an attractive new avenue for studying the reorganization of chromatin during development and for modeling human disease. I demonstrate that the completion of the gradual reprogramming process in murine somatic cells is temporally associated with epigenetic silencing of reprogramming retroviral vectors and acquisition of decondensed constitutive heterochromatin fibres. By specifically investigating H3K9me3-enriched heterochromatin, we provide structural evidence that compact heterochromatin domains are organized as densely packed 10 nm chromatin fibres. To model the effect of mutations in the methyl CpG-binding protein 2 (MECP2) in human neurons, we established iPS cells from a Rett syndrome patient with an exon 1 MECP2 frameshift mutation. Neuronal differentiation specification was unaffected in MECP2 mutant neurons when evaluated by single-cell Fluidigm analyses. Subtle changes in a subset of MECP2 target genes were found implicating MECP2e1 as a transcriptional modulator rather than a master regulator of gene transcription in neurons. Finally, MECP2e1 deficiency is sufficient to reduce neuronal soma-size even in presence of the alternate MECP2e2 isoform demonstrating that MECP2e1 is the dominant endogenous isoform that is essential for healthy neuron function.
In memory of my grandfather Vojislav J. Đurić
Acknowledgements

My growth as a scientist has been an exciting and fulfilling experience through the years. I will cherish the time I have spent in James Ellis’ lab for the rest of my life. I feel that my time here has made me a richer person, both due to the knowledge I have gained and the people I have met that will be a part of my life for the years to come. I am very grateful to James for all the trust that he has had in me during my PhD and the support and guidance he has provided. From our lab manager, Peter Pasceri, and postdoctoral fellow, Akitsu Hotta, I have learned the utmost dedication and attention to detail that is required to carry on a scientific career. To David Bazett-Jones, whose love of the nucleus is only matched by his amazing PhD student, Eden Fussner, working with both of you has been truly an enriching experience in every sense of the word. To Aaron Cheung and Alina Piekna, I will always be thankful for your hard work and contributions throughout the process.

My scientific endeavors have long been a part of my life. Exploration of ancient monasteries with my father and grandfather during my childhood provided a foundation for my quest in the search of the unknown. In my PhD career, analysis of painters’ brush strokes of awe-inspiring frescos of Studenica Monastery was replaced by studying the epigenetic phenomena of pluripotent cells. The appreciation for discovery has been cemented in me since early childhood and it was molded through the years by growing in a family of explorers and artists. From my father, an art historian, to my brother and sister, a trumpet genius and set and costume designer, to my mother, a master architect, I have been lucky enough to have a family environment that always encouraged thinking outside of the box and nurtured the pursuit of the impossible. I love you all dearly and I am very grateful for all the support you have given me.
Table of Contents

Abstract ............................................................................................................................... ii
Acknowledgements ........................................................................................................... iv
Abbreviations ................................................................................................................... viii
List of Figures ................................................................................................................... x
List of Tables ..................................................................................................................... xi

1. Chapter 1: Introduction ............................................................................................... 1
   1.1. Pluripotent Cells and Somatic Cell Reprogramming ............................................... 1
      1.1.1. Embryonic Stem Cells ......................................................................................... 1
      1.1.2. Induced Pluripotency ......................................................................................... 2
         1.1.2.1. Nuclear Reprogramming ................................................................................ 2
         1.1.2.2. Induced Pluripotency by Defined Factors ....................................................... 3
            1.1.2.2.1. Reprogramming Gradient and Partial iPSC Cells ..................................... 5
            1.1.2.2.2. Surrogate Markers of Full Reprogramming in the Human System ........... 7
            1.1.2.2.3. Tracking Pluripotency Using Fluorescent Reporters .................................. 8
         1.1.2.3. Applications of iPSC Cell Technology ............................................................ 9
      1.1.3. Epigenetic Organization in Pluripotent Cells ..................................................... 10
         1.1.3.1. Principles of Epigenetics .............................................................................. 10
         1.1.3.2. DNA Methylation ......................................................................................... 13
         1.1.3.3. H3K4 and H3K27 Methylation ....................................................................... 13
         1.1.3.4. H3K9 Methylation ....................................................................................... 14
         1.1.3.5. Readers of the Epigenetic Code .................................................................... 15
         1.1.3.6. Core Pluripotency Genes Activate and Repress Transcription ....................... 16
         1.1.3.7. Core Pluripotency Genes Target Heterochromatin ......................................... 17
         1.1.3.8. Pluripotency Gene Variegation ..................................................................... 18
         1.1.3.9. Epigenetic Silencing of Retroviruses .............................................................. 19
         1.1.3.10. Small Molecule Treatments Increase Reprogramming Efficiencies .......... 20
      1.1.4. Heterochromatin Features in Pluripotent Cells ................................................ 21
         1.1.4.1. Visualizing Chromatin Ultrastructure .......................................................... 21
         1.1.4.2. Heterochromatin in Pluripotent Cells is “Open” ........................................... 22
      1.1.5. Epigenetic Reorganization During Reprogramming ......................................... 23
   1.2. Modelling Rett Syndrome Using iPSC Technology .................................................. 25
      1.2.1. Disease Modelling Using iPSC Cell Technology ................................................. 25
      1.2.2. Rett Syndrome ................................................................................................ 26
         1.2.2.1. Clinical Features of Rett Syndrome ............................................................. 26
         1.2.2.2. MECP2 Mutations Cause RTT ................................................................. 27
      1.2.3. MECP2 Structure and Function ....................................................................... 28
         1.2.3.1. Distinct roles of MECP2 isoforms ............................................................... 30
      1.2.4. iPSC Cell Based Studies of RTT ....................................................................... 31
   1.3. Summary and Hypothesis of Thesis ....................................................................... 32

2. Chapter 2: Constitutive Heterochromatin Reorganization During Somatic Cell
   Reprogramming ............................................................................................................ 34
   2.1. Abstract ................................................................................................................... 34
   2.2. Introduction ........................................................................................................... 35
   2.3. Results .................................................................................................................. 39
      2.3.1. Distinction of Partial from Full iPSC Cell Lines ................................................. 39
      2.3.2. Chromocentre Compartment Differences Between Partial and Full iPSC Cells.... 44
      2.3.3. Chromocentres in Full iPSC Cells are Composed of 10 nm Chromatin Fibres .... 45
3. Chapter 3. The e1 Isoform of MECP2 Controls Gene Expression and Neuronal Morphology .......................................................... 72

3.1. Abstract .................................................................................. 72
3.2. Introduction ............................................................................. 73
3.3. Results .................................................................................... 74
  3.3.1. Characterization of RTTe1 iPS Cell Lines .......................... 74
  3.3.2. Identification of Mutant MECP2e1-expressing Human iPS Cell Lines ................................................................. 77
  3.3.3. Neuronal Differentiation and Lentiviral Rescue of RTTe1 Cells ................................................................. 78
  3.3.4. Single Cell Analysis of RTT iPS-derived Neuronal Differentiations ................................................................. 81
  3.3.5. MECP2e1 Affects Transcription Levels of Target Genes ........................................................................ 83
  3.3.6. MECP2e1 Mutation Causes Decreased Soma Size Phenotype in Neurons .................................................. 87

3.4. Discussion ................................................................................ 89
3.5. Materials and Methods .............................................................. 90
  3.5.1. Androgen Receptor Assay .................................................. 90
  3.5.2. Bisulfite Sequencing ......................................................... 91
  3.5.3. Generation and Transduction of Lentivirus ........................ 91
  3.5.4. qRT-PCR and cDNA Sequencing .................................... 91
  3.5.5. Immunocytochemistry ..................................................... 92
  3.5.6. Western Blot Analysis ....................................................... 92
  3.5.7. Single-cell Fluidigm Arrays ............................................... 93
  3.5.8. Chromatin Immunoprecipitation ..................................... 94

4. Summary of Significant Findings .................................................... 97

5. Discussion ................................................................................... 99

5.1. Regulation of Silent Chromatin Domains in Pluripotent Cells .......... 99
  5.1.1. Retroviruses are Silenced in Full iPS Cells ......................... 99
    5.1.1.1. Retrovirus Silencing: A Historical Perspective .......... 100
    5.1.1.2. Zfp809/Trim28 Target Exogenous pMX Retroviruses for Silencing ...................................................... 102
    5.1.1.3. Epigenetic Modifications in Exogenous and Endogenous Retroviral Silencing ..................... 104
  5.1.2. Heterochromatin Reorganization in Full iPS Cells ................. 107
5.1.2.1. Sall1 and Nanog Control Heterochromatin Decompaction in iPS Cells ................................................................. 107
5.1.2.2. 10 nm Fibre Configuration of Heterochromatin Domains .................................................................................. 108
5.1.2.3. DAPI Linescans as a Tool for Screening High Quality iPS Cell Lines .......................................................... 109
5.2. MECP2e1 Function in iPS Cell Model of RTT ........................................................................................................... 110
  5.2.1. MECP2e1 Mutation does not Alter Neuronal Identities ........................................................................................ 110
  5.2.1.1. X-Inactivation in Human RTT iPS Cells ............................................................................................................. 111
  5.2.2. MECP2e1 Function as Transcriptional Regulator ................................................................................................. 111
  5.2.3. Comment on iPS Cell Technology for Disease Modeling ..................................................................................... 115
5.3. Future Directions ....................................................................................................................................................... 117
  5.3.1. What are the initiating epigenetic events of pMX silencing? .............................................................................. 117
  5.3.2. What are the pluripotent-specific modes of heterochromatin decompaction? ..................................................... 118
  5.3.3. MECP2e1 Neuronal Cell-Type Specific Function? ................................................................................................. 120
5.4. Concluding Remarks .................................................................................................................................................... 123

6. References ........................................................................................................................................................................ 124
Abbreviations

Adenine- and thymine-rich ........................................ AT-rich
Chromatin immunoprecipitation ................................ ChIP
Conventional electron microscopy ............................ CTEM
4',6-Diamidino-2-phenylindole .............................. DAPI
Deoxyribonucleic acid ........................................ DNA
Electron microscopy .............................................. EM
ETn-Oct4-Sox2 pluripotency reporter vector .......... EOS
Embryonic stem cells ......................................... ESCs
Embryonic stem ................................................ ES
Electron spectroscopic imaging .............................. ESI
Fetal bovine serum ............................................. FBS
Fluorescence in situ hybridization ........................ FISH
Enhanced green fluorescence protein ................. EGFP
Glycogen synthase kinase-3 ................................. GSK3
Tri-methylated histone H3 on lysine residue 36 .......... H3K36me3
Tri-methylated histone H3 on lysine residue 9 .......... H3K9me3
Di-methylated histone H3 on lysine residue 9 .......... H3K9me2
Tri-methylated histone H3 on lysine residue 9 .......... H3K27me3
Tri-methylated histone H4 on lysine residue 20 .......... H4K20me3
Heterochromatin protein 1 ..................................... HP1
Induced pluripotent stem cell ................................. iPSC
Induced pluripotent stem ..................................... iPSC
Kilobase ........................................................... kb
Leukemia inhibitory factor .................................... LIF
Light microscopy ................................................ LM
MEK-GSK inhibitor ............................................. 2i
Methyl CpG binding protein 2 .............................. MECP2
Methyl CpG-binding domain ................................ MBD
Transcriptional repressor domain ........................ TRD
Mouse embryonic fibroblasts ................................. MEFs
Mitogen activated protein/Erk kinase .................. MEK
Phosphate buffer saline ....................................... PBS
Paraformaldehyde .............................................. PFA
Phosphate-buffered saline .................................... PBS
Quantitative polymerase chain reaction .......... qPCR
Qualitative reverse transcriptase polymerase chain reaction .... qRT-PCR
Transmission electron microscopy ......................... TEM
Tripartite motif containing 28 ............................. TRIM28
Zinc finger protein 809 ...................................... ZFP809
Suppressor of variegation 3-9 .............................. Suv39
Long terminal repeat ........................................... LTR
Intracisternal A particle ....................................... IAP
Endogenous retrovirus ........................................ ERV
Murine stem cell virus ......................................... MSCV
<table>
<thead>
<tr>
<th>Term</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endoribonuclease-prepared RNA</td>
<td>esiRNA</td>
</tr>
<tr>
<td>Histone deacetylase</td>
<td>HDAC</td>
</tr>
<tr>
<td>DNA methyltransferase</td>
<td>DNMT</td>
</tr>
<tr>
<td>Histone acetyltransferase</td>
<td>HAT</td>
</tr>
<tr>
<td>Rett syndrome</td>
<td>RTT</td>
</tr>
<tr>
<td>DNA methyltransferase</td>
<td>DNMT</td>
</tr>
<tr>
<td>Chromatin immunoprecipitation</td>
<td>ChIP</td>
</tr>
<tr>
<td>Transcription factor</td>
<td>TF</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1.1. Technical steps of iPS cell reprogramming ...........................................4
Figure 1.2. Features of partial and full iPS cells ......................................................7
Figure 1.3. Lentiviral EOS pluripotency reporter vector .........................................9
Figure 1.4. Schematic of common silent and active epigenetic modifications ............12
Figure 1.5. Hallmark epigenetic changes during somatic cell reprogramming ..........24
Figure 1.6. Schematic of MECP2e1 and e2 isoforms .............................................31
Figure 2.1. Characterization of induced pluripotent stem cells ..........................40
Figure 2.2. Characterization of iPS cell lines .......................................................42
Figure 2.3. Microarray analysis of full and partial iPS cells .................................43
Figure 2.4. Chromocentre organization differs between partial and full iPS cells ....45
Figure 2.5. Correlative LM/ESI with indirectly labeled H3K9me3 identifies chromocentres composed of 10 nm chromatin fibres ..................................................48
Figure 2.6. Sorting and characterization of 2i-treated EOS3F-24 iPS cells ............50
Figure 2.7. 2i Treatment converts partial EOS3F-24 cells to a full iPS cell state ....52
Figure 2.8. Retroviral silencing occurs in full iPS cells following 2i conversion ....53
Figure 2.9. 2i-mediated conversion of partial iPS cells results in H3K9me3-marked chromocentre decompaction ....................................................54
Figure 2.10. Nanog protein levels impact heterochromatin compaction in J1 ES cells ....56
Figure 2.11. Heterochromatin reorganizes in Nanog-high cells in vivo in mouse blastocysts ..............................................................58
Figure 2.12. Nanog overexpression in partial EOS3F-24 iPS cells .................60
Figure 2.13. Gene expression analysis following Nanog induction in partial iPS cells ....61
Figure 2.14. Nanog and Sall1 co-operate to bind major satellite repeats .........64
Figure 3.1. Androgen receptor methylation screen of RTTe1 iPS cell lines ..........75
Figure 3.2. Pluripotency characterization of RTTe1 iPS cells ..............................76
Figure 3.3. RTTe1 iPSCs and neurons retain an Xi and exclusively express the mutant MECP2e1 allele ...........................................................78
Figure 3.4. MECP2e1 rescue constructs specifically express in mature neurons following differentiation ..................................................80
Figure 3.5. RTTe1 neurons maintain normal balance of neuronal identity ............82
Figure 3.6. MECP2-mutant neurons maintain normal balance of neuronal identity ....83
Figure 3.7. Transcriptional consequences of MECP2 target genes upon neuronal differentiation of RTTe1 and Δ3-4 iPS cells ................................85
Figure 3.8. MECP2e1 mutation results in absence of MECP2 association with its genomic targets ..........................................................86
Figure 3.9. RTTe1 neurons exhibit a soma size defect that can be rescued in a cell autonomous manner .................................................88
Figure 5.1. Zfp809-mediated retroviral silencing in ES and full iPS cells .........103
Figure 5.2. 2i-Mediated conversion system for finding initiating epigenetic events in retroviral silencing ........................................118
Figure 5.3. Assessing MECP2e1 roles in transcriptional regulation ..................122
## List of Tables

Table 2.1. Primers used in chapter 2............................................................................................................. 70  
Table 3.1. Primers used in chapter 3............................................................................................................. 94  
Table 3.2 Calculation of AR assay, raw data................................................................................................. 95  
Table 3.3 Antibodies used in chapter 3........................................................................................................ 96
Chapter 1: Introduction

This chapter contains modified excerpts from the published review:

1.1. Pluripotent Cells and Somatic Cell Reprogramming

1.1.1. Embryonic Stem Cells

Embryonic stem cells (ESCs) are derived from the inner cell mass (ICM) of the mouse blastocyst at day 3.5 (e3.5) of development (Evans & Kaufman, 1981; Martin, 1981). ESCs have the unique ability to self-renew indefinitely in the presence of leukemia inhibitory factor (LIF) (Gearing et al, 1987; Smith et al, 1988; Williams et al, 1988) and differentiate into any embryonic cell type under appropriate differentiation conditions (Beddington & Robertson, 1989). The ability to temporally control their differentiation state in culture makes ESCs an ideal model system to study gene expression and chromatin organization events related to pluripotent and lineage committed cellular states. Expression of distinct hallmark genes distinguish different layers of e3.5 blastocyst including Cdx2, Gata6 and Nanog, which are restricted to the trophoblast, hypoblast and epiblast regions of the blastocyst, respectively (reviewed in (Rossant, 2008)). The pattern of expression of these genes is maintained in the in vitro counterparts of each cell layer with ESCs expressing high levels of the Nanog transcription factor (Chambers et al, 2003; Mitsui et al, 2003). The ESC transcriptional network, composed at its core by transcription factors Klf4, Oct4, Sox2 and Nanog, governs cellular identity in both mouse and human ES cell systems with various pluripotency-related genes involved in establishment, maintenance and exit from the pluripotent state upon differentiation (Boyer et al, 2005; Chew et al, 2005). Although cultured ESCs are pluripotent
in that they can contribute to all tissues of a developing embryo (Nagy et al., 1993), ESCs are in fact a tissue culture phenomenon, where differentiation cues that normally signal to the pluripotent ICM cells have been removed and pluripotency is maintained in the presence of LIF cytokine. Extensive studies of ES cells have greatly enhanced our understanding of development and deciphering their molecular blueprint has provided the groundwork for ingenious methods of generating pluripotent cells from somatic cell sources.

1.1.2. **Induced Pluripotency**

1.1.2.1. **Nuclear Reprogramming**

Historic nuclear transfer experiments demonstrated in the 1950s and 60s that amphibian nuclei of somatic cells, when placed into enucleated fertilized oocytes, can give rise to viable offspring (Briggs & King, 1952; Gurdon, 1962a; Gurdon, 1962b). Confirmation of these findings in the mammalian system came with cloning of Dolly the sheep (Wilmut et al., 1997) firmly establishing that developmental processes, which naturally occur in a unidirectional fashion, are not irreversible and genomes of differentiated cells remain capable of returning to a pluripotent cell state. Individual proteins required for reprogramming of transplanted nuclei seemed to reside within recipient oocytes and cell fusion experiments of pluripotent with differentiated cells solidified expectations that nuclear transcription factors (TFs) are responsible for guiding the “reprogramming” processes (Cowan et al., 2005; Miller & Ruddle, 1976; Tada et al., 1997). This was followed by demonstrations that introduction of single master transcription factors could induce cell fate changes where MyoD was used to generate myogenic cells (Davis et al., 1987), Pax5 can make hematopoietic cells (Nutt et al., 1999) and Cebpα can convert B and T cells to macrophages (Xie et al., 2004).
1.1.2.2. Induced Pluripotency by Defined Factors

For a long time, a defined factor reprogramming approach to establish the pluripotent state was thought to be unachievable due to the extent of molecular changes that had to take place to revert somatic to pluripotent cells. However, by screening combinations of 24 core ESC-related TFs Takahashi and Yamanaka finally succeeded in converting mouse embryonic fibroblasts (MEFs) to ESC-like cells, called induced pluripotent (iPS) cells, in a series of Nobel-prize winning experiments (Takahashi & Yamanaka, 2006). Through the use of cells with the βGeo knockin in the endogenous pluripotency-related Fbx15 locus, retroviral delivery of four TFs, c-Myc, Oct-4, Klf4 and Sox2, now known as “Yamanaka” factors, was found to be sufficient for MEFs to acquire pluripotency. These iPS cells gain ESC-like gene expression profiles, growth rates, ability to self renew and differentiate into cells of multiple lineages. Defined factor reprogramming was quickly replicated in cells of human origin (Nakagawa et al, 2007; Takahashi et al, 2007; Yu et al, 2007) and in somatic cells of all three germ layers including B lymphocytes (Hanna et al, 2008), liver and stomach (Aoi et al, 2008), pancreas (Stadtfeld et al, 2008a) and brain (Kim et al, 2008) demonstrating that cell types of all three germ layers maintain developmental reversibility. A battery of molecular characterization assays have emerged to assess similarities of iPS cells to ES cell counterparts, including transcriptomic and epigenetic analysis, as well as evaluations of differentiation potential through in vitro directed differentiations and in vivo teratoma assays (Figure 1.1). In the mouse system, the gold standard of pluripotency demonstration is that the cells are capable
iPS cell reprogramming is induced by four “Yamanaka” factors Oct4, Sox2, Klf4 and cMyc. Upon establishment of iPS cell lines through manual picking and expansion of individual colonies, epigenetic features, gene expression profiles and differentiation abilities are compared to ES cells.

However, the initial iPS cell reports failed to generate chimeras that developed to term and it was not until the use of an improved selection strategy for expression of endogenous Nanog or Oct-4 that iPS cells were shown to successfully generate live-born chimeras (Maherali et al, 2007; Okita et al, 2007; Wernig et al, 2007) and subsequently an “all-iPS-cell” mouse was made using the tetraploid complementation technique (Kang et al, 2009).
1.1.2.2.1. Reprogramming Gradient and Partial iPS Cells

Since the initial reports of retroviral-based iPS cell generation, alternative delivery methods of Yamanaka factors have been utilized for inducing the pluripotent state including lentiviral vectors (Sommer et al, 2009), inducible lentiviral vectors (Brambrink et al, 2008), removable transposons (Kaji et al, 2009), adenoviral vectors (Stadtfeld et al, 2008c), as well as protein transduction (Zhou et al, 2009) and RNA transfection (Warren et al, 2010). The reprogramming process proceeds over a period of 5 to 21 days depending on the reprogramming technique but is always characterized by a gradual change in molecular topographies with only subsets of cells undergoing reprogramming capable of overcoming every obstacle along the way (Brambrink et al, 2008; Polo et al, 2012; Stadtfeld et al, 2008b). The heterogeneous nature of reprogramming is initially manifest by downregulation of MEF-specific cell surface marker Thy1 in a large percentage of cells, followed by activation of pluripotent-specific Ssea1 and finally full activation of the pluripotent-specific transcriptional profile only in a subset of cells (Brambrink et al, 2008; Polo et al, 2012; Stadtfeld et al, 2008b). At the transcriptional level, early events in reprogramming are characterized by the downregulation of fibroblast-specific mesenchymal genes and a concomitant activation of epithelial genes in a mesenchymal-to-epithelial transition (MET) (Li et al, 2010; Samavarchi-Tehrani et al, 2010). Subsequent gradual activation of pluripotency genes culminates in the establishment of a stable pluripotent state (Brambrink et al, 2008; Li et al, 2010; Mikkelsen et al, 2008; Samavarchi-Tehrani et al, 2010; Sridharan et al, 2009; Stadtfeld et al, 2008b). However, stochastic events associated with this phase of reprogramming result in common isolation of two distinct pluripotent cell types. These two iPS cell types include those with robust activation of pluripotency transcriptomes that are capable of contributing to live-born
chimeras, termed full iPS cells, and those with limited differentiation potential, referred to as partial iPS cells (Figure 1.2). When retroviral-based reprogramming methods are used, partially reprogrammed iPS cells sustain their pluripotent state through maintained expression of retroviral transgenes (Mikkelsen et al, 2008; Silva et al, 2008; Stadtfeld et al, 2008b; Takahashi & Yamanaka, 2006), in the absence of endogenous pluripotency gene activation. This persistent retroviral expression is rarely observed upon endogenous Sox2 (Stadtfeld et al, 2008b) or Nanog (Nakagawa et al, 2007; Zhao et al, 2008) activation suggesting that it is a late event in reprogramming that coincides with the activation of endogenous pluripotency genes. As such retroviral silencing serves as an additional molecular mark of high quality iPS cell lines and a temporal indicator of the fully reprogrammed state. Although non-transcribed retroviruses mark fully reprogrammed iPS cells, the epigenetic modifications associated with silent retroviruses in iPS cell systems have not been characterized (retroviral silencing in ES cells is discussed further in section 1.1.3.9). These distinctions between partial and full iPS cells make them a useful tool in identifying chromatin organization events associated with acquisition of pluripotency.
Figure 1.2. Features of partial and full iPS cells.

Early phase of the reprogramming process is characterized by the mesenchymal-to-epithelial transition with subsequent establishment of either partial or full iPS cell lines with variable similarities to *bona fide* ES cells. Retroviral silencing is a hallmark feature of fully reprogrammed iPS cells.

1.1.2.2. Surrogate Markers of Full Reprogramming in the Human System

Markers of complete reprogramming in human cells share some but not all features of mouse cells. Given that human iPS cells cannot be subjected to chimera formation assays, establishing definitive molecular characteristics that distinguish full and partial iPS cells is more challenging. For example, *NANOG* reactivation by itself does not specifically mark full iPS cells, and alternative markers such as *DNMT3B*, *REXI* (Chan et al, 2009) and *hTERT* (Masaki et al, 2007) are better indicators of this developmental stage. A combination of selecting iPS cell colonies of desirable morphology, retroviral silencing (Chan et al, 2009; Papapetrou et al, 2009), gene expression and cell surface marker expression such as Tra-1-60...
(Chan et al, 2009), all need to be taken into consideration when selecting high quality human iPS cell lines.

1.1.2.2.3. Tracking Pluripotency Using Fluorescent Reporters

The use of fluorescent reporters was eloquently used in mouse reprogramming experiments to identify emerging pluripotent cells but knockin approaches are inherently inefficient and cumbersome in human cells. Through the use of Oct4/Sox2 enhancer elements combined with the early transposon (ETn) promoter, our lab designed a lentiviral reporter construct called EOS that is pluripotent-specific and is extinguished upon differentiation of mouse or human pluripotent cells (Hotta et al, 2009) (Figure 1.3). The potential uses of fluorescent reporters of pluripotency during reprogramming are three-fold: 1) To track the pluripotent state in vitro when differentiation leads to silencing of the fluorescent marker 1) Maintained reporter expression enables enrichment of high quality pluripotent iPS cell lines 3) Maintained expression in pluripotent cells that resist differentiation and could be utilized to deliver suicide genes in cell therapy applications where the presence of contaminating pluripotent cells is undesired.
Figure 1.3. Lentiviral EOS pluripotency reporter vector.

EOS contains Oct4/Sox2 binding motifs of pluripotency-related enhancers and early transposon (ETn) promoter which specifically drives expression of EGFP and puromycin-resistance cassettes in pluripotent cells upon reprogramming.

1.1.2.3. Applications of iPS Cell Technology

The ability to generate iPS cells from fibroblasts isolated from skin biopsies of patients carrying a genetic mutation and subsequent derivation of cell types of interest has vast potential for understanding developmental processes as well as disease mechanisms (Cheung et al, 2011; Ebert et al, 2009; Kim et al, 2011; Lee et al, 2009; Lee et al, 2010; Liu et al, 2011; Soldner et al, 2009; Wong et al, 2012). Differentiation of iPS cells in culture can uncover morphological or molecular phenotypes in live cells from a human patient and, as a result, drug screening assays that rescue the observed phenotypes will undoubtedly identify molecules with novel therapeutic applications. Furthermore, genetic rescue of mutations in patient-derived iPS cells may in the future become a cellular source for autologous cell
therapies, where the patient would be receiving his own cells and thus the need for immunosuppression of cell transplant recipients would be unnecessary. However, before such promises become a reality, the similarity of iPS cells to their ES cell counterparts as well as the safety of using tissue culture derived iPS cells for cell therapy purposes need to be rigorously investigated. As the reprogramming cascade is made possible by reconfiguration of cellular transcriptomes, one of the key questions of reprogramming is what mechanisms are involved in the establishment of proper gene expression and chromatin structure reflective of the pluripotent state.

1.1.3. Epigenetic Organization in Pluripotent Cells

1.1.3.1. Principles of Epigenetics

Eukaryotic genomes are packaged into chromatin, a complex of DNA and nucleosomes. A nucleosome is made up of a histone octamer, with two copies of each histone, H2A, H2B, H3 and H4 (Kornberg, 1974; Luger et al, 1997). Nucleosomes exist as mobile units along individual chromosomes with neighbouring nucleosomes separated by linker DNA, which is frequently occupied by linker histone H1. Nucleosomes serve a dual function of enabling extensive packaging of DNA within the confines of the nucleus and serving as structures upon which gene expression regulatory mechanisms can converge. A unifying concept for both of these functions is the phenomenon of epigenetics, loosely defined as heritable modifications of chromatin structure without alterations of DNA sequence. A large group of enzymes, termed “writers”, are responsible for establishing epigenetic modifications and those involved in removing the marks are known as the “erasers”. Epigenetic landscapes are further maintained by proteins that bind existing epigenetic
modifications, known as “readers” of the epigenetic code. Epigenetic modifying enzymes establish DNA methylation marks or an expansive amount of posttranslational modifications at amino acid residues of histone tails protruding out of the nucleosome. These histone marks include acetylation, methylation, phosphorylation, ubiquitination, glycosylation, and ribosylation (reviewed in (Kouzarides, 2007)). Epigenetic modifications can be classified into either active or silent marks based on the transcriptional outcome or structural conformation they elicit on the underlying genomic region (Figure 1). As such, epigenetic pathways are critical determinants of cellular identity and are involved in normal development and in the artificial somatic cell reprogramming process where suppression of various epigenetic modifying enzymes affects reprogramming efficiencies to various degrees (discussed in 1.1.3.10.). Chromatin immunoprecipitation, using antibodies to epigenetic marks of interest, combined with high throughput sequencing (ChIP-Seq) techniques, has remarkably advanced our understanding of connections between epigenetic modifications and pluripotency. ES and iPS cells serve as ideal cellular sources for the study of epigenetic control mechanisms as they express the vast majority of epigenetic modifying enzymes and their ability to differentiate is suitable for investigating developmentally-relevant dynamic epigenetic changes. Below are descriptions of some key features of the pluripotent cell “epigenome”.

Figure 1.4. Schematic of common silent and active epigenetic modifications.

Nucleosomes of silent genomic loci are modified by epigenetic modifying enzymes with histone deacetylases (HDACs) removing acetylation marks from histone tails, histone methyltransferases (HMTs) establishing methylation of histone H3K9 and H3K27. DNA methyltransferases lay down methylation of cytosine residues. Methyl binding proteins (MBP) recognize and bind methylated DNA, HP1 proteins bind H3K9me3 and linker histone H1 are at the exterior of nucleosomes generating compact DNA configuration not compatible with transcription. On the other hand H3K4me3 and acetylated histone tails, with hypomethylated DNA are common features of actively transcribed loci.
1.1.3.2. DNA Methylation

Perhaps the most heavily characterized epigenetic modification is that of DNA methylation. It is established by the de novo DNA methyltransferases Dnmt3a and 3b and is propagated by the maintenance methyltransferase Dnmt1 (Jeltsch, 2006). DNA methylation occurs at the cytosine nucleotide resulting in formation of 5-methylcytosine, most commonly in the context of CpG dinucleotides but non-CpG methylation occurs as well at much lower frequency, particularly in pluripotent cells (Ramsahoye et al, 2000). However, genome-wide DNA methylation patterns in ESCs, in a seemingly anti-intuitive fashion, reveal low levels of DNA methylation at CpG-rich promoters, known as CpG islands (Meissner et al, 2008). DNA methylation patterns inversely correlate with levels of gene expression where pluripotency-related and housekeeping genes are hypomethylated while differentiation-specific genes are hypermethylated at their promoters. Although generally associated with inactive loci, DNA methylation is not required for transcriptional silencing in ESCs as exogenously introduced retroviruses are aggressively silenced in ESCs lacking de novo methyltransferases (Pannell et al, 2000) (mechanisms of retroviral silencing are discussed in section 1.1.3.9.). Importantly, changes in distribution of DNA methylation coincide with establishment of iPS cell state and reprogrammed cells acquire ESC-like DNA methylation patterns (Doi et al, 2009; Maherali et al, 2007).

1.1.3.3. H3K4 and H3K27 Methylation

Formation of silent and active epigenetic modifications at gene promoters and enhancers is the mechanism through which reprogramming establishes the correct array of gene expression representative of a pluripotent cell. This is exemplified by the global
reconfiguration of methylation (me) levels of two lysine residues, at histone H3 lysine 27 (H3K27) and histone H3 lysine 4 (H3K4), catalyzed by histone methyltransferases (HMTs) belonging to polycomb (Pc) and trithorax (Trx) group proteins respectively. Generally, trimethylation of H3K4 (H3K4me3) is established at actively transcribed genes while H3K27me3 marks genes undergoing silencing during reprogramming (Mikkelsen et al, 2008). Further propagation of the silent epigenetic state by H3K27me3 occurs through recruitment of polycomb repressive complex 1 (PRC1) that includes RING1A-B (Ringrose & Paro, 2004), responsible for the ubiquitination of histone H2AK119 (de Napoles et al, 2004). Although H3K4me3 and H3K27me3 usually mark distinct sets of genes, they can also coexist in a “bivalent” state on the same nucleosome in pluripotent cells, specifically at developmentally-regulated genes (Bernstein et al, 2006). As a result, bivalent genes are silent in ES and iPS cells but poised for expression due to the underlying active H3K4me3 modification ensuring their quick reactivation upon differentiation. Aside from marking silent genes, H3K27me3 is globally distributed on the silent X-chromosome contributing to the highly compact structure of the inactive X within nuclei of female cells (Plath et al, 2003).

1.1.3.4. H3K9 Methylation

In addition to H3K27me3, methylation of H3K9 is a modification commonly associated with gene silencing. H3K9me1 and me2 are mostly euchromatic and catalyzed by G9a and GLP proteins (Tachibana et al, 2002; Tachibana et al, 2005) while the Suv39h1/h2 and Setdb1/Eset enzymes utilize these intermediaries as a template for establishing H3K9me3 (Peters et al, 2001). During embryogenesis, Oct4 is rapidly silenced in a H3K9 methylation-dependent manner as early lineage commitment is established (Feldman et al, 2006). Furthermore, removal of H3K9 methylation is a required step in achieving pluripotency.
demonstrating that it serves as a roadblock to activating appropriate gene expression patterns during reprogramming (Chen et al, 2013). Similar to DNA methylation, H3K9me3 is heavily concentrated at transcriptionally silent heterochromatin domains and manifests punctate staining pattern in immunostaining assays (Maison et al, 2002; Martens et al, 2005). As such, it has been extensively used in imaging-based approaches in defining nuclear organization of heterochromatin in pluripotent and lineage committed cells (discussed further in 1.1.4.)

1.1.3.5. **Readers of the Epigenetic Code**

Epigenetic organization involves interplay between a wide range of highly conserved histone modifying enzymes that control gene expression and the underlying chromatin ultrastructure. Silent genomic regions are typically characterized by deacetylated histone H3 and H4, high levels of histone H3K9 and H3K27 methylation and concentrated DNA methylation. In addition to modifications of DNA and histones, methyl-CpG-binding domain (MBD) proteins and heterochromatin protein 1 (HP1) family members provide another level of chromatin compaction by binding to methylated CpG sites and methylated histone H3K9, respectively (Fischle et al, 2005; Lachner et al, 2001). These chromatin modifying proteins are referred to as the readers of the epigenetic code with HP1-mediated compaction of nucleosomes occurring through dimerization of neighbouring H3K9me3-bound HP1 proteins. On the other hand, the MBD protein methyl-CpG binding protein 2 (MeCP2) can bind CpG methyl groups and recruit histone deacetylases to further establish a deacetylated histone state and induce gene silencing (Nan et al, 1998). Finally, linker histone H1 proteins bind to the exterior of nucleosomes at linker DNA and are most abundant in transcriptionally inactive loci (Misteli et al, 2000; Netzer et al, 2006).
1.1.3.6. Core Pluripotency Genes Activate and Repress Transcription

The manner in which epigenetic modifying enzymes identify their genomic targets remains poorly understood but is at least partly dependent on recruitment by transcription factors which directly bind DNA and interact with various chromatin modifiers. ES cells have a transcriptional control mechanism that maintains pluripotency and self-renewal through a transcriptional network governed by pluripotency-specific transcription factors. Through ChIP-Seq, genome-wide distribution of chromatin-bound transcription factors in ES cells has begun to be elucidated revealing a highly complex interaction system of many transcription factors at the centre of which reside the determinant pluripotency genes Nanog, Sox2, Klf4 and Oct4 (Boyer et al, 2005; Loh et al, 2006). Collectively, they maintain the pluripotent state by activating pluripotency genes and repressing lineage commitment genes in ES cells. The ability of these factors to influence gene expression in both directions resides in their collaborative binding to target genes in conjunction with epigenetic modifying enzymes. A subset of gene targets are bound by multiple transcription factors with Oct4 having the highest co-occupancy with Sox2 and Nanog. In fact, majority of loci were found to be occupied by more than 4 transcription factors at once. Their cooperative binding in conjunction with p300, a histone acetyltransferase, in ES cells suggests that these TFs are able to recruit epigenetic modifiers to chromatin regions (Chen et al, 2008). Indeed, knockdown of any one factor results in dissociation of p300 from its targets. Thus, Nanog, Oct4 and Sox2 can modulate gene expression by co-occupying target genes and recruiting a histone acetyltransferase to modify chromatin to a transcriptionally active state.

To exert their transcriptional repressor activity, Oct4 and Nanog assemble into the Nanog/Oct4 deacetylase (NODE) complex with HDAC1/2, Mta1/2 and members of NuRD
and Sin3A chromatin remodeling complexes (Liang et al, 2008). Knockdown of the components of the NODE complex induces ES cell differentiation due to upregulation of developmentally regulated target genes. Furthermore, a member of the NuRD complex, methyl CpG-binding domain protein 3 (Mbd3), is involved in the repression of primitive endoderm and trophoblast specific genes, Gata6 and Cdx2, respectively (Zhu et al, 2009). In addition, Oct4 interacts with SetDB1/Eset, a histone H3K9 methyltransferase, to restrict the expression of extraembryonic lineage genes in the inner cell mass of developing blastocysts (Yuan et al, 2009). Overall, the roles of core TFs in pluriopotent cells are manifest in a context-specific manner where they can play distinct roles as either transcriptional silencers or activators depending on the identity of accompanying epigenetic modifying enzymes that they recruit to their target genes.

1.1.3.7. Core Pluripotency Genes Target Heterochromatin

Although high throughput studies have greatly advanced our understanding of the regulation of the transcriptional circuitry maintaining the pluripotent state, such approaches generally filter out repetitive elements from their analysis, and thus may overlook some additional roles of core pluripotency factors. Even so, ChIP-seq data reveal that a large amount of genomic sites bound by the pluripotency TFs occur outside of annotated gene promoters and enhancer elements suggesting that they may be involved in general chromatin and nuclear organization. For instance, Sall1, a regulator of pluripotency, co-occupies a large number of both repressed and activated gene targets of Nanog (Karantzali et al, 2011) but Sall1 immunostaining demonstrates its pericentric heterochromatin subnuclear localization, with specific binding to major satellite repeats (Netzer et al, 2006; Yamashita et al, 2007).
Since Sall1 cooperates to bind its targets with Nanog, it is possible that this interaction is preserved at heterochromatin-associated Sall1. Intriguingly, Nanog binding is extensive at endogenous repetitive elements, although genomic sequences of pericentric heterochromatin were not investigated, in both mouse and human ES cells (Kunarso et al, 2010). Since transcription factors are rarely found to localize to highly dense heterochromatin foci, an interesting possibility exists that Sall1 in collaboration with other pluripotency genes and chromatin modifying enzymes may play a role in regulating heterochromatin organization.

1.1.3.8. Pluripotency Gene Variegation

Pluripotent stem cells within the ICM and cultured ES cells often display a heterogeneous pattern of expression of several key pluripotency genes, including Nanog, with subsets of cells having high or low protein levels, termed variegation (Chang et al, 2008; Hayashi et al, 2008). Although epigenetic mechanisms regulating Nanog variegation within pluripotent cells have not been assessed, this expression pattern is thought to reflect the requirement of pluripotent cells to rapidly commit to cellular lineages during blastocyst development (Chambers et al, 2007; Kalmar et al, 2009). Furthermore, fluctuations of Nanog expression is correlated with variegated gene expression of other pluripotency genes and is speculated to underlie the general heterogeneous nature of pluripotent cells (Faddah et al, 2013). Variegated Nanog expression is maintained in both ES cells and cells of the ICM and as such it serves as a model to determine chromatin reorganization events within cells undergoing the earliest phases of lineage commitment.
1.1.3.9. **Epigenetic Silencing of Retroviruses**

Retroviruses have long been known to be specifically silenced in pluripotent cells (Teich et al, 1977). Although reprogramming techniques routinely utilize unmodified Moloney Murine Leukemia virus (MMLV) for reprogramming, expression of improved retroviral vectors, with modified LTR promoters of the MMLV virus, also remain markedly lower in embryonic and some adult stem cells compared to differentiated cells (Cherry et al, 2000; Pannell et al, 2000). Our lab and others have described a repressive epigenetic configuration of non-expressing retroviral integration events in ESCs. Silent retroviruses are marked by high levels of linker histone H1, methyl CpG binding protein Mecp2 (Lorincz et al, 2000; Lorincz et al, 2001; Pannell et al, 2000; Yao et al, 2004), dimethylated histone H3K9 and heterochromatin protein Hp1γ (Dong et al, 2008; Wolf et al, 2008). Finally, reduced levels of acetylated H3, but not H4, are observed at silent retroviral transgenes (He et al, 2005; Pannell et al, 2000; Yao et al, 2004). In ESCs integrated MMLV viruses are targeted by the pluripotency-specific factor Zfp809 and Trim28 (Wolf & Goff, 2007; Wolf & Goff, 2009). Subsequent interaction between Trim28 and histone methyltransferases establishes methylation of the H3K9 residue, followed by the propagation of the repressive epigenetic modification by H3K9me3 binding-protein HP1 (See **Figure 4.1**). Overall, repression of retroviral transcription in ES cells requires an interplay of a large group of epigenetic pathways in ES cells and is discussed in further detail in section 5.1. Due to the ESC-specific silencing of retroviral vectors, their use as delivery vehicles of reprogramming factors was a clever exploitation of this feature with the anticipation that once pluripotency was reached the integrated MMLV (in the iPS cell context, pMX nomenclature is used but this virus is highly
similar to the classical MMLV) viruses would be “turned off”. The epigenetic pathways regulating retroviral silencing in iPS cells have not been investigated.

1.1.3.10. Small Molecule Treatments Increase Reprogramming Efficiencies

Generating iPS cells through reprogramming may be a predominantly stochastic event (Hanna et al, 2009), requiring precisely coordinated levels of transgene expression at the right time (Papapetrou et al, 2009; Stadtfeld et al, 2008b). Given their dynamic nature, epigenetic modifications are at least partly responsible for the stochastic features of reprogramming as they are required to establish silencing and activating marks on the correct gene promoters. The key role of epigenetics in reprogramming is convincingly demonstrated by the series of experiments using chemical inhibitors of chromatin remodeling enzymes such as G9a (histone H3K9 methyltransferase) (Shi et al, 2008), histone deacetylases (HDACs) (Huangfu et al, 2008), and DNA methyltransferases (DNMTs) (Mikkelsen et al, 2008; Shi et al, 2008), resulting in higher reprogramming efficiencies in a variety of cell types (Feng et al, 2009; Li & Ding, 2010). However, it is arguable whether increased efficiencies of reprogramming are actually desirable, given that low efficiencies of reprogramming yield sufficient amount of iPS cell lines to pursue downstream applications. The preferable outcome of chemical treatments of cells undergoing reprogramming is instead that they would result in generation of higher quality iPS cells, with simply higher frequency of fully reprogrammed colonies. To this end, treatment with a double chemical inhibitor (2i) cocktail targeting mitogen-activated protein kinase/extracellular signal-related kinase (Mek/Erk) and glycogen synthase kinase-3 (Gsk3) pathways converts mouse partial into full reprogrammed iPS cells (Silva et al, 2008; Sridharan et al, 2009). The ability to temporally regulate the achievement of full reprogramming, or
“ground state pluripotency”, makes the 2i small molecule conversion system ideal for investigating developmentally-relevant chromatin reorganization events.

1.1.4. **Heterochromatin Features in Pluripotent Cells**

1.1.4.1. **Visualizing Chromatin Ultrastructure**

Ever since the initial description of DNA as a double helix structure, elucidating the manner in which 2-meter long DNA is packaged within a nucleus became a topic of intense investigation. Electron microscopy and X-ray scattering demonstrated that chromatin consists of a 10 nm chromatin fibre made up of DNA and nucleosomes, giving rise to “beads-on-a-string” description of chromatin (Olins & Olins, 1974; Woodcock et al, 1976). The 10 nm fibre is capable of higher-order folding into 30 nm-thick solenoid and coiling of up to 300 nm chromatin fibres revealing intriguing mechanisms of chromatin fibre packaging (Rattner & Lin, 1985; Widom & Klug, 1985). However, due to technical limitations, the vast majority of these chromatin structure studies were performed using extracted chromatin or reconstituted nucleosome arrays. With the advent of Electron Spectroscopic Imaging (ESI), a technique capable of providing high-contrast imaging of nuclear fibre structure, in situ 30 nm fibres were observed in starfish sperm (Bazett-Jones, 1992). However, this chromatin structure was not found in highly condensed chromatin of mammalian mitotic chromosomes (Eltsov et al, 2008). Although the proposed 30 nm fibre structure helps explain the extensive packaging of chromatin, observation of an in vivo 30 nm fibre in mammalian chromatin remains elusive (Maeshima et al, 2010a; Maeshima et al, 2010b). Given that heterochromatin of mouse and human cells exists within compartmentalized H3K9me3 and H4K20me3-enriched chromocentre foci (Peters et al, 2001), the expectation is that 30nm fibres, if present at all,
would be manifest within this nuclear body. To this end, using correlative ESI in combination with light microscopy (LM/ESI) of immunolabeled sections for H3K9me3 is an invaluable approach to characterizing heterochromatin fibre behavior of a specific epigenetic composition (Dellaire et al, 2004).

1.1.4.2. **Heterochromatin in Pluripotent Cells is “Open”**

About 40% of mammalian genomes are made up of transcriptionally silent heterochromatin represented by centromeres, pericentromeric chromatin, telomeres, repetitive elements and endogenous retroviral elements. A few landmark findings over the years have created a generally accepted view that ES cells contain an “open” chromatin structure without detectable heterochromatin-specific condensation. This was initially demonstrated by utilizing ESI where general absence of highly compact heterochromatin domains was absent in nuclei of ES cells but readily identified in neural progenitor cells (Efroni et al, 2008). Subsequently, through ESI analysis of the developing mouse embryo, cells of the ICM were also found to generally lack regions of compact chromatin fibres, distinct from the primitive endoderm cell lineage at day 3.5 of development (Ahmed et al, 2010). Constitutive heterochromatin consists of major and minor satellite repeats that are almost always marked by high levels of H4K20me3, as well as H3K9me3 and its associated HP1 protein (Kubicek et al, 2007; Lachner et al, 2001; Lehnertz et al, 2003). Taking advantage of this feature, immunolabeling-based approaches reveal vast differences in heterochromatin organization where lower numbers and signal intensity as well as larger size of H3K9me3-enriched “chromocentre” foci were found in ES compared to neuronal lineage cells (Meshorer et al, 2006). In addition, fluorescent recovery after photobleaching (FRAP) reveal that linker histone H1 and HP1
proteins are rapidly exchanged in chromocentre regions of ES cells in contrast to differentiated cells which maintain tight association of these structural chromatin components (Phair et al, 2004). This general “euchromatinization” of heterochromatin in ES cells also has transcriptional consequences as genomic contents of H3K9me3 foci, including major and minor satellite repeats, have increased transcript levels in ES cells and are potently silenced in differentiated cells (Efroni et al, 2008). Collectively, these findings demonstrate striking differences of heterochromatin organization in pluripotent cells. Although functional consequences of abnormal heterochromatin organization is difficult to ascertain this plastic heterochromatin organization may reflect the need for pluripotent cells to appropriately respond to differentiation cues. Such a physical state would allow for quick reconfiguration of chromatin regions upon reception of natural differentiation cues during embryonic development. The extent to which iPS cell heterochromatin ultrastructure resembles that of “open” heterochromatin organization found in ES cells has not been investigated prior to the published work described in chapter 2 of this thesis.

1.1.5. Epigenetic Reorganization During Reprogramming

Robustness of somatic cell reprogramming using “Yamanaka” factors is evident when considering the wide range of somatic cells used in reprogramming experiments (Aoi et al, 2008; Hanna et al, 2008; Kim et al, 2008; Stadtfeld et al, 2008a). The requirement for the exogenous factors is variable depending on the starting cell type with reports of iPS cells reprogramming using only Oct-4 in neural stem cells (NSCs) (Kim et al, 2009). Collectively, these findings demonstrate that the epigenetic landscapes that characterize different somatic cell types can be reorganized using a similar cocktail of transcription factors as the initiator of reprogramming, with Oct4 playing a central role in the reprogramming cascade. Although
induced pluripotency is dependent on exogenously introduced transcription factors, the global epigenetic changes that take place during reprogramming are the driving force of a gradual transition to the pluripotent state. These remodeling steps involve: 1) Epigenetic reactivation of endogenous pluripotency genes, such as Oct4 and Nanog, 2) Establishment of bivalent chromatin domains at developmental loci and altered histone H3K4 and H3K27 trimethylation levels at ESC ‘signature’ gene promoters (Mikkelsen et al, 2008), 3) DNA hypomethylation of satellite repeats (Maherali et al, 2007) and 4) X-chromosome reactivation in female iPS cells, 5) Maintenance of DNA methylation marks of imprinted gene loci (Okita et al, 2007), 6) Retroviral transgene silencing upon pluripotency establishment and 7) in this thesis work we evaluate the possibility that reprogramming-associated epigenetic changes are accompanied by ultrastructural reorganization of heterochromatin fibres (Summarized in Figure 1.5).

**Figure 1.5. Hallmark epigenetic changes during somatic cell reprogramming.**

Hallmark epigenetic features that characterize the reprogramming process from fibroblast to intermediate to partial to full iPS cells (x-axis). In this thesis I aim to characterize whether chromatin fibre reorganization of heterochromatin
regions accompanies the reprogramming process to reflect ES cell-like “open”
fibre conformation.

1.2. **Modelling Rett Syndrome Using iPSC Technology**

1.2.1. **Disease Modelling Using iPS Cell Technology**

Somatic cell reprogramming using defined factors has revolutionized the way we study human disease and it also holds tremendous potential for the field of regenerative medicine. In order to study cell type-specific disorders, pluripotent iPS cells need to be properly instructed to differentiate down a lineage to a specific cell type of interest *in vitro*. Efforts to differentiate iPS and ES cells into somatic cell types are guided by recapitulating the *in vivo* differentiation cues *in vitro* (Wichterle et al, 2002; Yang et al, 2008) (reviewed in (Murry & Keller, 2008)). Signaling molecules, such as activin A and bone morphogenetic protein (BMP), used in differentiation studies are also expressed in the developing embryo (Bertacchi et al, 2013) and it is presumed that as such, these signaling molecules enhance the derivation of a cell type representative of what would occur during normal development. One of the primary goals of iPS cell-based models of disease is to generate cell types of interest using differentiation protocols and to seek out appropriate phenotyping assays that would reveal a functional defect. For instance, in the case of diabetes, generating pancreatic cells capable of producing and secreting insulin would be useful for developing assays in order to identify molecular pathways that are defective in cells derived from diabetic individuals (Zhang et al, 2009). On the other hand neurological disorders require generating the correct neuronal subtypes for molecular phenotyping assays.

iPS cell lines have been established from patients with multiple diseases such as amyotrophic lateral sclerosis (ALS) (Dimos et al, 2008), familial dysautonomia (Lee et al,
2009), Fanconi anemia (Raya et al, 2009), Rett syndrome (Hotta et al, 2009), spinal muscular atrophy (SMA) (Ebert et al, 2009), cystic fibrosis (Wong et al, 2012) and a spectrum of other diseases (Park et al, 2008). Several of these studies demonstrated directed iPS cell differentiation to affected cell types with Ebert et al. and Lee et al. reporting the first successful phenotype improvements using a chemical treatment. The general approach to iPS cell-based disease modeling is conserved regardless of the disease being investigated and includes: 1) generation and characterization of iPS cell lines, confirming the presence of disease-causing mutations in the newly derived pluripotent cells and absence of large-scale genetic mutations that may arise due to reprogramming (most commonly by karyotype analysis), 2) Directed differentiation to the affected cell type of interest and identification of molecular or physiological phenotypes compared to healthy controls and 3) Improvement of the phenotype by genetic rescue or the use of known or novel targeted compounds.

1.2.2. Rett Syndrome

1.2.2.1. Clinical Features of Rett Syndrome

Rett syndrome (RTT) is a female-specific neurodevelopmental disorder that was first characterized based on the progressive loss of mental development, encephalopathy, microcephaly and loss of voluntary hand movements manifest around 6-18 months in age (Hagberg et al, 1983). Upon initial onset of the disorder during early development, the disorder is maintained throughout life and has devastating effects marked by general mental retardation, epilepsy, physical disability including scoliosis, premature neuromuscular aging and peripheral atrophy (Hagberg, 2005). Although classification of RTT as an autism spectrum disorder has been recently altered, reduced social interactions and repetitive
behaviors are present in both and suggest commonality between the two distinct disorders. RTT is a second leading cause of mental retardation among newborn girls and occurs in 1 in 10000 live births (Chahrour & Zoghbi, 2007). Although clinical trial is under way to assess the efficacy of the IGF1, the efficacy of which has been reported in a small cohort of patients (Pini et al, 2012), RTT patients remain without approved drug treatments. The molecular mechanisms underpinning the disorder have been greatly advanced with the identification of genetic causes of RTT.

1.2.2.2. MECP2 Mutations Cause RTT

Since the first description that RTT affects exclusively females, it was postulated that the genetic cause of RTT resides on the X-chromosome and that mutations in hemizygous males are incompatible with life. RTT is a predominantly sporadic disorder but rare familial cases were utilized to map the disease-causing mutation to chromosome Xq28 and subsequent candidate screening approach identified Methyl CpG binding protein 2 (MECP2) as the genetic cause of RTT (Amir et al, 1999). Mutation of only one MECP2 allele is sufficient to cause RTT deeming it an X-linked dominant disorder. Currently MECP2 mutations account for more than 90% of RTT cases with missense, nonsense and frameshift mutations being the most common (Amir et al, 1999; Chahrour & Zoghbi, 2007; Christodoulou et al, 2003). Additional genetically identifiable RTT-causing mutations occur in genes with gene or protein regulatory functions in the brain including CDKL5 and FOXG1 (Ariani et al, 2008; Tao et al, 2004) but their involvement in modulating MECP2 function further implicate MECP2 mutation as the primary cause of RTT (Bertani et al, 2006; Dastidar et al, 2012). The heterozygous nature of MECP2 mutations and random X-inactivation patterns render RTT
patients mosaics with subpopulations of cells expressing either wild-type or mutant MECP2-containing X-chromosome. Given that X-inactivation status is often the victim of erosion in tissue culture where inactive X-chromosome can become re-activated (Mekhoubad et al, 2012), any iPSC-based models of RTT require rigorous assessment of the X-chromosome inactivation status in both starting iPS cell lines and their differentiated neuronal cultures.

1.2.3. MECP2 Structure and Function

Mecp2 was initially discovered as CpG methylation-dependent chromatin binding protein in mouse nuclear extracts (Lewis et al, 1992; Meehan et al, 1992). Through deletion analysis of Mecp2, methyl-binding domain (MBD) and transcriptional repressor domain (TRD) were defined as domains required for binding methylated cytosines and recruitment of epigenetic modifying enzymes, respectively (See Figure 1.6) (Nan et al, 1997; Nan et al, 1993). Co-immunoprecipitation experiments identified members of a deacetylase complex mSin3A, Hdac1 and Hdac2, as interacting partners required for the MECP2-mediated transcriptional repressor activity (Jones et al, 1998; Nan et al, 1998). These experiments provided some of the initial evidence for cross-talk between methylated DNA and acetylation state of the underlying nucleosomes during establishment of a silent transcriptional state at a given locus. Although Mecp2 is a methyl CpG-binding protein that co-localizes with transcriptionally inactive heterochromatin foci in nuclei of differentiated cells (Brero et al, 2005; Kumar et al, 2008), it can also function as a transcriptional activator (Chahrour et al, 2008; Tudor et al, 2002; Yasui et al, 2007). Brain tissue in Mecp2<sup>−/−</sup> mice exhibit global transcriptional changes in both directions compared to wild-type cells where activated genes are targets of MECP2 and the transcriptional coactivator CREB (Chahrour et al, 2008). Collectively, these data support a model that MECP2 binds genes in a complex with Sin3A
deacetylase complex or the co-activator CREB when repressing and activating gene transcription, respectively. Although MECP2 binding and transcriptional effects have been found at multiple target genes, widespread genomic distribution of Meep2 suggests that it possesses a more global function of chromatin organization and gene expression regulation (Ghosh et al, 2010; Skene et al, 2010). This alternative model of MECP2 distribution, one without specific genomic targets, proposes that MECP2 competes with linker histone H1 in binding methyl-CpG rich genomic regions throughout the genome, including intergenic regions (Skene et al, 2010).

MECP2 is subject to post-translational modifications that influence its binding dynamics. In response to calcium-mediated neuronal activation, Mecp2 is phosphorylated and decoupled from DNA enabling the transcriptional machinery to transcribe activity-dependent genes (Cohen et al, 2011). Activity dependent phosphorylation of three residues (S86, S274 and T308) has recently been reported with the T308 phosphorylation causing MECP2 decoupling from the nuclear receptor co-repressor (NCoR) resulting in the inability to properly repress gene expression in an activity dependent manner (Ebert et al, 2013). The role of NCoR in mediating MECP2 function was further corroborated by an independent study demonstrating that R306C mutations ablate interaction with NCoR and result in RTT phenotypes in the mouse model (Lyst et al, 2013). These findings demonstrate that posttranslational modifications of MECP2 are instrumental for its function in gene expression regulation. Overall, MECP2 is a gene that interprets complex environmental signals and serves as a mediator of neuronal activity and epigenetic structure of its genomic targets.
1.2.3.1. **Distinct roles of MECP2 isoforms**

Alternative splicing of exon 2 results in two MECP2 isoforms, MECP2e1 and MECP2e2, with distinct N-terminal amino acid sequences (Kriaucionis, 2004; Mnatzakanian et al, 2004) (**Figure 1.6**). MECP2e2 was the original isoform upon which most of the knowledge of MECP2 function was founded but the identification of the e1 isoform raised the possibility of unique isoform-specific functions. Demonstration by Mnatzakanian and colleagues of an 11 bp deletion in exon 1 that specifically disrupts the MECP2e1 isoform in a Rett syndrome patient, and the higher level of expression of this isoform in postnatal brain tissue (Dragich et al, 2007; Zachariah et al, 2012), imply that the e1 isoform is more relevant to RTT. Furthermore, no MECP2e2-specific mutations have been identified to date and most mutations are concentrated in the mutational “hotspot” of exons 3-4 that affect both isoforms. In fact, MeCP2e2-specific mutations in exon 2 do not recapitulate RTT phenotypes in the mouse model but instead elicit a placental defect (Itoh et al, 2012). This finding of an early developmental defect of Mecp2e2-specific mutations may explain the general absence of MECP2e2 mutations among RTT individuals. However, in mouse models transgenic MeCP2e1 or MeCP2e2 can rescue behavioural and physiological RTT phenotypes in mutant Mecp2 mice, with Mecp2e1 transgenes generally resulting in a higher degree of improvement (Koh et al, 2011). Collectively, these findings suggest a potential functional equivalence with the unique expression patterns underlying the distinct roles of MECP2 isoforms in neuronal function. The ability to differentiate human pluripotent cells lacking only the MECP2e1 isoform to neuronal cells described in chapter 3 of this thesis can help elucidate the isoform-specific functions of MECP2 in transcriptional regulation of putative target genes in human neurons and their role in disease.
Figure 1.6. Schematic of MECP2e1 and e2 isoforms.

A) MECP2 contains 4 exons with alternative splicing giving rise to MECP2e1 and MECP2e2 isoforms. Translation start site (ATG) is in exon 1 and 2 for MECP2e1 and e2, respectively. B) MECP2e1 and e2 proteins differ only in the N-terminus. The amino acid locations of common MBD (methyl CpG binding domain) and TRD (transcriptional repressor domain) regions are indicated.

1.2.4. iPS Cell Based Studies of RTT

Our lab was the first to report reprogramming of human RTT patient fibroblasts to iPS cells (Hotta et al, 2009) with subsequent reports investigating some of the hallmark RTT-related neuronal phenotypes in human iPS cell-derived neurons (Ananiev et al, 2011; Cheung et al, 2011; Kim et al, 2011). The identified neuronal phenotypes recapitulated those observed in mouse models and RTT patients and include morphological defects of reduced neuronal/soma size and dendritic spine density (Ananiev et al, 2011; Cheung et al, 2011; Marchetto et al, 2010). Reduction of calcium transients as well as spontaneous excitatory and
inhibitory postsynaptic currents further reveal electrophysiological defects in iPS cell-derived RTT neurons (Farra et al., 2012; Marchetto et al., 2010). Collectively, these findings provide the proof-of-principle of modeling RTT using iPS cells with a range of MECP2 mutations. With a recent publication of MeCP2e1-specific mouse model of RTT, the sufficiency of this of mutations of this isoform in development of RTT have been further corroborated (Yasui et al., 2014). However, to date there has been no report of isoform-specific iPS cell models of RTT and the transcriptional consequences of MECP2 mutations in the differentiated neurons have not been addressed in previous studies.

1.3. Summary and Hypothesis of Thesis

Chromatin remodeling is the driving force of somatic cell reprogramming. Although loci-specific epigenetic changes gradually establish the pluripotent transcriptional profile during reprogramming, reprogramming-mediated reorganization of silent chromatin domains has not been assessed. Through identifying reprogrammed mouse iPS cell lines at partial and full stages of reprogramming I established a system for investigating heterochromatin ultrastructure reorganization and mechanisms of retroviral silencing during the reprogramming process. Furthermore, we explored the possibility that members of the core pluripotency transcription factors Nanog and Sall1 are directly involved in the observed chromatin remodeling through the use of KO ES cell lines and a Nanog overexpression system in partial iPS cell lines.

MECP2 is a heterochromatin-associated protein that plays a causative role in the development of Rett syndrome. The vast majority of MECP2 studies have been performed in murine models or post mortem patient tissues and it remains to be investigated whether human MECP2 mirrors the activity observed in the mouse model system. Human iPS cell-based
models are perhaps the ideal system to examine human MECP2 function in live cells, as mutant cells can be generated from patients and directed to differentiate to neuronal cell lineages. Using this approach, I set out to investigate the functional outcome of MECPe1-specific mutations in RTT derived neurons. By evaluating iPS cell lines expressing only mutant MECP2e1 through X-chromosome inactivation, neurons lacking MECP2e1 are probed to assess their differentiation efficiency and level of expression of MECP2 target genes. Finally, the sufficiency of MECPe1 mutations to cause morphological defects associated with RTT is assessed by neuronal soma size analysis.

**Hypotheses**

1. Reprogramming is characterized by chromatin fibre structural reorganization concurrent with retroviral silencing as the cells reach the fully reprogrammed iPS cell state.
2. MECP2e1-specific mutations do not impact neuronal specification but are sufficient to induce transcriptional and morphological perturbations associated with global MECP2 deficiency.
2. Chapter 2: Constitutive Heterochromatin Reorganization During Somatic Cell Reprogramming

The data and text in this chapter is based on the publication:

Constitutive heterochromatin reorganization during somatic cell reprogramming.


Ugljesa Djuric performed all the ES and iPS cell tissue culture, gene expression, ChIP and bisulfite sequencing assays, the 2i chemical conversion of partial iPS cells and prepared the RNA for microarray analysis. Eden Fussner performed correlative LM/ESI and DAPI linescan analyses. Figures 2.1-2.11 are largely identical to those in the publication while figures 2.12-2.14 are unpublished data.

2.1. Abstract

Induced pluripotent stem (iPS) cell reprogramming is a gradual epigenetic process that reactivates the pluripotent transcriptional network by erasing and establishing heterochromatin marks. Here, we characterize the physical structure of heterochromatin domains in full and partial mouse iPS cells by correlative Electron Spectroscopic Imaging (ESI). In somatic and partial iPS cells, constitutive heterochromatin marked by H3K9me3 is highly compartmentalized into chromocentre structures of densely packed chromatin fibres. In contrast, chromocentre boundaries are poorly defined in pluripotent ES and full iPS cells, and are characterized by unusually dispersed 10 nm heterochromatin fibres in high Nanog-
expressing cells, including pluripotent cells of the mouse blastocyst prior to differentiation. This heterochromatin reorganization accompanies retroviral silencing during conversion of partial iPS cells by Mek/Gsk3 2i inhibitor treatment. Thus, constitutive heterochromatin reorganizes into dispersed 10 nm chromatin fibres in pluripotent stem cells at a very late stage of reprogramming.

2.2. Introduction

The cascade of events in somatic cell reprogramming to a pluripotent state involves large-scale epigenetic remodeling to establish heterochromatin marks on tissue-specific genes, and to erase these marks on key members of the pluripotency network (Maherali et al, 2007; Takahashi & Yamanaka, 2006; Wernig et al, 2007). Reprogramming proceeds through at least three stages: intermediate, partial and full iPS cell states (Brambrink et al, 2008; Jaenisch & Young, 2008; Yamanaka, 2009). Partial mouse iPS cells attain some aspects of pluripotency including Embryonic Stem (ES) cell-like colony morphology, teratoma-forming ability, and partial activation of pluripotency genes accompanied by downregulation of differentiation-specific genes (Mikkelsen et al, 2008; Okita et al, 2007; Sridharan et al, 2009; Stadtfeld et al, 2010; Stadtfeld et al, 2008b). Partial iPS cells can be converted into full iPS cells using cell signaling or epigenetic inhibitors. As they complete reprogramming, full mouse iPS cells acquire epigenetic marks of pluripotency including X chromosome reactivation and bivalent histone domains indicative of open chromatin structure (Maherali et al, 2007).

ES cells epigenetically inactivate exogenous retroviruses and, as full iPS, robustly activate endogenous pluripotency genes during the reprogramming process while silencing the retroviral reprogramming factors (Okita et al, 2007; Silva et al, 2008; Stadtfeld et al, 2008b). In contrast partial iPS cells maintain retroviral gene expression indicating that an ES cell-
specific transcriptional network is required for successful silencing of retroviruses in pluripotent cells. Clearly, expression of Oct4, Sox2 and Klf4 in primary mouse embryonic fibroblasts (MEFs) is not sufficient for retrovirus silencing, nor is their persistent expression in partial iPS cells, suggesting that additional endogenous pluripotency factors are required for retroviral silencing in full iPS cells. For example, Nanog, which controls the pluripotency ground state (Silva et al, 2009), is reactivated in full mouse iPS cells (Okita et al, 2007), and is capable of silencing gene expression (Liang et al, 2008). Epigenetic mechanisms normally employed in establishing heterochromatin have prominent functional roles in silencing retrovirus vectors in ES cells. Viral DNA sequences are recognized by a ZFP809/TRIM28 complex that establishes H3K9me2 marks bound by HP1 (Wolf & Goff, 2007; Wolf & Goff, 2009). In addition, H3 is deacetylated (Lorincz et al, 2001; Pannell et al, 2000) and H3K9me3 is deposited by Eset/Setdb1 (Matsui et al, 2010). Additional epigenetic involvement includes recruitment of Swi/Snf components (Golding et al, 2010) and participation of de novo DNA methylases (Cherry et al, 2000; Dodge et al, 2002). These enzymes also reorganize chromatin structure on a global scale. Thus retrovirus silencing is a late event in reprogramming and a useful marker for timing heterochromatin reorganization in iPS cells.

Despite major changes in heterochromatin modifications in the late stages of reprogramming, the physical structure of heterochromatin domain organization during nuclear reprogramming remains unexplored. Pericentric constitutive heterochromatin, comprised of major and minor satellite repeat sequences, cluster into chromocentre domains within the nucleus (Guenatri, 2004; Joseph et al, 1989; Wong & Rattner, 1988). Chromocentres are readily identified in mouse nuclei by their DAPI-rich staining, and are specifically marked by H3K9me3 and H4K20me3 (Peters et al, 2001). Chromocentre organization has typically been
investigated by measuring the amount of clustering or changes in the number of observable chromocentre foci within differentiating nuclei (Brero et al, 2005; Meyer-Ficca et al, 1998; Tessadori et al, 2007). Notably, the number of these H3K9me3 foci increases when Chd1 is knocked down in ES cells and pluripotency is simultaneously lost (Gaspar-Maia et al, 2009). Thus, like retrovirus silencing, heterochromatin organization in ES cells also correlates with the pluripotent state, but the timing of its reorganization during iPS cell reprogramming and the specific structures of the heterochromatin fibres remain to be identified.

As an alternative to using molecular biology approaches or visible light imaging, chromatin organization has been studied using conventional transmission electron microscopy (CTEM). Somatic cell nuclei imaging by CTEM reveals condensed chromatin domains along the nuclear envelope and at the nucleolus periphery (Belmont et al, 1989; Kireev et al, 2008). These silenced compartments including chromocentres are widely accepted to be comprised of 30 nm chromatin fibre assemblies (Rego et al, 2008), though evidence for 30 nm fibres in vivo is largely restricted to non-mammalian cell types (Maeshima et al; Tremethick, 2007; van Holde & Zlatanova, 1995). When visualized by electron spectroscopic imaging (ESI) (Ahmed et al, 2009), the only technique that provides high contrast of unstained chromatin, ES cells display a meshwork of 10 nm chromatin fibres throughout the nucleus, and a paucity of the blocks of condensed chromatin observed in somatic cells (Efroni et al, 2008). Upon differentiation of ES cells to neural progenitor cells, however, some of the dispersed chromatin becomes organized into compact heterochromatin domains, particularly along the nuclear envelope (Hiratani et al, 2009). In addition to cell culture models, pluripotent pre-implantation embryos have also been shown to have globally decondensed chromatin (Ahmed et al, 2010) most strikingly after the 8-cell stage. These observations lead us to ask whether
reprogramming is accompanied by the loss of compact heterochromatin domains at chromocentres in somatic cells to the more dispersed mesh of 10 nm fibres in iPS cells that is typical of pluripotent ES cells.

The significant difference in global nuclear architecture between somatic and ES cells suggests that large-scale reorganization events may be involved in iPS cell reprogramming. Our primary goal was to identify the timing and structural features of heterochromatin reorganization during reprogramming. We compare H3K9me3 marked constitutive heterochromatin organization in full and partial iPS cells to that of the parental MEFs and the J1 ES cell line. We demonstrate that chromocentres in full iPS cells with high Nanog levels are characterized by dispersed 10 nm chromatin fibres, like those of ES cells and high Nanog expressing cells of the inner cell mass (ICM) of the mouse blastocyst prior to differentiation. Failure to disrupt heterochromatin domains by retaining tightly packed 10 nm fibres is a characteristic of partial iPS cells and MEFs. Conversion of partial iPS cells with the mitogen activated protein/Erk kinase and glycogen synthase kinase-3 (Mek/Gsk3) inhibitor (2i) cocktail (Silva et al, 2008) shows that heterochromatin reorganization is a novel signature of a very late stage in reprogramming that accompanies retrovirus silencing. By analyzing the heterochromatin structure of Nanog and Sall1 knockout ES cells, we further demonstrate that the absence of these pluripotency-related genes results in compaction of heterochromatin domains similar to the structure observed in partially reprogrammed iPS cells. Both of these transcription factors are found to bind major satellite repeat regions, which reside within the chromocentre nuclear structures. Overexpression of Nanog in established partial iPS cell line is sufficient to induce heterochromatin fibre decompaction but the binding of both factors to major satellite repeats occurs in a co-operative fashion.
2.3. Results

2.3.1. Distinction of Partial from Full iPS Cell Lines

To study the relationship between global chromatin organization and cellular reprogramming events we compared MEFs, iPS and J1 ES cells. We previously reported the derivation of three mouse iPS cell lines, EOS3F-24, EOS3F-28 and EOS3F-29, using pMX-based retroviral delivery of three “Yamanaka” factors, Klf4, Sox2 and Oct4 (Hotta et al, 2009). All three iPS cell lines were shown to maintain expression of the EOS-EGFP pluripotency reporter vector and have a pluripotent capacity to differentiate in vitro and form teratomas in vivo (Hotta et al, 2009). Differentiation of pluripotent cells usually results in the loss of the ability to give rise to teratomas. Interestingly, the EOS3F-24 cell line maintained its teratoma-forming capacity even after embryoid body-mediated differentiation, indicative of a partial iPS cell state (Jaenisch & Young, 2008; Okita et al, 2007). In contrast, differentiation of both EOS3F-28 and -29 is accompanied by EOS-EGFP inactivation and subsequent loss of their teratoma-forming capacity (Hotta et al, 2009). These results were the first indication that EOS3F-24 could be a partial iPS cell line.

A reliable molecular marker for full reprogramming is retrovirus silencing with partially reprogrammed cell lines failing to silence the retroviral transgenes. To determine this possibility in the iPS cell lines, I performed qRT-PCR demonstrating that EOS3F-24 iPS cells maintain high level of expression of all three exogenous reprogramming factor transgenes compared to EOS3F-28 and -29 cells (Figure 2.1A). Through bisulfite sequencing, I demonstrated that the long terminal repeat (LTR) promoters of the reprogramming vectors are devoid of DNA methylation while promoters of Oct4 and Nanog are hypermethylated in EOS3F-24 iPS cells. On the other hand, both EOS3F-29 and EOS3F-28 iPS cells have high
levels of DNA methylation at the LTR promoters with demethylated Oct4 and Nanog promoters (Figure 2.1B). In order to investigate the acetylation status of histone H3 (H3Ac) in the iPS cell lines, I carried out chromatin immunoprecipitation (ChIP) analysis with an anti-H3Ac antibody. These experiments show that the Nanog and Oct4 promoters are hypoacetylated while the viral transgenes are hyperacetylated in the EOS3F-24 iPS cells compared to J1 ES and EOS3F-29 iPS cells (Figure 2.1C). These data demonstrate that, in contrast to pluripotent-like epigenetic state of EOS3F-28 and -29, the epigenetic state of the viral transgenes and endogenous pluripotency genes were not remodeled in EOS3F-24.

Figure 2.1. Characterization of induced pluripotent stem cells.

A) Retroviral gene expression of pMX-Klf4, pMX-Oct4 and pMX-Sox2 was analyzed by qRT-PCR on RNA from infected MEFs (MEF3F), uninfected
controls (J1 ES cells), EOS3F-29, -28 and -24 iPS cell lines. Expression is presented relative to MEF3F control. Error bars are S.E.M. B) Bisulfite sequencing of CpG sites of endogenous Oct4 and Nanog promoters and pMX LTRs. Open and closed circles indicate unmethylated and methylated CpG sites, respectively. C) ChIP analysis of histone H3 acetylation of the Oct4 and Nanog promoters and the pMX LTR promoters in EOS3F-29 and EOS3F-24 cells. J1 cells are included as control for ChIP analyses of endogenous promoters. Error bars represent ± SEM. Figure reused with permission, EMBO press (Fussner et al, 2011b).

I then performed flow cytometry to determine the levels of Ssea1, a pluripotency-specific cell surface marker. ES cell-specific cell surface marker is observed only in EOS3F-28 and -29 iPS cells (Figure 2.2A). Nanog is a strong marker of full reprogramming in the mouse system and is absent in EOS3F-24 but is abundant in EOS3F-28 and -29 iPS cell colonies (Figure 2.2B and C). Furthermore, the vast majority of pluripotency-related genes are expressed at significantly reduced levels compared to J1 ES cells only in EOS3F-24 iPS cells (Figure 2.2D). Collectively, these data show that EOS3F-28 and EOS3F-29 have molecular marks of full whereas EOS3F-24 has molecular features of partial iPS cells.

To directly examine global gene expression patterns in the iPS cell lines in comparison to reference pluripotent J1 ES cells and MEFs, microarray analysis was performed. These results show that the EOS3F-29 expression profile is most similar to J1 ES cells (354 genes differentially expressed) in contrast to EOS3F-24 cells (4987 genes differentially expressed) (Figure 2.3A). All the iPS cell lines have greater than 5000 differentially expressed genes
relative to MEFs demonstrating that some level of reprogramming has occurred. This is most clearly shown by Principle Component Analysis (PCA) and hierarchical clustering (Figure 2.3B) where EOS3F-24 cells cluster independently from MEFs and J1 ES cells, in contrast to the clustering of EOS3F-28 and -29 with J1 ES cells.

**Figure 2.2. Characterization of iPS cell lines.**

A) Flow cytometry analysis of Ssea1 expression in EOS3F-24, -28 and -29 iPS cell lines. B) Nanog immunocytochemistry of iPS cell lines with DAPI counterstain shown in insets. Scale bar, 50 µm. C) Nanog western blot analysis and in indicated iPS cells. β-actin and J1 ES cells are positive
controls for loading and Nanog expression, respectively. D) Pluripotency gene expression analysis by qRT-PCR relative to reference J1 ES cell control. Representative qRT-PCR is shown and error bars are S.E.M. from triplicate PCR reactions. Figure reused with permission, EMBO press (Fussner et al, 2011b).

Figure 2.3. Microarray analysis of full and partial iPS cells.

A) Microarray data summary of differentially expressed genes in iPS cell lines shows that EOS3F-29, -28 and 2i-treated but not EOS3F-24 iPS cells have expression profiles that are very close to J1 cells. Table indicates the number of genes over- or under-expressed in different iPS cells compared to J1 ES cells and MEF controls. B) Principal component analysis was performed to compare the microarray expression profiles of genes in MEFs, iPS (EOS3F-24, EOS3F-24 +2i, EOS3F-28, and EOS3F-29), and J1 ES cells. The left panel represents a scatter plot of the microarray expression profiles on the planes spanned by the first and second principal components. The right panel represents a dendrogram of the hierarchical clustering of
expression profiles. Figure reused with permission, EMBO press (Fussner et al, 2011b).

2.3.2. Chromocentre Compartment Differences Between Partial and Full iPS Cells

Constitutive heterochromatin in mouse cells is largely present in silenced genomic domains represented by pericentric repeat elements comprised of major and minor satellite sequences. These genomic elements cluster into chromocentres (Guenatri, 2004). To determine whether the pluripotent versus differentiated state was reflected in global chromatin organization we characterized the constitutive heterochromatin fibre structures of parental MEFs, partial and full iPS cells as well as J1 ES cells. These cells all have physically confined heterochromatin regions enriched in H3K9me3 as seen by fluorescence microscopy (Figure 2.4A). These domains were identified as chromocentres on the basis of H4K20me3 enrichment, DAPI density and proximity to centromeres, visualized with CREST antisera (Fussner et al, 2011b). Analysis of optical z-stack images confirmed that H3K9me3 enrichment is always associated with DAPI enrichment in these cells. These data are consistent with previously published results indicating that both pluripotent and differentiated mouse cells have compartmentalized heterochromatin, which form chromocentres (reviewed in Ahmed et al, 2009). However, further analysis of the DAPI signal revealed significant differences between the different cell types. MEF and partial iPS cell chromocentres appear as bright foci while those of J1 ES and full iPS cells contain lower DAPI signal relative to nucleoplasmic background (Figure 2.4B). Thus, individual AT-rich stretches of DNA, bound by the DAPI compound, seem to exist in closer proximity to each other in partial and differentiated MEF cells compared to full iPS and J1 ES cells.
Figure 2.4. Chromocentre organization differs between partial and full iPS cells.

A) H3K9me3 immunofluorescence and DAPI counterstain in iPS, J1 ES and MEF cells. White arrows through DAPI images indicate intensity line scan plot and direction. DAPI (red) and H3K9me3 (blue) line scans are shown below image panels. Scale bar, 5 µm. B) Box plot analysis of 100 chromocentres from at least 30 different cells and 3 independent trials of the background to chromocentre DAPI ratio. *, P < 0.001. Figure reused with permission, EMBO press (Fussner et al, 2011b).

2.3.3. Chromocentres in Full iPS Cells are Composed of 10 nm Chromatin Fibres

To interpret these differences in heterochromatin organization observed by DAPI counter stain we employed Electron Spectroscopic Imaging (ESI), the only technique currently available to provide high contrast images of unstained chromatin fibres in situ (Dellaire et al, 2004). Indirect labeling of cells with antibodies directed against H3K9me3 and H4K20me3 was used to demark the constitutive heterochromatin domains for imaging the underlying chromatin fibre organization by correlative light microscopy (LM)/ESI (Figure...
Line scan analysis demonstrates that the chromocentres of ES and full iPS cells are difficult to delineate from the surrounding euchromatin (Figure 2.5B). In contrast, chromocentres in partial iPS cells and MEFs were densely packed, displaying a significantly higher fibre density than the surrounding euchromatin. At higher resolution we observed an abundance of dispersed 10 nm chromatin fibres within chromocentre domains of ES and full iPS cells (Figure 2.5C). Surprisingly, we also observed a prevalence of 10 nm fibres within the compact heterochromatin domains of MEFs, where only 30 nm or higher-level fibre organization might be expected. These observations can be quantitatively confirmed since the phosphorus signal is a direct measure of chromatin density (Bazett-Jones et al, 1999). Since the H3K9me3-marked chromatin is completely disrupted in full-reprogrammed iPS cells, we conclude that a transition between 10 nm and 30 nm fibres is not required. Bisulfite sequencing shows that alpha satellite DNA repeats located in chromocentres are still highly methylated in the full iPS cells (Figure 2.5D), indicating that DNA methylation is also compatible with dispersed chromocentre formation.

To confirm that these dispersed fibres were in fact 10 nm we combined ESI with tomography to generate a three-dimensional representation of the phosphorus content in ES cells (Figure 2.5E). This analysis clearly shows the 10-nm chromatin fibre comprising both nucleosomes and intervening linker sequences. Both the H3K9me3-defined heterochromatin regions and the surrounding euchromatin were comprised exclusively of 10 nm chromatin fibres. Although we detect no 30 nm fibres in these dispersed chromatin regions, 30 nm fibres can be detected by ESI. For example, 30 nm chromatin fibres have been imaged by ESI in starfish sperm (Bazett-Jones, 1992). Fourier transformed analysis of five tomograms containing both H3K9me3-enriched regions and surrounding chromatin demonstrates an
average particle size of 10.8nm with a complete absence of higher-order chromatin fibres of greater diameters (Figure 2.5E). The 3D analysis clearly demonstrates that 10 nm chromatin fibres exclusively populate both the heterochromatin regions and surrounding chromatin domains in pluripotent J1 ES cells.

Taken together these data indicate that dispersion of blocks of constitutive heterochromatin correlates with complete reprogramming, whereas partial iPS cells maintain compact constitutive heterochromatin domains resembling those in MEFs. Furthermore, this is the first demonstration of 10 nm chromatin fibres within constitutive heterochromatin domains.
Figure 2.5. Correlative LM/ESI with indirectly labeled H3K9me3 identifies chromocentres composed of 10 nm chromatin fibres.

A) Quantitative phosphorus and nitrogen ratio images were segmented to show chromatin in yellow and protein-based structures in blue. White circles delineate H3K9me3 enrichment and white arrows indicate length and direction of phosphorus line scan analysis (shown below image panels). Scale bar, 0.5 \( \mu \text{m} \).

B) Box plot analysis of the phosphorus density shows the distribution of chromatin compaction within these cells, 50 chromocentres were measured from at least 30 different nuclei from 3 independent trials. *, P
C) 10 nm fibres within constitutive heterochromatin in a pluripotent cell and a differentiated cell, chromatin fibres are shown in yellow with zoomed panels to the right. White arrowheads in zoomed images indicate 10 nm fibres. Scale bar, 30 nm. D) Bisulfite sequencing of CpG sites in major satellite repeats, open and closed circles represent unmethylated and methylated CpG sites, respectively. E) Perspective image of Chimera-generated model of tomographic reconstruction of H3K9me3-enriched region and surrounding chromatin of a J1 ES cell pseudo-coloured by electron density. Connected nucleosomes and intervening linker sequence of these 10 nm fibres are clearly visualized in the zoomed panel to the right. Representative fourier transform analysis of the average chromatin fibre size at 10.84 nm. Figure reused with permission, EMBO press (Fussner et al, 2011b).

2.3.4. Chromocentres Reorganize During 2i Conversion of Partial iPS Cells

To address whether constitutive heterochromatin reorganization is indeed a feature of the acquisition of the fully reprogrammed state, I induced complete reprogramming of EOS3F-24 iPS cell line using a MEK/GSK double inhibitor (2i) cocktail. Preliminary readouts of successful conversion after one-week chemical treatments were evaluated by Ssea1 and EOS-EGFP activation in EOS3F-24 iPS cells. Three concentrations of the 2i cocktail ([1X], [2X] and [4X], see Materials and Methods) resulted in dosage-dependent emergence of a high EOS-EGFP expressing EOS3F-24 subpopulation of cells that was also Ssea1-positive (Figure
2.6A and B). To investigate whether molecular changes at the DNA/RNA level were taking place in the newly derived Ssea1+ EOS3F-24 cells, I sorted the high EGFP expressing cells and cultured them for an additional week in the presence of 2i.

![Graph and Figure](image)

Figure 2.6. Sorting and characterization of 2i-treated EOS3F-24 iPS cells.

A) Levels of EOS-EGFP expression in response to three concentrations of 2i inhibitor cocktail. The bar graph indicates the percent of EOS3F-24 iPS cells expressing high levels of EOS-EGFP, as gated in the left panel (gate shown in green). B) Scatterplot of flow cytometry analysis of Ssea1 expression in 2i-treated EOS3F-24 before and 7 days after sorting for high EGFP expressing population. Ssea1-positive, high EGFP expressing cells were sorted and cultured in presence of 2i inhibitor cocktail for additional 7 days, confirming the stability of high EOS-EGFP
expressing cells. Figure reused with permission, EMBO press (Fussner et al, 2011b).

Following 2i treatment, the converted EOS3F-24 cells robustly activated pluripotency-related loci (Figure 2.7A), and the newly established iPS colonies were positive for Nanog (Figure 2.7B). Bisulfite sequencing and ChIP demonstrate that Oct4 and Nanog promoters are hypomethylated (Figure 2.7C) and hyperacetylated (Figure 2.7D) following the 2i treatment to reflect the newly acquired pluripotent state. The fully reprogrammed expression patterns following 2i-mediated conversion were reproduced in independently converted [4X] 2i-treated EOS3F-24 cells, and microarray analysis on these samples demonstrates that they now cluster with ES cells and are as similar to them (368 differentially expressed genes) as the EOS3F-29 iPS cells (Figures 2.3A and B). As a result of the newly established full iPS cell state, EOS3F-24 cells silenced the reprogramming retroviral transgenes (Figure 2.8A). Appropriately, this silent retroviral state is accompanied by newly methylated LTR promoters (Figure 2.8B) which become devoid of histone acetylation marks and acquire high levels of the silent H3K9me3 (Figure 2.8C).
Figure 2.7. 2i Treatment converts partial EOS3F-24 cells to a full iPS cell state.

A) qRT-PCR of endogenous pluripotency-associated genes upon sorting and culturing of 2i-converted, Ssea1-positive, EOS3F-24 iPS cells. Gene expression is calculated relative to J1 ES cells and error bars are S.E.M. of representative triplicate reactions. B) Nanog immunostaining in EOS3F-24 iPS cell colonies after 2i conversion of EOS3F-24 iPS cells. Nanog is shown in red and DAPI inset in blue. Scale bar, 40 µm. C) Bisulfite sequencing of CpG sites of endogenous Nanog and Oct4 promoters. Open and closed circles are unmethylated and methylated CpGs, respectively. DMSO-treated EOS3F-24 cells are included as a negative control. D) Chromatin immunoprecipitation of acetylated H3 (H3Ac) at the endogenous Oct4 and Nanog promoters. Error bars represent ± SEM. Figure reused with permission, EMBO press (Fussner et al, 2011b).
To determine whether conversion of partial iPS cells was accompanied by global heterochromatin changes, we performed LM/ESI. H3K9me3-enriched regions were significantly more disrupted in the 2i treated cells compared to DMSO-treated controls (Figure 2.9A). These trends were consistent with our LM observations, where the chromocentre DNA density relative to nucleoplasmic background, as measured by DAPI intensity, was significantly different between the control and 2i treated cells (Figure 2.9B). Thus, DAPI line scan analysis supports the LM/ESI data and is a simple method to identify dispersed heterochromatin in converted and fully reprogrammed iPS cells. 2i treatment caused dispersion of densely packed heterochromatin fibres upon the conversion of partial iPS cells.
This heterochromatin reorganization is coincident with late events in reprogramming including epigenetic retroviral transgene silencing and activation of pluripotency genes.

Figure 2.9. 2i-mediated conversion of partial iPS cells results in H3K9me3-marked chromocentre decompaction.

A) Correlative LM/ESI analysis of constitutive heterochromatin in DMSO control and 2i-treated EOS3F-24 iPS cells. H3K9me3 enriched-regions are outlined in white. Scale bar 0.5 µm. B) DAPI line scan analysis of 2i- versus DMSO-treated control cells indicates a significant reduction in chromocentre partitioning. *, P < 0.001. Figure reused with permission, EMBO press (Fussner et al, 2011b).

2.3.5. Heterochromatin Reorganizes in Nanog-high Pluripotent Cells

Disruption of constitutive heterochromatin in the 2i-treated EOS3F-24 cells was accompanied by reactivation of endogenous Nanog expression. Nanog is a core pluripotency transcription factor whose protein levels vary from cell to cell within the inner cell mass of a
developing embryo (Chambers et al, 2007). Nanog protein fluctuations are also maintained within cultured iPS and ES cell colonies with low levels of Nanog expression associated with the cell’s predisposition to undergo differentiation (Kalmar et al, 2009). To determine whether Nanog levels correlate with heterochromatin reorganization I identified ES cell colonies exhibiting Nanog variegation (Figure 2.10A) and assessed chromatin compaction. We measured DAPI intensity of chromocentres relative to nucleoplasmic intensity in high and low Nanog-expressing J1 ES cell nuclei. I observed a correlation of Nanog expression with heterochromatin compaction, where high Nanog-expressing cells exhibited the most disrupted and least partitioned chromocentres (Figure 2.10B). The lowest Nanog-expressing ES cells had significantly more compacted chromocentres, suggesting that heterochromatin reorganization initiates when Nanog is downregulated as, or even before, ES cells begin to differentiate. It should be noted, however, that Nanog-low ES cells still have more disrupted chromocentres than differentiated cells and withdrawal of LIF from J1 ES cells for 72 hours results in reappearance of fully compact chromocentres detected by LM/ESI (Figure 2.10C). These results indicate that dispersed heterochromatin in pluripotent cells is reorganized into the compact structures seen in somatic cells as Nanog expression is lost at the onset of differentiation.
Figure 2.10. **Nanog protein levels impact heterochromatin compaction in J1 ES cells.**

**A)** Low magnification immunofluorescence of variegated Nanog expression in a J1 ES cell colony. Scale bar, 20 µm. **B)** High-magnification and resolution representative image of a J1 ES cell field with high and low Nanog expressing cells (top left panel) and DAPI counterstain of a 0.3 µm z-stack series of the same nuclei. DAPI line scan analysis quantification is adjacent, 30 nuclei of each expression level were analyzed, *, $P < 0.001$. White arrows indicate direction of line scan (shown above) through optimal DAPI z-stack. Scale bar, 5 µm. **C)** Withdrawal of LIF from J1 ES cell culture results in
loss ESC-like colony morphology (phase image) indicating exit from pluripotency and formation of compact chromocentres (ESI micrograph) within 72 hours. Scale bar, 0.5 µm. Figure reused with permission, EMBO press (Fussner et al, 2011b).

To assess the relationship between Nanog expression and heterochromatin organization in vivo we took advantage of Nanog variegation that briefly occurs in the ICM of E3.75 blastocysts (Yamanaka et al, 2010). Cells were binned according to Nanog expression levels, excluding lineage committed Gata6- and Cdx2- positive cells, which are no longer pluripotent (Figure 2.11A). Using a DAPI line scan analysis we found that Nanog levels correlated directly with heterochromatin compartmentalization and significant differences were observed between high and low Nanog-expressing nuclei. Chromatin in both high and low Nanog-positive nuclei was significantly less compartmentalized than the lineage committed Cdx2- or Gata6-positive cells (Figure 2.11B). Together, these data indicate that in the absence of Nanog, lineage-committed cells during early development possess heterochromatin that is organized into compact structures. In contrast, high Nanog protein levels in nuclei of pluripotent cells of the blastocyst directly correlate with their dispersed heterochromatin organization. These findings are consistent with the global analysis of chromatin structures associated with embryonic development and with the striking chromatin compaction of the ICM nuclei in the Oct4-null embryos (Ahmed et al, 2010). The implications of these data combined are two fold. 1) The pluripotent transcription factor network regulates chromatin structure and dictates unique chromatin architecture of the pluripotent state and 2) chromatin reorganization precedes differentiation both in vitro and in vivo.
Figure 2.11. Heterochromatin reorganizes in Nanog-high cells *in vivo* in mouse blastocysts.

**A)** Immunofluorescence of variegated Nanog (red), GATA-6 (green) and Cdx2 (blue) expression in E3.75 day early mouse blastocyst. Scale bar, 20 \( \mu m \).

**B)** DAPI images of 3 optical sections above and below the optimal image plane, representative chromocentres in these cells are adjacent to IF image, \( z = 2.1 \, \mu m \). White arrows indicate direction of DAPI line scans (lower panels). 40 nuclei per category were analyzed. \(*\), \( P < 0.005 \). Figure reused with permission, EMBO press (Fussner et al, 2011b).

### 2.3.6. Overexpression of Nanog in Partial iPS Cells

Given the inverse relationship of Nanog levels in pluripotent cells and chromocentre chromatin fibre compaction, I wanted to determine whether manipulating Nanog levels in partial iPS cells can directly result in heterochromatin decompaction. I used the piggyBac delivery system to induce overexpression of *Nanog* in partially reprogrammed EOS3F-24 iPS
cells upon doxycycline addition (Figure 2.12A). The Nanog-GFP transgene expression results in a shift in GFP+ cellular component of the EOS3F-24 population detected by flow cytometry that correctly corresponds to the Nanog-positive cells in the partial iPS cell colonies (Figure 2.12B). I then performed ChIP analysis demonstrating that Nanog binds to correct target genomic loci revealing the functionality of the piggyBac system in partial iPS cells (Figure 2.12C). By performing qRT-PCR, I demonstrate that the addition of Nanog by itself only slightly affects gene expression of differentiation- and pluripotency-associated genes Gata6 and Stella, respectively (Figure 2.13A) while maintaining high levels of retroviral transgenes (Figure 2.13B). Thus, Nanog induction establishes high levels of Nanog expression in EOS3F-24 iPS cells but does not alter the partial iPS cell state making this system appropriate for the study of Nanog effects on heterochromatin fibre architecture.
Figure 2.12. Nanog overexpression in partial EOS3F-24 iPS cells.

A) Experimental strategy for piggyBac transfection and induction of Nanog.

B) Flow cytometry analysis of Nanog-GFP transgene following transfection and doxycycline induction of Nanog transgene. The population of EGFP positive cells in absence of DOX induction reflects expression of EOS reporter vector in EOS3F-24 iPS cells. Immunostaining and western blot analysis of Nanog protein following induction is shown in right panel with piggyBac-GFP transfected control iPS colonies.

C) Nanog chromatin immunoprecipitations indicate that the Nanog transgene is binding its
endogenous genomic targets, Oct4 promoter and Nanog enhancer, following doxycycline induction in EOS3F-24 iPS cells.

**Figure 2.13.** Gene expression analysis following Nanog induction in partial iPS cells.

qRT-PCR analysis of **A)** pluripotency-related genes, Nanog and Stella, and differentiation-specific gene Gata6 and **B)** retroviral transgenes pMX-Oct4, pMX-Sox2 and pMX-Klf4. Gene expression is analyzed in piggyBac-Nanog or -GFP control vector transfected cells in presence or absence of DOX and shown relative to non-transfected EOS3F-24 cells.
2.3.7. **Nanog and Sall1 Co-operatively Bind Major Satellite Repeats and Induce Heterochromatin Fibre Decompaqction**

To determine whether Nanog can directly affect heterochromatin fibre decompaction, I compared partial EOS3F-24 cells with or without *Nanog* induction. Compared to PB-GFP transfected control cells, PB-Nanog induction in partial iPS cells results in emergence of chromocentres with decompacted heterochromatin fibre structure, not observed in partially reprogrammed iPS cells lacking *Nanog* expression (Figure 2.14A). We further investigated the apparent decompaction found in Nanog-positive partial iPS cells by comparing Nanog-null and WT ES cells (Chambers et al, 2007). We find abundance of compact fibre chromocentre organization in Nanog-null cells similar to those found in partially reprogrammed iPS cells suggesting that Nanog may be sufficient to induce heterochromatin decompaction observed in pluripotent iPS and ES cells (Figure 2.14B). Nanog ChIPs demonstrate that this transcription factor is enriched at major satellite repeats, genomic regions comprising parts of chromocentre structures, upon doxycycline induction of the *Nanog* transgene in partial iPS cells (Figure 2.14C). This finding raises the possibility that Nanog can interact with pericentric repetitive DNA sequences. It is not clear whether Nanog is capable of directly binding its genomic loci on its own since major co-occupancy in large complexes with other TFs has been found at target genes, including members of the Sall protein family (Karantzali et al, 2011). Several Sall proteins have been found to occupy chromocentre regions by directly binding major satellite repeat regions (Yamashita et al, 2007). In my PB-mediated expression system, *Nanog* induction also results in Sall1 binding to major satellite repeats, and this association is absent in GFP-induced control partial iPSCs (Figure 2.14C). Conversely, absence of Sall1 in Sall1 KO ESCs, results in lack of binding of Nanog to major satellite repeats (Figure 2.14D). The absence of Nanog and Sall1 binding to major satellite repeats is also reflected in the
heterochromatin structure of Sall1 KO ESCs, where highly compact chromatin fibres of chromocentres are reminiscent of the organization observed in partial iPSCs and somatic cells (Figure 2.14E). Collectively, these results indicate that the pluripotency-related transcription factors Sall1 and Nanog co-operate to bind major satellite repeat sequences within pluripotent mouse cells and can induce decompaction of heterochromatin fibres observed in fully reprogrammed iPS cells.
Figure 2.14. Nanog and Sall1 co-operate to bind major satellite repeats.

A) ESI micrographs of partially reprogrammed EOS3F-24 iPS cells in presence of either GFP (PB-GFP+DOX) or Nanog (PB-Nanog+DOX) induction with doxycycline and B) Nanog-null ES cell lines. C) ChIP assays of Nanog and Sall1 in Nanog- or GFP-expressing EOS3F-24 iPS cells following 7 days of doxycycline treatment. Primers specific to the major satellite repeats were used to evaluate Nanog and Sall1 enrichment and quantified based on enrichment relative to input. Error bars represent S.E.M. from triplicate PCR reactions. D) ChIP assays of Nanog and Sall1 in Sall1-/- and Sall1 +/- ES cells. Primers specific to the major satellite repeats were used to evaluate Nanog and Sall1 enrichment and quantified based on enrichment relative to input. E) ESI micrographs of Sall1-null and WT ES cell lines.
2.4. Discussion

We used LM/ESI to identify novel heterochromatin structure reorganization during reprogramming into iPS cells. We demonstrate that constitutive heterochromatin in chromocentres transitions from a very compact chromatin fibre density in MEFs and partial iPS cells to a more open domain of loosely packed chromatin fibres in ES and full iPS cells. We expand on previous observations of generally dispersed chromatin in ES cells (Efroni et al, 2008) and demonstrate that heterochromatin specifically enriched in H3K9me3 is composed of 10 nm fibres in ES and full iPS cells. We propose that heterochromatin fibres in ES and full iPS cells have a euchromatin-like dispersed 10 nm fibre organization. This is compatible with the general concept that pluripotent stem cells have more open chromatin structure, and conceivably makes the cells more responsive to differentiation cues that they receive. We were surprised to observe 10 nm chromatin fibres in the very densely packed chromocentres of MEFs and partial iPS cells and failed to detect obvious 30 nm fibres, although we cannot exclude the possibility that 30 nm fibres exist in these cells. Nevertheless, these results challenge the textbook view that constitutively silent heterochromatin domains are composed exclusively of higher order 30 nm fibres. Instead, we propose that the increased heterochromatin compartmentalization in MEFs and partial iPS cells could be largely regulated through degrees of bending or modulating inter-fibre distances between adjacent 10 nm fibres.

To directly demonstrate that heterochromatin reorganization is a late event in reprogramming, I converted partial iPS cells into full iPS cells via 2i treatment. I thus confirm that 2i conversion is an effective means to complete reprogramming and that heterochromatin reorganization coincides with other late molecular events in reprogramming including
retrovirus silencing and Nanog activation. These events are specific for ES and full iPS cells, and therefore are not induced by the combined expression of Oct4, Sox2 and Klf4 in the primary transduced MEFs or the partial iPS cells. Since endogenous Nanog expression is a reliable marker of fully reprogrammed mouse iPS cells, we took advantage of its natural variegation in ES cells and blastocysts. These analyses show that high Nanog ES cells have the most dispersed heterochromatin and are the most likely to silence retrovirus vectors, whereas low Nanog cells have less dispersed heterochromatin. These findings are confirmed in vivo during normal development, where heterochromatin is dispersed most in Nanog high cells and is more compartmentalized in Nanog negative differentiated cells. Thus, expression of Nanog or its targets in the pluripotency network correlate with late events of heterochromatin reorganization and retrovirus silencing during reprogramming. This finding is consistent with the known ability of Nanog to both activate pluripotent gene expression and silence genes normally expressed in somatic cells. Collectively these findings imply that heterochromatin reorganization is dependent on establishment of the endogenous pluripotency transcriptional network or perhaps due to a few pluripotency-associated factors. Due to the variable heterochromatin fibre compaction observed when analyzing bulk population of both ES cells and cells of the ICM, I explored the involvement of a variegated transcription factor, Nanog, and its DNA binding partner Sall1. I demonstrate that absence of either Nanog or Sall1 in KO ES cells leads to heterochromatin fibre compaction and overexpression of Nanog alone in partially reprogrammed iPS cells can induce heterochromatin decompaction. However, we cannot exclude the involvement of other epigenetic pathways in concert with these two pluripotency-associated TFs and experimental manipulation of such pathways in
partial iPS cells could promote retrovirus silencing and/or direct chromocentre reorganization during the completion of the reprogramming process.

2.5. Materials and Methods

2.5.1. ES/iPS cell culture

ES and iPS cell lines were grown on irradiated feeders, provided by SickKids Embryonic Stem Cell Facility, in presence of recombinant LIF. iPS cell lines were maintained under 1µg/ml puromycin selection (EOS pluripotency reporter contains a puromycin resistance cassette). 2i conversion was performed using MEK inhibitor (PD0325901, 0.5 µM) and GSK3 inhibitor (CHIR99021, 3 µM) purchased from StemGenet, represented as [1x] concentration in the text. Blastocysts were collected from uteri of ICR outbred mice at embryonic day 3.75 and fixed in 4% formaldehyde solution in PBS for 15 min at RT.

2.5.2. qRT-PCR

1 µg of RNA, collected using TRIZOL (Invitrogen), was reverse transcribed using SSII RT kit with oligo dT and random hexamer primers (Invitrogen). 50 ng of cDNA was used for qRT-PCR using SYBR green master mix (ABI). The primer sequences are in table 2.1.

2.5.3. Immunocytochemistry and Western blot analysis

Cells were prepared for immunolabelling as described (Ahmed et al, 2009), except for blastocysts which were washed in PBS-Tween and permeabilized with 0.2% Triton-X 100. Primary antibodies used were: rabbit anti-H3K9me3, rabbit anti-H4K20me3, human anti-CREST (Immunovision), rabbit anti-Nanog (Abcam), Cdx2 (BioGenex), Gata6 (R&D) and SSEA1 (Developmental Studies Hybridoma Bank). Secondary antibodies used were: donkey anti-rabbit or mouse Cy2, Cy3 and Cy5 (Jackson Laboratories) and PE-Cy5 (eBiosciences),
Alexa488 and Alexa633 (Molecular Probes). Images were collected on a Leica Microsystems DMRA2 microscope equipped with a Hamamatsu ORCA-ER camera or an Olympus IX81 inverted microscope equipped with a Cascade II CCD (Photometrics) camera with or without the spinning disc. OpenLab 3.5.1 (Improvision) software was used to collect images.

2.5.4. **Microarray**

Total RNA isolated from different clones was labeled and hybridized to the Affymetrix GeneChIP Mouse Gene 1.0 ST gene expression microarrays using standard conditions. Intensity values from biological triplicates were processed by variance stabilization and normalization (Huber et al, 2002) and summarized by RMA (Irizarry et al, 2003), iPS expression profiles were contrasted both to MEFs and ES J1s profiled. Differentially expressed transcript in each contrast with an adjusted p-value less than 0.05 where determined using limma (Smyth, 2004).

2.5.5. **Chromatin Immunoprecipitation**

Chromatin immunoprecipitation was performed using the Upstate chromatin immunoprecipitation kit. 1.0x10^6 cells were crosslinked in 1% paraformaldehyde, cells were washed, scraped into a conical tube and sonicated for 25 minutes in the Bioruptor™ XL at the high setting with cycles of 30 seconds ON, 30 seconds OFF. 4 µg of anti-acetyl histone H3 (Upstate), anti-Nanog (Cosmobio) or Sall1 (Santa-Cruz) was used for the precipitation and DNA was purified using phenol/chloroform extraction method and DNA was resuspended in 50µl of H2O. 1 µl of DNA was subjected to qPCR using SYBR green (Primers are listed in Table 2.1).
2.5.6. Bisulfite Sequencing

DNA was isolated from iPS cell lines using phenol/chloroform extraction method and 1µg of DNA was bisulfite converted using EZ DNA methylation Gold Kit (Zymo Research). 50 ng of converted DNA was subjected to PCR (Table 1). PCR reactions were subcloned into TOPO-TA vectors (Invitrogen) and sequenced at the Toronto Centre for Applied Genomics (TCAG).

2.5.7. Correlative LM/ESI Microscopy

Following immunolabeling cells H3K9me3 or H4K20me3 cells were post fixed in 1% glutaraldehyde, dehydrated and embedded in Quetol (Electron Microscopy Sciences). Samples were sectioned by an ultramicrotome (Leica) 70-nm. Energy filtered electron micrographs were taken on a transmission electron microscope (Tecnai 20, FEI) operated at 200 kV collected using a GATAN post column imaging filter at 120 and 155, and 385 and 415 eV to generate the phosphorus and nitrogen images, respectively. Pre- and post-edge images were recorded on a CCD camera. Digital micrograph software was used to collect images. Nitrogen images were subtracted from phosphorus images, so that the net nitrogen signal in chromatin structures were normalized to zero. These phosphorus subtracted nitrogen images are colored blue. The phosphorus images are pseudo-colored yellow and overlaid onto the phosphorus subtracted nitrogen image. Images were processed with Digital micrograph and Photoshop 7.0 (Adobe).

2.5.8. PiggyBac-Nanog Transfection and Induction

PiggyBac-Nanog transfections into EOS3F-24 iPSCs were performed using Lipofectamine 2000 (Invitrogen) and using 500 ng of three constructs: pTET-Nanog-IRES-EGFP, pCAG-Rtta-Puromycin and piggyBac transposase. Selection using 1.5 µg/µl puromycin was applied
for three days followed by 200 ng/µl doxycycline induction of the Nanog transgene driven by the TET promoter. Nanog transgene expression was assessed by qRT-PCR, flow cytometry for EGFP and immunostaining after 3 days of doxycycline induction.

2.5.9. DAPI linescan analysis

Cells used for DAPI line scan analyses were mounted in 1 mg/mL paraphenlenediamine and 1 µg/mL 4’,6-Diamidino-2-phenylindole (DAPI) or soaked in 35 ng/mL DAPI. Images were collected on a Leica Microsystems DMRA2 microscope equipped with a Hamamatsu ORCA-ER camera or an Olympus IX81 inverted microscope equipped with a Cascade II CCD (Photometrics) camera with or without the spinning disc. OpenLab 3.5.1 (Improvision) software was used to collect images. Analyses of the DAPI linescans were performed using Adobe Photoshop, each linescan was drawn to maximize the number of captured chromocentres and the compaction levels were estimated relative to surrounding DAPI-poor nuclear regions.

Table 2.1. Primers used in chapter 2.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rex-1</td>
<td>GGAAGAAATGCTGAGGTTGAGGAGAC</td>
<td>AGTCCCCCATCCCCCTTCAATAGC</td>
<td>qRT-PCR</td>
</tr>
<tr>
<td>Oct4</td>
<td>TCTTCCACGGCCCCCGGCTC</td>
<td>TGCGGGCGACATGGGAGATCC</td>
<td>qRT-PCR</td>
</tr>
<tr>
<td>Nanog</td>
<td>CAGGAGTTTAGGCTAGGTC</td>
<td>CGGTCTCATCTGAGTACGTC</td>
<td>qRT-PCR</td>
</tr>
<tr>
<td>Dnmt3b</td>
<td>TGTTAACCTTTCAGTTGTC</td>
<td>GAAGCAATGAGGAGCTCTCTTG</td>
<td>qRT-PCR</td>
</tr>
<tr>
<td>Dppa3</td>
<td>CTITTGTGTGGGCTGCTGAA</td>
<td>GGAAAGAAATGCTGAGGAGGAGAC</td>
<td>qRT-PCR</td>
</tr>
<tr>
<td>Sox2</td>
<td>CATGAGAGCAAGTACTGGAAG</td>
<td>CCAACGATATCAAACCTGCATGG</td>
<td>qRT-PCR</td>
</tr>
<tr>
<td>c-Myc</td>
<td>GCTTCGAAACCTTGGTGCTG</td>
<td>CCCCTTTCGAGACTGAATAAA</td>
<td>qRT-PCR</td>
</tr>
<tr>
<td>pMX-Klf4</td>
<td>CCACTTTTCCTTACATGAGA</td>
<td>CTTTTATATATCGTGACC</td>
<td>qRT-PCR</td>
</tr>
<tr>
<td>pMX-Oct4</td>
<td>TCTCCCATGACATTCAAACACTG</td>
<td>CTTTTATATATCGTGACC</td>
<td>qRT-PCR</td>
</tr>
<tr>
<td>pMX-Sox2</td>
<td>CTCGGGACATGACGAGC</td>
<td>CTTTTATATATCGTGACC</td>
<td>qRT-PCR</td>
</tr>
<tr>
<td>Stella</td>
<td>AGGGCTGAGAAGGAAATGGAGTTG</td>
<td>TCCCTATATTCTCCCCGATTG</td>
<td>qRT-PCR</td>
</tr>
<tr>
<td>Gata6</td>
<td>ACAGCCCATCTCTGTGACC</td>
<td>GCTGGTGAGCAGGTAACG</td>
<td>qRT-PCR</td>
</tr>
<tr>
<td>Nanog-bis</td>
<td>GATTAGTTATGTGGATATCAATTTG</td>
<td>ACCATATATATATCTCAATACG</td>
<td>bisulfite seq</td>
</tr>
<tr>
<td>Oct4-bis</td>
<td>GTAATGCTGTTGTTGTTGTTTGTGTTTT</td>
<td>CCACCTCCTACTCTAACCCTCTGTAAC</td>
<td>bisulfite seq</td>
</tr>
<tr>
<td>pMX-bis</td>
<td>TTTATTGTAGGTGGTAAGTTAGTTTAA</td>
<td>ATATCCACAAAATAAACTTTCC</td>
<td>bisulfite seq</td>
</tr>
</tbody>
</table>
2.5.10. Flow Cytometry

Cells in tissue culture were trypsinized (Gibco Trypsin-EDTA, 0.05%) and suspended in PBS with 5% FBS. For Ssea1 analyses, trypsinized cells were stained with anti-SSEA1 antibody (Developmental Studies Hybridoma Bank) at a 1:500 dilution in 5% FBS in PBS, washed three times, and stained with the AlexaFluor555 secondary antibody (Invitrogen) and analyzed by FACScan. Data was collected on the LSRII flow cytometer (Becton Dickinson) using CellQuest software (Becton Dickinson) at the SickKids Flow Cytometry Facility and analyzed with FlowJo (Tree Star).
3. **Chapter 3. The e1 Isoform of MECP2 Controls Gene Expression and Neuronal Morphology**

Work presented in this chapter is a part of a manuscript in preparation titled “The E1 isoform of MECP2 controls neuronal form and function in Rett syndrome” with the following authors: Ugljesa Djuric, Aaron Y. L. Cheung, Wenbo Zhang, Alina Piekna, Jason A. Hendry, P. Joel Ross¹, Peter Pasceri, Dae-Sung Kim, Michael W. Salter, James Ellis

**Contributions:**

I performed DNA methylation, AR assays, gene expression, western blot and Fluidigm single-cell analysis. AYL performed the soma size analysis. AP generated the iPS cell lines and performed all the neuronal differentiations. JAH generated the R-software scripts for analysis of Fluidigm data.

3.1. **Abstract**

MECP2 mutations cause the neurodevelopmental disorder Rett Syndrome (RTT) by consistently altering the protein encoded by the MECP2e1 alternative transcript. While mutations on the X chromosome that inactivate both MECP2 isoforms have been widely studied, the function of the obligate MECP2e1 isoform remains unknown. Here we report the first isoform-specific human induced pluripotent stem cell model of RTT and identify MECP2e1-specific roles in patient-derived neurons. Single cell profiling of MECP2e1 mutant neurons demonstrate that they have a molecular signature of cortical neurons, retain an inactive X-chromosome and misexpress MECP2-target genes. By expressing MECP2e1 from a lentiviral transgene we elicited a cell-autonomous soma size increase comparable to normal
MECP2-expressing neurons. We conclude that MECP2e1 binds gene promoters and regulates transcription that ultimately modulates soma size in human neurons that are defective in RTT.

3.2. Introduction

Rett Syndrome (RTT) is a neurodevelopmental disorder characterized by repetitive hand motions, intellectual disability, loss of acquired language and general lack of social behavior (Chahrour & Zoghbi, 2007). Heterozygous loss-of-function mutation in the X-linked gene encoding Methyl-CpG Binding Protein 2 (MECP2) is the prime cause of RTT in girls (Amir et al, 1999). This gene is alternatively spliced into MECP2e1 and MECP2e2 isoforms that encode distinct proteins differing at the N-termini due to exclusion or inclusion of exon 2 respectively (Kriaucionis, 2004; Mnatzakanian et al, 2004). Mutations that affect both isoforms have been widely studied, and the role of MeCP2 in binding methylated and hydroxy-methylated cytosine genome-wide (Mellen et al, 2012) to recruit chromatin-remodeling proteins that modulate global transcription is now well established (Chahrour et al, 2008; Skene et al, 2010).

All common RTT patient mutations affect both MECP2 isoforms but since its discovery in 2004 the distinct function of the MECP2e1 isoform remains unknown due to a lack of a MECP2e1-specific human cell model. However, a MECP2e1-specific mutation that does not alter MECP2e2 is sufficient to cause RTT in human patients (Mnatzakanian et al, 2004) and Mecp2e1-specific mutation recapitulates some of the known RTT phenotypes (Yasui et al, 2014). In contrast, no MECP2e2-specific mutations have been described in RTT patients and Mecp2e2-mutant mice lack neurological phenotypes (Itoh et al, 2012). Appropriate levels of expression of either Mecp2e1 or Mecp2e2 can improve RTT phenotypes in Mecp2-null mice (Kerr et al, 2012) suggesting that MECP2e1 is essential for normal brain
function but manipulation of total MECP2 expression levels can ameliorate MECP2e1-associated phenotypes.

We, and others, reported the generation of human and mouse induced pluripotent stem cells (hiPSCs and miPSCs, respectively) from RTT patients and mouse models that carry pathogenic mutations in both MECP2 isoforms (Ananiev et al, 2011; Cheung et al, 2011; Kim et al, 2011). Neuronal differentiations of RTT iPSCs reveal neuronal maturation and electrophysiological defects reminiscent of those seen in RTT patients and mouse models (Cheung et al, 2012; Farra et al, 2012) and are amenable to rescue by introduction of exogenous MECP2 or compounds such as IGF1 (Marchetto et al, 2010). Generally, female RTT-hiPSCs retain an inactive X-chromosome (Xi) (Pomp et al, 2011; Tchieu et al, 2010) and express either the wild-type (WT) or mutant MECP2 allele and this expression pattern is conserved upon differentiation into neuronal cell types (Cheung et al, 2012). Here, we generated hiPSC-derived neurons that express mutant MECP2e1. Using this system, we show that MeCP2e1 control target gene expression and neuronal soma size indicating that mutations in the e1 splice variant alone leads to production of neurons with altered morphologies consistent with phenotypes observed in RTT patients.

3.3. Results

3.3.1. Characterization of RTTe1 iPSC Cell Lines

RTTe1-hiPSCs were generated via retroviral transduction of OCT4, SOX2, KLF4, and c-MYC into fibroblasts acquired from a RTT patient with an 11 bp deletion in exon 1 of MECP2 (RTTe1). This mutation causes a frameshift resulting in truncated MECP2e1 but an intact MECP2e2 (Mnatzakanian et al, 2004). I used the Androgen Receptor (AR) assay to determine the X-inactivation status in established RTTe1-hiPSCs, revealing that all 22
RTTe1-hiPSC lines exhibited XCI skewing (80:20 ~ 100:0) towards the same parental X-chromosome (Figure 3.1). We selected four RTTe1-hiPSC lines for further characterization and quantitative RT-PCR analyses demonstrate silencing of exogenous reprogramming factors and activation of pluripotency-related genes in the established hiPS cell lines (Figure 3.2A and B). Silencing of retroviral transgenes is indicative of a fully reprogrammed iPS cell state and I performed bisulfite sequencing demonstrating that silencing is accompanied by DNA methylation of CpG dinucleotides within the retroviral LTR promoters (Figure 3.2C). Directed in vitro differentiation of all four RTTe1 iPS cell lines demonstrates their differentiation capacity towards cell types of all three germ layers, mesoderm, ectoderm and endoderm (Figure 3.2D). Furthermore, the cell lines are able to give rise to all three germ layers in in vivo differentiation assays of teratoma formation (Figure 3.2D) demonstrating that the selected iPS cell lines are pluripotent both in vivo and in vitro. Finally, karyotype analysis confirms the appropriate chromosomal profile of all selected iPS cell lines (Figure 3.2E).

**Figure 3.1. Androgen receptor methylation screen of RTTe1 iPS cell lines.**

Androgen receptor methylation screen of RTTe1-hiPSCs demonstrates skewed XCI ratios in all examined lines. Bar graph depicts XCI ratio of two X chromosomes (X1 and X2). Fib., fibroblasts.
Figure 3.2. Pluripotency characterization of RTTe1 iPS cells.

qRT-PCR analyses of A) pMX reprogramming retroviral transgenes and B) endogenous pluripotency genes in RTTe1-hiPSC lines. Data are expressed as mean ± SEM. C) Bisulfite sequencing of retroviral pMX-LTR reprogramming vectors in RTTe1-hiPSC lines. Open and closed CpG sites
indicate unmethylated and methylated CpG sites, respectively. D) RTTe1-hiPSCs differentiate into the three germ layers *in vitro* and *in vivo* (representative images of RTTe1-hiPSC #39). Scale bars, 50 µm (immunocytochemistry) and 100 µm (histology). E) G-banding analysis demonstrate normal karyotypes of RTT-hiPSCs.

3.3.2. Identification of Mutant MECP2e1-expressing Human iPS Cell Lines

AR assay screen demonstrated that the same X-chromosome is active in the established RTTe1 iPS cells and *MECP2*. I performed cDNA sequencing with primers spanning the 11 bp deletion demonstrating that the four selected RTTe1-hiPSCs express the mutant allele (*Figure 3.3A*). To ensure that neither a change in the inactive X chromosome nor Xi erosion (Mekhoubad et al, 2012) occurred during differentiation, we performed AR assay and cDNA sequencing on RTTe1-neurons (see Figure 3.4 and 3.5 for neuronal characterization), confirming expression of only mutant *MECP2e1* in the chosen RTTe1-hiPSC lines following differentiation (*Figure 3.3B*).
Figure 3.3. RTTe1 iPSCs and neurons retain an Xi and exclusively express the mutant MECP2e1 allele.

A) cDNA sequencing of MECP2e1 transcripts in selected RTTe1-hiPSCs reveals that the expressing X-chromosome contains the 11 bp RTTe1 deletion (left panel), with representative chromatogram (right panel).

B) Differentiated RTTe1-neurons maintain expression of the MECP2e1 mutant X-chromosome based on MECP2e1 cDNA sequencing (right panel) and the AR assay (right panel).

3.3.3. Neuronal Differentiation and Lentiviral Rescue of RTTe1 Cells

To generate neurons in vitro, we identified three protocol variations that differentiate each of the four RTTe1-hiPSC lines into neuronal progenitor cells (NPCs) and neurons based on previously described protocols (Brennand et al, 2011; Kim et al, 2012). I observed a 10-fold increase in MECP2 mRNA upon transition of RTTe1-NPCs into neurons (Figure 3.4A). Since we were unable to establish RTTe1-hiPSCs expressing the wild-type allele, I compared RTTe1-neurons to those transduced by a lentiviral vector expressing MECP2e1-MYC under
the control of endogenous *MeCP2* (MeP) promoter (Rastegar et al, 2009). The MeP lentiviral rescue construct specifically expresses in mature neurons but not in NPCs, in contrast to the ubiquitously-expressing EF1α construct (Figure 3.4C). I performed western blot analysis and while MECP2e2 protein levels remain unaffected, MeP vector transduction resulted in specific increase of *MECP2e1* mRNA and protein in differentiated neurons (Figure 3.4B and D). These results demonstrate that RTTe1 cells have normal levels of MeCP2e2 while only MeP-transduced cells have detectable MECP2e1 protein.
Figure 3.4. MECP2e1 rescue constructs specifically express in mature neurons following differentiation.

A) MECP2 transcription of both isoforms increases over the course of neuronal differentiation. B) MeP-MECP2e1 rescue results in increased MECP2e1 transcripts in mature RTTe1 neurons. Data are expressed as mean ± SEM. C) Immunocytochemistry of MECP2 and MYC in MECP2e1 transduced RTTe1-NPCs and upon differentiation into RTT-neurons. Scale bars, 44 μm. D) Western blot analysis shows that MeP-MECP2e1 rescue results in recovery of detectable levels of MECP2e1 and overexpression by
transduction with EF1α-MECP2e1 constructs, densitometry quantification is shown (right panel). Histone H3 western blot is used as a loading control.

3.3.4. Single Cell Analysis of RTT iPS-derived Neuronal Differentiations

To determine the neuronal identity and differentiation efficiency of RTTe1 iPS cell-derived neurons, I performed single cell expression profiles using Fluidigm arrays (Pașca et al, 2011). Single-cell sorting was carried out by sorting collagenase-treated neurons directly into the reverse transcriptase mix in individual wells of 96-well plates. Data for the four RTTe1 mutant cell lines was pooled and compared to the RTTe1#27 neurons rescued with the MECP2e1 lentiviral construct (RTTe1-MeP). More RTTe1-MeP cells were expressing the mature neuronal markers DCX, NCAM and MAP2 compared to RTTe1 mock neurons (Figure 3.5A). The expression of these neuronal markers is upregulated following differentiation of neuronal precursor cells (NPCs) as the progenitor-specific markers (NES, PLZF, ZO1) are downregulated (Figure 3.5B). Importantly, of the neurons produced both RTTe1 mock and RTTe1-MeP had comparable neuronal types (Figure 3.5C). These included neurons with a dorsal forebrain identity indicated by PAX6 expression and an equal (~35:45%) mixture of glutamatergic and GABAergic neurons. I also analyzed control MECP2-null and isogenic WT-neurons (Cheung et al, 2011) and show that complete absence of MECP2 does not alter the type of neurons generated from hiPSCs (Figure 3.6). Together, these results indicate that lack of MECP2e1 or both isoforms has minimal effects on neuronal differentiation fate in vitro and that cortical neurons relevant to RTT phenotypes were generated using our differentiation protocols. Separate work by our collaborator, Dr. Mike Salter, demonstrates
that these RTTe1 neurons exhibit electrophysiological features representative of mature functional neurons.

Figure 3.5. RTTe1 neurons maintain normal balance of neuronal identity.

A) Bar graphs show comparable percentage of cells expressing majority of neuronal markers as determined by Fluidigm arrays in RTTe1-mock and RTTe1-MeP rescued cells. Data are expressed as mean ± SEM; * p-value < 0.05. B) Single-cell (x-axis) heatmaps of Ct values for neuronal
progenitor \( (NES, PLZF, ZO1) \) and mature markers \( (MAP2, NCAM, DCX) \) (y-axis) in RTTe1 NPCs and differentiated neurons. C) Similar mixture of GABA and Glutamatergic neurotransmitter neurons are produced from both RTTe1-mock and MeP rescued cells. Data are expressed as mean ± SEM.

Figure 3.6. MECP2-mutant neurons maintain normal balance of neuronal identity.

Fluidigm array on independently derived Δ3-4 WT and null neurons. A) Bar graph depicts the percentage of cells expressing indicated genes. Data are expressed as mean ± SEM. * p <0.05.

3.3.5. MECP2e1 Affects Transcription Levels of Target Genes

To investigate whether RTTe1 neurons exhibit dysregulation of MECP2 target genes, I next examined candidate genes that are regulated by Mecp2 in mouse neurons (Chahrour et al, 2008; Skene et al, 2010). Most of these genes were upregulated during differentiation into neurons (Figure 3.7A). By comparing MeP rescued relative to RTTe1-neurons, and WT relative to \( MECP2 \)-null neurons, we observed transcriptional changes of up to three-fold activation of \( OPRK1, GAMT \) and \( EYA2 \), whereas the other genes were relatively unchanged (Figure 3.7B). These data are consistent with MECP2e1 being an activator of some genes in
human neurons, but the fold change in target gene expression levels is relatively subtle. To examine whether these transcriptional changes are due to altered MECP2 binding, I performed chromatin immunoprecipitation (ChIP) assays in mature RTTe1 and Δ3-4 neurons. ChIP analysis revealed enrichment of MECP2 at all investigated loci in WT neurons, and this binding was reduced in RTTe1 neurons that have normal MECP2e2 levels (Figure 3.8A). Furthermore, RTTe1 MeP neurons had an increase in MYC antibody enrichment at most loci indicating direct binding by tagged MECP2e1 (Figure 3.8B). As a negative control Δ3-4#20 mutant neurons, lacking any MECP2 protein, were used indicating the specificity of the MECP2 antibody used as no enrichment was detected on chromatin isolated from these neurons compared to WT Δ3-4#37 neurons. These data support the concept that MeCP2e1 is a transcription modulator that binds to and regulates target genes in human neurons similar to what has been previously shown in mouse models (Chahrour et al, 2008; Urdinguio et al, 2008).
Figure 3.7. Transcriptional consequences of MECP2 target genes upon neuronal differentiation of RTTe1 and Δ3-4 iPS cells.

A) qRT-PCR analysis of putative MECP2 target genes during the 4-week neuronal differentiation period in MECP2 WT hiPSC-derived neurons. Data are presented as relative to NPC as mean ± SEM. B) qRT-PCR analysis of MECP2 target genes in differentiated RTTe1 MeP or Δ3-4#37 WT neurons, expressed relative to either RTTe1 Mock and Δ3-4#20 (MECP2 null), respectively. Data are expressed as mean ± SEM.
Figure 3.8. MECP2e1 mutation results in absence of MECP2 association with its genomic targets.

Chromatin immunoprecipitation analysis in MECP2 null (Δ3-4#20), WT (Δ3-4#37) and RTTe1 neurons, shows A) MECP2 and B) MYC-tagged MECP2e1 binding to all investigated loci. qRT-PCRs were carried out using primers in the 5’ region of indicated genomic loci. Data are expressed as mean ± SEM.
3.3.6. *MECP2e1* Mutation Causes Decreased Soma Size Phenotype in Neurons

RTT is thought to be a neurodevelopmental disorder with defects in neuronal maturation and/or maintenance (Kishi & Macklis, 2004; Nguyen et al, 2012). To determine whether the loss of *MECP2e1* alone results in a neuronal phenotype, we measured soma size in RTTe1-neurons. RTTe1-hiPSCs were able to differentiate into MAP2-positive neurons, which expressed negligible MECP2, consistent with continued low-level expression of *MECP2e2* (Figure 3.9A). RTTe1-neurons exhibited a significant decrease in soma size compared to WT-neurons (Figure 3.9B).

To test whether the soma size defect is due to *MECP2e1* mutation, RTTe1-NPCs were infected with *MECP2e1* lentiviral vectors to rescue the soma size phenotype in differentiated RTTe1-neurons. Transduced NPCs expressed the ubiquitous EF1α-vector while the MeP construct was only activated upon neuronal differentiation (See Figure 3.4). These results demonstrate temporal regulation of the MeP promoter during maturation of human neurons in vitro. Co-staining with MYC revealed that roughly half of the RTTe1-neurons expressed the lentiviral transgene (Figure 3.9A). Therefore, soma size was scored in MYC-positive relative to MYC–negative neurons (Figure 3.9B). MeP transduced MYC positive neurons showed a soma size increase equivalent to WT-neurons, while adjacent MYC negative neurons had a smaller soma size equivalent to untransduced RTTe1-neurons. These results reveal that soma size rescue by *MECP2e1* is cell autonomous. In contrast, EF1α transduced MYC positive neurons displayed a minor increase in soma size that may be a consequence of transgene overexpression from the strong promoter. We conclude that MeCP2e1 is a cell autonomous regulator of soma size, and the rescue experiment demonstrates that the defect in neurons is due solely to a lack of MeCP2e1 isoform.
Figure 3.9. RTTe1 neurons exhibit a soma size defect that can be rescued in a cell autonomous manner.

A) Immunocytochemistry for MAP2 and MYC or MAP2 and MECP2 in RTTe1-neurons with or without MECP2e1-vectors and WT-neurons. Scale bars, 44 μm for large image, 10 μm for inset. Arrows, MYC-positive neurons. Arrowheads, MYC-negative neurons. B) Soma size analysis of RTTe1-
neurons with or without MECP2e1-vectors compared to WT-neurons (**, \(P < 0.0001\); *, \(P < 0.001\); Student’s t-test; \(n = \) independent differentiations indicated at the bottom of each bar). Data are expressed as mean ± SEM.

3.4. Discussion

In this work I investigated the function of the MECP2e1 isoform in human neurons through the use of a RTT iPSC cellular model. RTTe1-hiPSCs retained an Xi allowing the generation of mutant MECP2e1 neurons upon differentiation. This finding serves as another example of human iPS cell lines retaining an inactive X chromosome upon reprogramming. Although no iPS cell lines expressing the wild-type MECP2e2 allele were isolated, it is possible that this is due to passaging of the starting fibroblasts prior to reprogramming (Pomp et al, 2011). Single cell Fluidigm arrays demonstrated that the majority of neurons were cortical in nature with an equal mixture of glutamatergic and GABAergic neurons, and the relative proportion of cell types was unaffected by RTTe1 or MECP2 null mutations. These equal proportions of cell types were appropriate for investigating the consequences of MECP2 mutation on transcription of target genes. Subtle transcriptional dysregulation I identified in a subset of target genes is similar to expression level changes found in Mecp2 target genes in mouse models of RTT (Chahrour et al, 2008; Urdinguio et al, 2008). ChIP experiments revealed MECP2 binding to most investigated loci in WT human neurons, and MeP-rescued neurons also demonstrated enrichment of MYC tagged MECP2e1 binding. We conclude that MECP2e1 directly binds and regulates a set of target genes in human neurons with dysregulated expression in the presence of normal MECP2e2 levels.
We next investigated the effect of MECP2e1 on neuronal form. RTTe1-neurons displayed a soma size defect in comparison to WT-neurons which was rescued with an MECP2e1 transgene. The rescue effect was cell autonomous as only RTTe1-neurons that received the vector and not their uninfected neighbours exhibited a soma size increase. These results are consistent with the finding that nuclear size is cell autonomously regulated by MECP2 in mouse ES cell derived neurons (Yazdani et al, 2012). Taking the rescue experiments together with the reproducibility of the soma size defect in neurons derived from all 4 iPSC lines, we conclude that MECP2e1 controls soma size in human neurons.

In conclusion, we show that the cellular consequences of disease-causing alternatively spliced transcripts can be defined using patient iPS cells. Given that MECP2e2 transgenes rescue RTT in mice (Kerr et al, 2012), it may be possible to identify pathways that regulate alternative splicing (Braunschweig et al, 2013) to enhance exon 2 inclusion, resulting in high levels of the normal MECP2e2 isoform being transcribed from the MECP2e1 mutant allele in this patient.

3.5. Materials and Methods

3.5.1. Androgen Receptor Assay

To obtain the methylated Xi, 200 ng of DNA was digested overnight at 37°C with methylation-sensitive enzymes HpaII and HhaI (Invitrogen). To differentiate between the two parental X-chromosomes, 20 ng of digested and undigested DNA was amplified with primers (Table 3.1) spanning the heterozygous polymorphic trinucleotide repeat in the first exon of the AR gene for 32 cycles. The 5’ end of the forward primer is labeled with FAM fluorescein (Invitrogen). Relative abundance of PCR products were analysed at the Centre for Applied
Genomics (TCAG). In brief, PCR products were separated on an ABI3100 Genetic Analyzer with 500 LIZ size standard and analysed by Peak Scanner software (all from Applied Biosystems). XCI ratio (Table 3.2) was calculated by using a correction factor (calculated using peak areas of the undigested samples) to normalize for preferential amplification of one of the two AR alleles. The XCI ratio was then calculated using the corrected peak area in the digested samples.

3.5.2. Bisulfite Sequencing

Bisulfite conversion was performed as previously described (Fussner et al, 2011b). Briefly 1 µg of DNA was subjected to conversion using the DNA Methylation Gold Kit (Zymo Research) and the DNA was purified and recovered using kit-supplied spin columns. 50 ng of converted DNA was subjected to PCR with appropriate primers (Table 3.1).

3.5.3. Generation and Transduction of Lentivirus

Plasmids containing cDNA of interest were transfected into 293T cells using Lipofectamine 2000 (Invitrogen). The supernatant containing virus was collected 48 hours posttransfection and lentiviruses were concentrated by ultracentrifugation at 4 °C at 30,000 RPM for 2 hours. The viral pellet was soaked in 40 µl HBSS (Invitrogen) overnight at 4 °C and resuspended. Transduction of cells was performed in the presence of 8 µg/ml Hexadimethrine bromide (Sigma). 293T cells were maintained in virus production medium: 10% FBS DMEM, 0.1mM MEM NEAA, and 50 U/ml Penicillin and 50 mg/ml Streptomycin (Invitrogen).

3.5.4. qRT-PCR and cDNA Sequencing

RNA was isolated using Trizol extraction method (Invitrogen) and SSII RT (Invitrogen) was used for the reverse transcription following manufacturer’s instructions. Primers for the Real-
Time PCR assays were designed using Primer3 online primer design software and qPCR was carried out using SYBR green (Applied Biosystems) on an ABI 7900HT PCR System (Applied Biosystems). To sequence RTTe1-hiPSC cDNA, RT-PCR was performed using V1 primers (Table 3.1). The MECP2e1 amplicon (382 bp or 371 bp for WT or mutant allele, respectively) was gel purified from the MECP2e2 amplicon (506 bp) using QIAquick PCR Purification Kit following manufacturer’s instructions and sequenced using the V1-f for amplification.

3.5.5. Immunocytochemistry

Cells were fixed with 4% formaldehyde (EMD Biosciences) for 10 min at room temperature (RT), permeabilized with 0.1% Noniodet P-40 (Sigma). Blocking was performed for 3 hr at RT, primary antibodies diluted in block solution and incubated overnight at 4°C (See Table 3.3 for antibodies). Appropriate AlexaFluor secondary antibodies (Invitrogen) diluted (1:500) in block solution were incubated for one hour at RT. Images were made using Leica DMI4000B for hiPSCs or Zeiss Axiovert 200M for neurons. Soma size of neurons was scored using Improvision Volocity software on 40X images blinded to the observer.

3.5.6. Western Blot Analysis

Nuclear proteins were collected using nuclear extraction protocol, samples were aliquoted and snap frozen in liquid nitrogen for storage until western blots were performed. 5 ug of total protein was loaded for western blots, transferred to nitrocellulose membranes overnight at 4 degrees. Membranes were blocked in 5% Milk PBS-T and incubated overnight in primary antibodies (see Table 3.3), washed 5 times in PBS-T and incubated in appropriated HRP-conjugated secondary antibodies (Invitrogen) for 1 hour at room temperature. Densitometry
measurements and normalization (to Histone H3 signal) was performed using ImageJ software.

3.5.7. Single-cell Fluidigm Arrays

iPS cell lines were differentiated to neurons and neuronal differentiation was assessed by morphological analysis and independent MAP2 immunostaining assays. Neurons were dissociated to single cells using accutase for 10 min at room temperature (Innovative Cell Technologies) treatment and suspended in ice-cold PBS containing 5% FBS. Live cells were stained by propidium iodide (eBioscience) and sorted into a 96 well plate, with preamplification mix at The Flow Cytometry Facility at The Hospital for Sick Children (Toronto, Ontario, Canada) using MoFlo cell sorter (Beckman Coulter). The preamplification mix contains 40 nM of appropriate primers (Table 3.1) and CellsDirect One-Step qRT-PCR Kit (Invitrogen). After sorting, samples were reverse transcribed and pre-amplified for 18 cycles and processed for Fluidigm gene expression analysis. Samples were mixed with 20X DNA Binding Dye Sample Loading Reagent (Fluidigm Corp.), 20X EvaGreen (Biotium), TE buffer (10mM tris-HCl, 0.1mM EDTA, pH8.0 [TEKnova]) and 2X TaqMan Gene Expression Master Mix (Applied Biosystems). Assays were mixed with 2X Assay loading reagent (Fluidigm Corp.) and TE buffer (10mM tris-HCl, 0.1mM EDTA, pH8.0) to a final concentration of 5 µM. We primed the 48.48 Fluidigm Dynamic Arrays (Fluidigm Corp.) and loaded on an IFC Controller HX (Fluidigm Corp.), and ran the GE 48X48 PCR protocol for the qRT-PCR experiments on a Biomark HD System for Genetic Analysis (Fluidigm Corp.). Raw data was analyzed using R statistical software (R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0) and data presented is a combination of at least 3 independent differentiation experiments.
3.5.8. Chromatin Immunoprecipitation

ChIP was performed using EZ Chromatin Immunoprecipitation Kit (Millipore). Briefly, 1x10^6 cells were crosslinked using 1% final concentration of formaldehyde (BIORAD) and crosslinking reaction was quenched using 125 µM of glycine. Cells were sonicated using bioruptor (Diagenode) for 27 minutes using the high setting at 30 seconds ON, 30 seconds OFF, intervals. ChIP was performed using appropriate antibodies for either cMYC tag or MECP2 (Table 3.3). Following precipitation and washes, precipitated DNA was suspended in 20 µl of H₂O and 1 µl was used for each individual qPCR reaction (Table 3.1).

Table 3.1. Primers used in chapter 3.

<table>
<thead>
<tr>
<th>Primers used for Fluidigm, cDNA sequencing and qRT-PCRs:</th>
<th>Forward (5' → 3')</th>
<th>Reverse (5' → 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>MECP2-V1</td>
<td>GTAAAAGCCGGTCCGGAAAT</td>
<td>GCTTAAGCTTCCGGTGTCCA</td>
</tr>
<tr>
<td>NES</td>
<td>GGGCGCATTCAAGATGTC</td>
<td>CTTGCGTCTCGAAGACTG</td>
</tr>
<tr>
<td>PLZF</td>
<td>GGAGCTTTGCGAGTATGGA</td>
<td>AATGCGGTGAAAGAGATCTC</td>
</tr>
<tr>
<td>ZO1</td>
<td>AGTCCCTACCTTTGGCCGTA</td>
<td>TCCCTTAGCATTTTGGAGCTGC</td>
</tr>
<tr>
<td>FOXG1</td>
<td>GCCAACAATCGTCTCCCAACA</td>
<td>CGGGTCCAGACATCCGATAG</td>
</tr>
<tr>
<td>OTX2</td>
<td>CAGGACCCTTACACAGTCA</td>
<td>GGGAGTCAGCAAGTCCATAC</td>
</tr>
<tr>
<td>ZIC1</td>
<td>GTTCGGACGACTATGCTG</td>
<td>TCGACGACTTTTGGGTTG</td>
</tr>
<tr>
<td>PAX6</td>
<td>ATGTGGTGAATATAACTCGGCA</td>
<td>GCTTACACTTGGAGCTGCA</td>
</tr>
<tr>
<td>TBR2</td>
<td>CCGGGCACCTATCATGACAG</td>
<td>GGTTCAGACGGTAGAGCTG</td>
</tr>
<tr>
<td>ASCL1</td>
<td>TCTTCGCCCGAAGACTGATC</td>
<td>CAAAGCCCGGATGACCAAT</td>
</tr>
<tr>
<td>DLXI</td>
<td>CCCATGCAACAAAGACTCTCAACA</td>
<td>GGGCCAAAATCCTCAAATAACCC</td>
</tr>
<tr>
<td>NXX2.1</td>
<td>AGCAAGCAGCTTCTCCCTC</td>
<td>GCCCCACTTTCTTGAGCTTCC</td>
</tr>
<tr>
<td>LHX6</td>
<td>TGAGAGTGAGGGACTACTGCG</td>
<td>GCGCCATCATATCGCTTGAG</td>
</tr>
<tr>
<td>EN1</td>
<td>GACCGAGCGCATCAAATAA</td>
<td>AAAAAGTGTCGTAACCACCC</td>
</tr>
<tr>
<td>MSX2</td>
<td>CACCTCTGGAAGAACAAAGAC</td>
<td>AACTCTGACGGCTGCAAT</td>
</tr>
<tr>
<td>NURR1</td>
<td>TGTGTTCAGCGCGCATGATG</td>
<td>TCCCGGAGGGTTGTAACCTG</td>
</tr>
<tr>
<td>HNK1</td>
<td>CCGAGCGGCTGTCTACTTCCG</td>
<td>GCAATTTGAGCGGAAATCC</td>
</tr>
<tr>
<td>DCX</td>
<td>CTTGCTGTAAGGACGAAAGT</td>
<td>CCACTGCGGATGAGTGAAG</td>
</tr>
<tr>
<td>NCAM</td>
<td>ACATCCACCTGCTCTCTCTG</td>
<td>CTTGGACTCATCTTTGGAGG</td>
</tr>
<tr>
<td>MAP2</td>
<td>CTTGCTTACAGGGTGACAGCAA</td>
<td>TTGAGTAGCTGGAAAAAGCTG</td>
</tr>
<tr>
<td>CAMK2</td>
<td>AAAGCTGAAAGGAGCGATCTCA</td>
<td>GAGGATTCTCAATGAACAGC</td>
</tr>
<tr>
<td>VGLUT1</td>
<td>GCAGCGACGAGCCCTTGTGGT</td>
<td>GCCGTAGACGTAAGAAAGAG</td>
</tr>
<tr>
<td>VGLUT2</td>
<td>GGAGGACAAAATCGAGCTGACG</td>
<td>CAGCGAGATCAGGAAAGAGATG</td>
</tr>
<tr>
<td>VGLUT3</td>
<td>AAACCGGGCAATTCGAGCTACG</td>
<td>CCAAGCACTCCCTGTGACCA</td>
</tr>
<tr>
<td>GAD67</td>
<td>GGGCAAGACAGATGATGATG</td>
<td>CACGTTCAGCCATTGTTGAT</td>
</tr>
<tr>
<td>VGLU1</td>
<td>CCGGTGGAGTGACAGTACGACG</td>
<td>GTGCGCAATIAAGCCAGCACG</td>
</tr>
<tr>
<td>TH</td>
<td>GCCTCTACCAAGACGAGTA</td>
<td>CGTAGAGGCATAGCTCTGA</td>
</tr>
<tr>
<td>SHT2-2C</td>
<td>TCTTTAATGGCTCCTAGGAGTGT</td>
<td>TCCTGATCCCGAGTATAAGC</td>
</tr>
<tr>
<td>FOXP1</td>
<td>AGACAGAAAAATCACGGTGTCG</td>
<td>GCGACATCTGTAAAGTGTCG</td>
</tr>
<tr>
<td>ET1V</td>
<td>CGTGATGGACCCGGCAAAACTT</td>
<td>CCTCTGACGGCTCAATCGTT</td>
</tr>
<tr>
<td>SAB2</td>
<td>TCTCCCTCAGTATGATGAC</td>
<td>AGGCAAGTCTCCAACTTGGAG</td>
</tr>
<tr>
<td>CTIP2</td>
<td>TGGGTGCTGCTATGCAAG</td>
<td>GGTCGGGACACTTCTCTGAG</td>
</tr>
<tr>
<td>CUX1</td>
<td>GCTCTCTATGCGCAATCTCATC</td>
<td>TCTATGGCCGCTCCACTG</td>
</tr>
<tr>
<td>REELIN</td>
<td>TCCGGGACAAGAATACCAGT</td>
<td>CCAAATCCGAAAAGCATGGA</td>
</tr>
</tbody>
</table>

Table 3.3. Primers used for Fluidigm, cDNA sequencing and qRT-PCRs:

<table>
<thead>
<tr>
<th>Primers used for Fluidigm, cDNA sequencing and qRT-PCRs:</th>
<th>Forward (5' → 3')</th>
<th>Reverse (5' → 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>MECP2-V1</td>
<td>GTAAAAGCCGGTCCGGAAAT</td>
<td>GCTTAAGCTTCCGGTGTCCA</td>
</tr>
<tr>
<td>NES</td>
<td>GGGCGCATTCAAGATGTC</td>
<td>CTTGCGTCTCGAAGACTG</td>
</tr>
<tr>
<td>PLZF</td>
<td>GGAGCTTTGCGAGTATGGA</td>
<td>AATGCGGTGAAAGAGATCTC</td>
</tr>
<tr>
<td>ZO1</td>
<td>AGTCCCTACCTTTGGCCGTA</td>
<td>TCCCTTAGCATTTTGGAGCTGC</td>
</tr>
<tr>
<td>FOXG1</td>
<td>GCCAACAATCGTCTCCCAACA</td>
<td>CGGGTCCAGACATCCGATAG</td>
</tr>
<tr>
<td>OTX2</td>
<td>CAGGACCCTTACACAGTCA</td>
<td>GGGAGTCAGCAAGTCCATAC</td>
</tr>
<tr>
<td>ZIC1</td>
<td>GTTCGGACGACTATGCTG</td>
<td>TCGACGACTTTTGGGTTG</td>
</tr>
<tr>
<td>PAX6</td>
<td>ATGTGGTGAATATAACTCGGCA</td>
<td>GCTTACACTTGGAGCTGCA</td>
</tr>
<tr>
<td>TBR2</td>
<td>CCGGGCACCTATCATGACAG</td>
<td>GGTTCAGACGGTAGAGCTG</td>
</tr>
<tr>
<td>ASCL1</td>
<td>TCTTCGCCCGAAGACTGATC</td>
<td>CAAAGCCCGGATGACCAAT</td>
</tr>
<tr>
<td>DLXI</td>
<td>CCCATGCAACAAAGACTCTCAACA</td>
<td>GGGCCAAAATCCTCAAATAACCC</td>
</tr>
<tr>
<td>NXX2.1</td>
<td>AGCAAGCAGCTTCTCCCTC</td>
<td>GCCCCACTTTCTTGAGCTTCC</td>
</tr>
<tr>
<td>LHX6</td>
<td>TGAGAGTGAGGGACTACTGCG</td>
<td>GCGCCATCATATCGCTTGAG</td>
</tr>
<tr>
<td>EN1</td>
<td>GACCGAGCGCATCAAATAA</td>
<td>AAAAAGTGTCGTAACCACCC</td>
</tr>
<tr>
<td>MSX2</td>
<td>CACCTCTGGAAGAACAAAGAC</td>
<td>AACTCTGACGGCTGCAAT</td>
</tr>
<tr>
<td>NURR1</td>
<td>TGTGTTCAGCGCGCATGATG</td>
<td>TCCCGGAGGGTTGTAACCTG</td>
</tr>
<tr>
<td>HNK1</td>
<td>CCGAGCGGCTGTCTACTTCCG</td>
<td>GCAATTTGAGCGGAAATCC</td>
</tr>
<tr>
<td>DCX</td>
<td>CTTGCTGTAAGGACGAAAGT</td>
<td>CCACTGCGGATGAGTGAAG</td>
</tr>
<tr>
<td>NCAM</td>
<td>ACATCCACCTGCTCTCTCTG</td>
<td>CTTGGACTCATCTTTGGAGG</td>
</tr>
<tr>
<td>MAP2</td>
<td>CTTGCTTACAGGGTGACAGCAA</td>
<td>TTGAGTAGCTGGAAAAAGCTG</td>
</tr>
<tr>
<td>CAMK2</td>
<td>AAAGCTGAAAGGAGCGATCTCA</td>
<td>GAGGATTCTCAATGAACAGC</td>
</tr>
<tr>
<td>VGLUT1</td>
<td>GCAGCGACGAGCCCTTGTGGT</td>
<td>GCCGTAGACGTAAGAAAGAG</td>
</tr>
<tr>
<td>VGLUT2</td>
<td>GGAGGACAAAATCGAGCTGACG</td>
<td>CAGCGAGATCAGGAAAGAGATG</td>
</tr>
<tr>
<td>VGLUT3</td>
<td>AAACCGGGCAATTCGAGCTACG</td>
<td>CCAAGCACTCCCTGTGACCA</td>
</tr>
<tr>
<td>GAD67</td>
<td>GGGCAAGACAGATGATGATG</td>
<td>CACGTTCAGCCATTGTTGAT</td>
</tr>
<tr>
<td>VGLU1</td>
<td>CCGGTGGAGTGACAGTACGACG</td>
<td>GTGCGCAATIAAGCCAGCACG</td>
</tr>
<tr>
<td>TH</td>
<td>GCCTCTACCAAGACGAGTA</td>
<td>CGTAGAGGCATAGCTCTGA</td>
</tr>
<tr>
<td>SHT2-2C</td>
<td>TCTTTAATGGCTCCTAGGAGTGT</td>
<td>TCCTGATCCCGAGTATAAGC</td>
</tr>
<tr>
<td>FOXP1</td>
<td>AGACAGAAAAATCACGGTGTCG</td>
<td>GCGACATCTGTAAAGTGTCG</td>
</tr>
<tr>
<td>ET1V</td>
<td>CGTGATGGACCCGGCAAAACTT</td>
<td>CCTCTGACGGCTCAATCGTT</td>
</tr>
<tr>
<td>SAB2</td>
<td>TCTCCCTCAGTATGATGAC</td>
<td>AGGCAAGTCTCCAACTTGGAG</td>
</tr>
<tr>
<td>CTIP2</td>
<td>TGGGTGCTGCTATGCAAG</td>
<td>GGTCGGGACACTTCTCTGAG</td>
</tr>
<tr>
<td>CUX1</td>
<td>GCTCTCTATGCGCAATCTCATC</td>
<td>TCTATGGCCGCTCCACTG</td>
</tr>
<tr>
<td>REELIN</td>
<td>TCCGGGACAAGAATACCAGT</td>
<td>CCAAATCCGAAAAGCATGGA</td>
</tr>
</tbody>
</table>
Table 3.2 Calculation of AR assay, raw data.

<table>
<thead>
<tr>
<th>± HpaII &amp; HhaI</th>
<th>Sample</th>
<th>Peak Area</th>
<th>Corrected Peak Area</th>
<th>XCI Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Allele 1</td>
<td>Allele 2</td>
<td>Allele 1</td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-Fibroblasts</td>
<td>1392</td>
<td>1414</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-hiPSC #1</td>
<td>294</td>
<td>859</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-hiPSC #7</td>
<td>451</td>
<td>295</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-hiPSC #11</td>
<td>257</td>
<td>381</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-hiPSC #16</td>
<td>237</td>
<td>603</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-hiPSC #22</td>
<td>585</td>
<td>870</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-hiPSC #24</td>
<td>243</td>
<td>414</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-hiPSC #27</td>
<td>235</td>
<td>469</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-hiPSC #29</td>
<td>188</td>
<td>390</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-hiPSC #39</td>
<td>632</td>
<td>1277</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-hiPSC #46</td>
<td>901</td>
<td>1627</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-hiPSC #48</td>
<td>1088</td>
<td>1457</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-hiPSC #55</td>
<td>713</td>
<td>1815</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-hiPSC #57</td>
<td>1299</td>
<td>692</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-hiPSC #62</td>
<td>276</td>
<td>516</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-hiPSC #65</td>
<td>441</td>
<td>717</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-hiPSC #68</td>
<td>774</td>
<td>1581</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-hiPSC #75</td>
<td>245</td>
<td>594</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-hiPSC #79</td>
<td>418</td>
<td>1362</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-hiPSC #90</td>
<td>541</td>
<td>1114</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-hiPSC #95</td>
<td>1200</td>
<td>1748</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-hiPSC #96</td>
<td>337</td>
<td>679</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-hiPSC #108</td>
<td>896</td>
<td>1815</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-hiPSC #109</td>
<td>1479</td>
<td>659</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-hiPSC #115</td>
<td>1319</td>
<td>2100</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-neurons #27</td>
<td>251</td>
<td>356</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-neurons #39</td>
<td>815</td>
<td>1071</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-neurons #48</td>
<td>1023</td>
<td>1368</td>
<td></td>
</tr>
<tr>
<td>Undigested</td>
<td>RTTe1-neurons #96</td>
<td>149</td>
<td>315</td>
<td></td>
</tr>
<tr>
<td>Digested</td>
<td>RTTe1-Fibroblasts</td>
<td>4240</td>
<td>1756</td>
<td>4307</td>
</tr>
</tbody>
</table>

Table 3.2 Calculation of AR assay, raw data.
<table>
<thead>
<tr>
<th>Antibody</th>
<th>Company</th>
<th>Catalogue #</th>
<th>Dilution</th>
</tr>
</thead>
<tbody>
<tr>
<td>TUJ1</td>
<td>Chemicon</td>
<td>MAB1637</td>
<td>1:200</td>
</tr>
<tr>
<td>SMA</td>
<td>Invitrogen</td>
<td>18-01106</td>
<td>1:200</td>
</tr>
<tr>
<td>GATA4</td>
<td>Santa Cruz</td>
<td>SC-9053</td>
<td>1:200</td>
</tr>
<tr>
<td>MECP2</td>
<td>Sigma</td>
<td>M6818</td>
<td>1:1000</td>
</tr>
<tr>
<td>c-MYC</td>
<td>Santa Cruz</td>
<td>Sc-789</td>
<td>1:200</td>
</tr>
<tr>
<td>MAP2</td>
<td>Millipore</td>
<td>AB5622</td>
<td>1:1000</td>
</tr>
<tr>
<td>MAP2</td>
<td>Sigma</td>
<td>MI406</td>
<td>1:3000</td>
</tr>
<tr>
<td>MECP2 (ChIP and W. Blot)</td>
<td>Abcam</td>
<td>Ab2828</td>
<td>1:2000</td>
</tr>
<tr>
<td>c-MYC (ChIP and W. Blot)</td>
<td>Millipore</td>
<td>05-724</td>
<td>1:2000</td>
</tr>
<tr>
<td>MECP2 (W. Blot)</td>
<td>Millipore</td>
<td>07-013</td>
<td>1:1500</td>
</tr>
</tbody>
</table>
4. **Summary of Significant Findings**

The work presented in this thesis utilizes iPS cell reprogramming technology to provide novel insights into silent chromatin in pluripotent cells and neuronal characterization in the first isoform-specific iPS cell-based model of RTT.

In chapter 2 of this thesis, I demonstrated that retroviral reprogramming factors are silenced during the final stages of the reprogramming process. Retroviral silencing mechanisms associated with ES cells are conserved in full iPS cells since they establish ES cell-like silent epigenetic marks on the reprogramming transgenes. This retroviral silencing serves as a temporal landmark to delineate the timing of heterochromatin fibre reorganization and we demonstrate that chromocentre regions experience a wave of decompaction in full iPS cells. By applying 2i chemical treatment to partial iPS cells I took advantage of a controllable model to study epigenetic events associated with retroviral silencing and the structural features of chromocentre domains. Furthermore, I found that overexpression of the core pluripotency transcription factor Nanog is sufficient to induce chromocentre decompaction in partial iPS cells and that it cooperates with Sall1 to bind major satellite repeats within chromocentres. This cell culture model also provided a visualization platform to demonstrate for the first time that compact H3K9me3-enriched chromocentres are composed of 10 nm chromatin fibres.

In chapter 3 of this thesis, we utilized RTT patient fibroblasts harboring an 11 bp deletion in exon 1 of the MECP2 gene that specifically ablates MECP2e1 isoform to generate the first isoform-specific iPS cell-based disease model of RTT. I demonstrated that neuronal specification is not compromised in absence of MECP2e1 but the mutant neurons have slightly altered gene expression levels of several MECP2-associated genes. By expressing
MECP2e1 lentiviral vector and using a cell line lacking both isoforms, I demonstrate that MECP2 binds at most investigated loci. Finally, MECP2e1 rescue successfully recapitulates normal soma-size of differentiated neurons to WT levels in a cell autonomous fashion. Collectively, these data demonstrate that mutation of MECP2e1 is sufficient to induce subtle transcriptional changes of MECP2 target genes and abnormalities in regards to neuronal morphology.
5. **Discussion**

5.1. **Regulation of Silent Chromatin Domains in Pluripotent Cells**

Expression of Yamanaka factors during the reprogramming process guides the establishment of silent and active epigenetic modifications of somatic- and pluripotency-related genes, respectively. With particular focus on partial and full iPS cells, the findings in this thesis add to understandings of relationships between pluripotency, silent chromatin of retroviruses and the ultrastructure of constitutively silent chromocentre regions. In this section, I discuss the known epigenetic pathways of retroviral silencing in ES cells and the implications of my findings of decompaction of heterochromatin fibres on chromatin biology.

5.1.1. **Retroviruses are Silenced in Full iPS Cells**

Retroviral delivery of Yamanaka factors was instrumental in making iPS cell reprogramming a reality. Maintained expression of retroviral transgenes is required for establishment of endogenous transcriptional profile and upon acquisition of the full iPS cell state, retroviruses are silenced. I demonstrate in chapter 2 that silent pMX promoters in full iPS cells are epigenetically marked by high levels of histone H3K9me3, hypoacetylated histones and DNA methylation. These epigenetic marks are absent in retroviruses of partial iPS cells but are acquired upon 2i chemical treatment, which is known to promote establishment of full reprogramming. The importance of retroviral silencing during iPS cell reprogramming is two-fold. Firstly, retroviral silencing is a well-known feature of ES cells (discussed below) and thus the ability of iPS cells to silence retroviral vectors is a necessary characteristic of similarity between reprogrammed iPS and ES cells. Secondly, maintained high expression of pluripotency genes within the retroviruses may prevent establishment of a steady-state full iPS cells since even a twofold overexpression of Oct4 is sufficient to induce
differentiation in ES cells (Niwa et al, 2000). On the other hand, forced expression of all four reprogramming factors in established iPS cells was shown to prevent differentiation in an inducible system of iPS cell reprogramming (Brambrink et al, 2008), demonstrating that retroviral silencing is important in establishing the pluripotent state as well as enabling downstream differentiation-related applications of iPS cells. In both human and mouse system silencing of retroviral transgenes correlates with the quality of generated iPS cell lines, with those capable of silencing during the right phase of reprogramming conferring germline-competent and karyotypically intact iPS cell lines (Maherali et al, 2007; Okada & Yoneda, 2011; Okita et al, 2007; Ramos-Mejia et al, 2010; Wernig et al, 2007). iPS cell reprogramming lends itself to be an excellent model to study pluripotent-specific retroviral silencing mechanisms and our publication was the first pMX epigenetic characterization of full and partial iPS cells. Although, the direct role remains to be assessed my findings implicate a recently identified ES-cell specific targeting mechanism of retroviruses by Zfp809/Trim28 silencing complex (Wolf & Goff, 2007; Wolf & Goff, 2009) that is activated in the full iPS cell. The identity of this silencing mechanism was a long pursued missing link between known silencer elements within the retroviral backbone and the epigenetic marks that silence their expression. Its relevance in further understanding mechanisms of retroviral silencing in pluripotent cells is described in the following sections.

5.1.1.1. Retrovirus Silencing: A Historical Perspective

Although a useful marker of full reprogramming in iPS cells, retroviral silencing in pluripotent cells has historically restricted their use for long-term transgenic and gene-therapy applications (Asche et al, 1984; Challita & Kohn, 1994; Gautsch, 1980; Robertson et al, 1986). This block in expression is a result of a combination of cis-acting regulatory elements
within the genome of the classical moloney murine leukemia virus (MMLV, the LTR regions of MMLV are identical to the retroviral pMX vectors used in reprogramming) and the trans-acting factors in pluripotent cells that block retroviral transcription. MMLV silencing is mainly initiated at the viral replication-related sequence, close to the LTR promoter, known as the primer binding site (PBS) (Barklis et al, 1986; Loh et al, 1987). Realizations that mutations of the PBS (Grez et al, 1990; Petersen et al, 1991) and the LTR promoter itself (Tsukiyama et al, 1991) relieves silencing led to the development of a virus that can express in a subset of embryonic stem cells (Grez et al, 1990; Hawley et al, 1994; Laker et al, 1998). However, this murine stem cell virus (MSCV) is still silenced at most integration sites (Pannell et al, 2000) and with additional deletions in the LTR region our lab has generated the HSC1 virus which is capable of expressing in the highest percentage of infected ES cells (Osborne et al, 1999). Engineering a stably expressing retrovirus with additional deletions may prove difficult due to the overlap of remaining silencer regions with elements necessary for retroviral reverse transcription and integration. Instead, including insulators and strong internal promoters in the retroviral backbone, as we have done recently (Rival-Gervier et al, 2013), may lead to the design of viral constructs that can confer stable long-term expression in pluripotent cells and their differentiated progeny. However, even this, so far most optimal, retroviral design reveals integration events which are rapidly silenced suggesting that there are silencer elements remaining that are targeted by the ES cell. Understanding the epigenetic mechanisms and ES specific trans-factors that rapidly silence the pMX virus in ES cells has important consequences in understanding the related mechanisms in retroviral silencing in iPS cells.
5.1.1.2. Zfp809/Trim28 Target Exogenous pMX Retroviruses for Silencing

For a long time the identity of the pluripotency factors targeting the MMLV PBS in pluripotent cells remained a mystery. With a series of studies Stephen Goff and Daniel Wolf eventually determined the PBS-mediated silencing mechanism. MMLV retroviruses are targeted by pluripotent-specific zinc finger protein 809 (Zfp809) which directly binds the PBS element and through recruitment of Trim28, initiates a cascade of epigenetic modifications that ultimately establish a silent retroviral state (Wolf & Goff, 2007; Wolf & Goff, 2009) (Figure 5.1). The order in which the epigenetic marks are deposited is of particular interest for understanding how transcriptional regulation is modulated in general in ES cells. Furthermore, deciphering this complex regulatory mechanism will aid in identifying epigenetic pathways important for establishing and maintaining silent endogenous retroviruses (ERVs), which make up 10% of the mouse genome (Mouse Genome Sequencing et al, 2002), during embryonic development when extensive genome-wide epigenetic remodeling is taking place (Feng et al, 2010).
Partial and somatic cells maintain high expression of retroviruses at all integration sites. Upon transition to the full iPS cell state, retroviral primer binding site (PBS) is recognized by pluripotency-specific factor Zfp809 which recruits Trim28. As a result histone methyltransferases (Eset/G9a) establish H3K9me2/3 marks in Hp1-dependent manner on nucleosomes and signal to Dnmts for the secondary repressive mark of DNA methylation which, although abundant, is not required for retroviral silencing in ES cells.
5.1.1.3. Epigenetic Modifications in Exogenous and Endogenous Retroviral Silencing

Zfp809-dependent silencing of MMLV is relieved when the heterochromatin protein 1 (Hp1)-binding domain of Trim28 is mutated, demonstrating that Hp1 interaction is a required intermediary in the silencing cascade (Wolf et al, 2008). The functional importance of this interaction has not been addressed but the well-established roles of Hp1 proteins in H3K9 methyltransferase recruitment (Chin et al, 2007; Fuks, 2003) suggest that Trim28-associated HP1 acts as a bridge to induce histone methylation marks on nucleosomes of integrated retroviruses. This multifaceted combination of protein-protein interactions collectively reveals an intriguing possibility that histone methylation is the initial silencing epigenetic mark of newly integrated proviral DNA sequences. Although additional silent histone modifications, such as H4K20me3, deacetylated histones and DNA methylation, are abundant at transcriptionally silent retroviruses, this mechanism implies that these marks occur downstream of initially deposited H3K9me3 histone modification. This is supported by recent studies demonstrating that mouse ES cells lacking H3K9 methyltransferases Setdb1/Eset and Suv39h1/h2 have a higher level of expression of exogenous and endogenous retroviruses (ERVs). Class I ERVs, which are most similar to the MMLV virus, are derepressed in mouse ESCs lacking Eset (Matsui et al, 2010). This derepression is accompanied by reduced levels of DNA methylation at proviral promoters confirming that H3K9me3 is deposited prior to DNA methylation. Conversely, DNA methyltransferase (Dnmt1−/−, Dnmt3a−/−, Dnmt3b−/−) triple null (DNMT TKO) ES cells normally repress class I ERVs and have unchanged H3K9me3 and H4K20me3 levels (Dong et al, 2008; Matsui et al, 2010). Recruitment of Dnmtts to retroviruses in ES cells is a secondary repressive mark that is mediated in part by H3K9 dimethyltransferase G9a (Dong et al, 2008; Leung et al, 2011). When exogenous
retroviruses are considered, DNA methylation is a consequence of epigenetic silencing and de novo Dnmt-null ES cells efficiently silence retroviral transcription (Pannell et al, 2000).

The data presented in this thesis further strengthen these arguments but from the iPS cell perspective. Upon 2i-induced conversion, retroviruses in full iPS cells are stably silenced and marked by deacetylated histones and high levels of H3K9me3 with only low levels of DNA methylation at the retroviral promoter. Differentiated Dnmt3l-null cells of the testis and MeCP2-null neurons, however, have increased class I ERV (Matsui et al, 2010) and LINE1 elements (Muotri et al, 2011), respectively, suggesting that DNA methylation is required for repression of endogenous retroviruses in differentiated cell types.

Overall the mechanisms in pluripotent stem cells described here paint an intricate picture of epigenetic interactions that are taking place in establishing and maintaining the silent retroviral state. H3K9 methyltransferase Eset is implicated in being required for both of these phases of retroviral transcriptional regulation while DNA methylation is an accompanying mark that is not required for silencing. These models of retroviral regulation are also maintained in ERV elements in the mouse genome. The long-held concept of host defense, which holds that CpG methylation evolved in order to protect the organism from deleterious retrotransposition events (Walsh & Bestor, 1999), now contains another layer in the form of H3K9 methylation in pluripotent cells. Evolutionarily there would have been selective pressure for conservation of developmentally-regulated epigenetic pathways that maintain these parasitic events from occurring and thus retroviral silencing mechanisms may be one of the defining features of pluripotency. Interestingly, maintaining silent epigenetic marks on interspersed ERV elements has a role to play in regulation of the pluripotency-specific transcriptomes as well. Recently, depletion Trim28 was found to result in accumulation of
active epigenetic marks at ERV elements which can in turn act as enhancers to aberrantly activate transcription of proximal genes (Maksakova et al, 2013; Quenneville et al, 2012; Rowe et al, 2013). Trim28 does not have a DNA binding domain and the sequence specificity of its targets is mostly mediated by Zfp proteins (reviewed in (Iyengar & Farnham, 2011)). Although Zfp809 specifically targets Trim28 to pMX retroviruses it would be of interest to determine whether additional Zfp proteins target retroviruses such as MSCV and HSC1 lacking the PBS sequence. However in contrast to pMX, expression of these more developed retroviruses is actually maintained in a high number of integration sites and the apparent silencing in a subset of ES cells may be a consequence of epigenetic landscapes of the integration sites.

As I was starting the work reported in chapter 2 of this thesis, pMX delivery of Yamanaka factors was the most widely used method of somatic cell reprogramming. We anticipated that by dissecting the reprogramming process into different phases, we would be able to study pluripotent-specific modes of retroviral silencing. Detailed temporal studies have not been performed in iPS cells and the 2i conversion platform is an attractive tool to use to dissect the precise order of epigenetic events that establish silencing in pluripotent cells. Although histone methylation seems to be the driving force of retroviral silencing, additional pathways independent of DNA methylation may have important roles to play in the epigenetic sequence of events. For instance deacetylation and ATP-dependent chromatin remodeling during this process has not been assessed in detail and some outstanding questions with approaches to address them are discussed as future direction plans in section 5.3.1.
5.1.2. **Heterochromatin Reorganization in Full iPS Cells**

Prior to our publication from chapter 2, no *in situ* investigations of chromatin fibre structure changes during reprogramming have been conducted. By defining the partial and full iPS cell state based on retroviral silencing and activation of the pluripotency transcriptional profile, I demonstrate that full reprogramming results in striking decompaction of H3K9-enriched chromocentres with 10 nm chromatin fibre density indistinguishable from the surrounding euchromatin. Partial iPS cells maintain compact heterochromatin fibre structure similar to parental fibroblast cells and they only experience heterochromatin fibre decompaction upon transition to a 2i-mediated full iPS cell state. Complete DNA methylation of major satellite repeats, ubiquitous genetic component of chromocentres, is maintained in full iPS cells demonstrating that DNA demethylation is not required for the observed heterochromatin fibre reorganization.

5.1.2.1. **Sall1 and Nanog Control Heterochromatin Decompaction in iPS Cells**

Dispersed chromatin fibre structure is a feature commonly associated with pluripotent cells, both in ES cell culture models (Efroni et al, 2008) and in the ICM cells of the blastocyst (Ahmed et al, 2010). When one captures the chromatin structure using ESI, where chromatin fibres can be visualized in high contrast, it displays a homogenous distribution of chromatin without distinct blocks of tightly packed domains characteristic of differentiated cell types. FRAP experiments further characterize this general open conformation by the abundance of “loosely” bound core histones and the associated Hp1 proteins in ES cells suggesting that histones in gneraly exist in a hyperdynamic state within pluripotent nuclei (Bhattacharya et al, 2009; Meshorer et al, 2006). Our findings provide further evidence that heterochromatin domains are in an “open” configuration only upon transition to full iPS cells implicating
factors that are activated during this transition as necessary components for the decompaction to occur. This is corroborated by the findings in mouse blastocysts lacking Oct4, a core pluripotency-related TF, where cells of the ICM lose their pluripotency and the homogeneous chromatin fibre structure reverts to a more differentiated cell-like structure with regions of compact heterochromatin (Ahmed et al, 2010). Clearly, Oct4 is not the necessary factor in the iPS cell system since partial iPS cells are expressing high levels of Oct4 from the retroviral transgenes. Instead, we focused on Sall1 and Nanog due to known Sall1 association with chromocentre domains (Kiefer, 2003; Netzer et al, 2006; Sakaki-Yumoto, 2006; Yamashita et al, 2007) and our initial findings of inverse relationship between Nanog variegation and chromocentre compaction levels. We analyzed Nanog−/− and Sall1−/− ES cells and found abundant chromocentre compaction. Overexpression of Nanog was sufficient to induce heterochromatin decompaction in partial iPS cells and both of these TFs are found to cooperatively bind major satellite repeats. Collectively, these findings demonstrate that chromatin assumes a homogeneous arrangement without distinct ultrastructural heterochromatin fibre domains in full iPS cells and this structure may be directly maintained by pluripotency TFs Nanog and Sall1.

5.1.2.2. 10 nm Fibre Configuration of Heterochromatin Domains

In our publication we used correlative LM/ESI in order to distinguish between “open”, euchromatin, and “closed”, H3K9me3-enriched, chromatin domains. Prevailing models of chromatin organization maintain that the 30 nm fibre is one of the core structural components of mammalian chromatin (Belmont & Bruce, 1994; Finch & Klug, 1976; Gilbert et al, 2004; Naughton et al, 2010). However, findings of the 30 nm fibre in situ has been largely restricted to non-mammalian cells and support is mounting that it may not exist at all in mammalian
nuclei (Fussner et al, 2011a; Maeshima et al, 2010a; Maeshima et al, 2010b; Nishino et al, 2012). If one is to speculate about its existence, however, 30 nm fibres would be expected to exist within the “closed” chromocentre domains. Our findings provide some of the first correlative ESI images of H3K9me3 foci and we find no evidence of 30 nm fibres in chromocentres of ES, iPS or fibroblast cells. Although extensive compaction in fibroblast cells does exist, the fibres are predominantly 10 nm and our follow-up publication demonstrates using 3-dimensional electron tomography technique that 30 nm fibres are completely absent in mouse embryonic fibroblasts (Fussner et al, 2012).

5.1.2.3. DAPI Linescans as a Tool for Screening High Quality iPS Cell Lines

With the identification of defined genetic factors necessary for reprogramming to occur, iPS cell technology has emerged as the more efficient method to isolate high numbers of reprogrammed iPS cell lines. However, as is expected reprogramming yields a heterogeneous mixture of iPS cell lines with variable similarities to ES cells. In order to obtain the highest quality iPS cell lines it is important to determine all the molecular blockades of reaching the fully reprogrammed iPS cell state. It is clear that epigenetic reprogramming of pluripotency-related loci is required for the activation of genes necessary to establish pluripotency with our study and others (Gaspar-Maia et al, 2010; Mattout et al, 2011) providing evidence that remodeling of constitutive H3K9me3 also correlates with the pluripotent state. Given that iPS cell reprogramming requires selection of iPS cell lines with which to perform subsequent experiments, we propose that characterization of the heterochromatin compaction state can serve as a valuable tool to prioritize mouse iPS cell lines of the highest quality. Although in this study, we use correlative LM/ESI, a highly specialized technique, the surrogate assay of DAPI linescan analysis serves as a robust
approach to assess levels of heterochromatin compaction in a larger number of mouse iPS cell lines.

5.2. MECP2e1 Function in iPS Cell Model of RTT

5.2.1. MECP2e1 Mutation does not Alter Neuronal Identities

The ability to study MECP2 function in live human neurons from RTT patients has become a reality with the development of iPS cell reprogramming. By utilizing iPS cells from a patient containing a mutation in only MECP2e1, with an intact MECP2e2, we set out to investigate the role of this more relevant isoform to RTT. By using single cell Fluidigm arrays, I found that differentiation of MECPe1 mutant iPS cells yielded comparable mixture of GABAergic and glutamatergic neurons without a difference in the cortical neuronal identities produced. However, the overall number of cells expressing mature neuronal markers DCX, NCAM and MAP2 was slightly increased when mutant neurons were supplemented with lentiviral MECP2e1. This is consistent with some of the overall findings of current iPS cell models implicating a neuronal maturation defect where expression of mature neuronal marker of Tuj1 was found to be decreased in absence of both MECP2 isoforms (Kim et al, 2011).

Both mouse (Farra et al, 2012) and human (Ananiev et al, 2011; Cheung et al, 2011; Kim et al, 2011; Marchetto et al, 2010) iPS cell-derived neurons recapitulate neuronal phenotypes associated with mouse models of the disease and known patient phenotypes. One of the defining features of RTT neurons is their decreased morphological complexity displayed by decreased neuronal soma size, reduction in dendritic branching and spine densities. This is further exemplified by the myriad of electrophysiological phenotypes and expression and immunolabeling analysis showing a slight decrease in expression of a few mature marker genes in RTT neurons (Kim et al, 2011). Collectively, these findings suggest
that RTT neurons have a defect in neuronal maturation exemplified by both neuronal function and morphology.

5.2.1.1. X-Inactivation in Human RTT iPS Cells

A point of slight controversy in the RTT field has been the issue of X-chromosome inactivation status in human iPS cells (Cheung et al, 2012). This is of particular interest in RTT as the presence of the WT allele on one of the X-chromosomes could confound approaches to identify mutant MECP2-related phenotypes. Several methods are available to assess the X-inactivation status but in general cDNA sequencing of MECP2 with primers flanking the mutation combined with the AR assay accurately assess the X-inactivation status. Here we add evidence, in addition to our previous report (Cheung et al, 2011), that inactive X-chromosome is not reactivated in human iPS cells upon reprogramming and differentiated neurons do not experience X erosion. This allowed us to prioritize cell lines for phenotyping experiments that are solely expressing the mutant MECP2e1-containing chromosome.

5.2.2. MECP2e1 Function as Transcriptional Regulator

We demonstrate that reduced levels of expression of MECP2-target genes found in MECP2-null mutant neurons are maintained when only the MECP2e1 isoform is affected. Furthermore, MECP2e1 lentiviral vectors specifically express in differentiated neurons, correctly bind to promoters of target genes and result in restoration of normal gene expression of several target genes. This effect on target gene expression is concomitant with the lentiviral-mediated soma size rescue in RTTe1 neurons. Overall, the data presented here solidify the expectations that MECP2e1 is the deterministic isoform in RTT development and that the human iPS cell-based models of RTT can be useful in determining the chromatin
positioning of MECP2. Although common MECP2 targets have been difficult to ascertain in the past (Zachariah & Rastegar, 2012), MECP2 function is increasingly being defined as a global regulator of gene regulation that tracks methyl CpGs and responds to activity dependent stimulation to enable transcription of activity-dependent genes (Cohen et al, 2011; Goffin et al, 2012; Lyst et al, 2013; Skene et al, 2010). Although, isoform specific distribution within neurons has not been assessed, it is suspected that enrichment of MECP2e1 should be found at most investigated loci if it is also globally distributed in human neurons. We provide the first evidence that MECP2e1 dysfunction in human neurons can lead to transcriptional alterations of some genes previously shown to associate with Mecp2 in mouse models of RTT.

It is of worthy note that as this thesis is being compiled a publication by the Jaenisch group reported a novel human ES cell-based model of RTT using transcription activator-like effector nuclease (TALEN) approach to target exon 3 of MECP2 (Li et al, 2013). These mutant neurons exhibited morphological and electrophysiological phenotypes reminiscent of those with true disease-related mutations. However, this study has provided a lot of food for thought as it demonstrates, using a novel gene expression analysis technique which takes total RNA content into consideration when performing normalization calculations (Loven et al, 2012), that MECP2 acts as a global activator of gene expression in neurons. MECP2 action as a transcriptional activator has been raised by numerous studies in the past (Chahrour et al, 2008; Loven et al, 2012; Tudor et al, 2002; Yasui et al, 2007), but they did not appreciate the dramatic abundance of genes that are affected in mutant neurons. The discrepancy may reside in the fact that these previous reports were done in the mouse model which is thought to be an underrepresentation of the human RTT condition. Global analysis by the Jaenisch group
reveals that mutant MECP2 neurons had a 50% decrease of total RNA, with 60% of genes being downregulated and only 7% upregulated. Importantly, genes involved in synaptogenesis and neuronal development were downregulated while astrocyte-specific genes were largely unchanged, showing the expected cell-type specificity for MECP2 function. The fact that MECP2 is involved in activating gene expression is further supported by the findings that the genes that are the most highly expressed are the ones that experience the highest level of reduction in mutant neurons. The authors correlated 5-hydroxymethylcytosine levels, which are more commonly found at actively transcribed genes, with the downregulated targets of MECP2 in contrast to 5-methylcytosine marked genes which were less frequently affected in the mutant neurons. These findings require special attention with respect to the data presented in chapter 3 of this thesis. The qRT-PCR and Fluidigm approach that I employed in chapter 3 to investigate MECP2e1 effects on gene expression are consistent with MECP2e1 acting as an activator of gene expression. The genes that I found to be differentially expressed were all upregulated in MECP2e1-rescued neurons as well as the Δ3-4 WT compared to their mutant counterparts. Four (MAP2, DCX, NCAM, CUX1) out of six genes that were different in the single-cell Fluidigm experiments were expressed in a higher number of rescued cells compared to mutant neurons. Fluidigm approach used in this thesis measures the absolute number of cells expressing the transcripts and is thus immune from the normalization techniques that may normally mask gene expression differences. Importantly, the study by Jaenisch group is the first to assess gene expression analysis of mature neuronal and cortical layer markers, many of which overlap with the gene set I analyzed in Fluidigm. Consistent with my findings in RTTe1 neurons, the corical layer cellular specification during neuronal differentiation is not changed in neurons with artificially generated MECP2 mutations. The
authors went on to demonstrate that global translation levels in the mutant neurons were reduced due to the diminished activity of the AKT/mTOR pathway, which is positively regulated by the MECP2 target BDNF. Treatment with BDNF and suppression of PTEN, a negative regulator of AKT/mTOR, increased translation levels and rescued the soma size and dendritic branching phenotypes. This comprehensive study really addressed many of the familiar questions of RTT modeling and importantly it has almost single handedly redefined the function of MECP2 as a transcriptional activator. However, as is always the case there are many questions yet to be addressed. Primarily, lack of MECP2 binding experiments in this novel human model of RTT is somewhat puzzling given the extensive data presented in this single study. How is the MECP2 protein distributed in human neurons? Is there a high degree of correlation between MECP2 chromatin association between human and mouse models? At the transcriptionally activated targets of MECP2, what are the co-activators and at the transcriptionally repressed, what are the co-repressors? Are the phenotypes investigated and shown to be ameliorated by pTEN knockdown reproducible in the neurons harboring RTT-causing MECP2 mutations? And most relevant to the work presented in this thesis, how do the findings relate to isoform-specific functions of MECP2? The authors made no mention of MECP2 isoforms nor did they demonstrate that the MECP2 expressed in their system is MECP2e1, the highly expressed neuronal isoform. Indeed, in all the studies performed to date, iPS cells were generated from RTT patients with mutations affecting both MECP2 isoforms and our e1 model is the first time isoform-specific dysfunctions are explored for any iPS cell-based disease model.
5.2.3. Comment on iPS Cell Technology for Disease Modeling

Development of iPS cell technology has led to tremendous amount of cell-type specific models of various diseases. However, some concerns are yet to be addressed for widespread modeling of human diseases in vitro. In particular, each of the studies so far has reported a very limited number of established lines and for disease modeling to be reliably explored, more control and patient lines need to be established in parallel using highly reproducible reprogramming methods. A wide range of innovative techniques has resulted in successful reprogramming using pMX-based retroviral vectors, lentiviruses, non-integrating adenoviruses, episomal vectors, plasmid and piggyBac transposon systems. Although integration-free reprogramming methods are preferable, our group has previously suggested that the use of retroviral transgenes provides a beacon of the completion of reprogramming process and may be useful in cell lines not used for clinical applications (Hotta & Ellis, 2008). This feature has been exploited in numerous iPS cell studies investigating the dynamics of the reprogramming process (Stadtfeld et al, 2008b; Takahashi & Yamanaka, 2006). Even with non-integrating reprogramming technologies, significant heterogeneity between iPS cell lines may prove to be the biggest obstacle to reliable evaluation of disease phenotypes in vitro.

It is yet to be determined, in multiple disease model settings, the range of the normal phenotype of a particular functional assay. Only when the full range has been determined and reproducible differentiation protocols have been established will we be able to model human diseases in culture with a high level of confidence. Overcoming this obstacle may involve determining the inherent variability in the starting population of undifferentiated ES and iPS cells. It is unknown whether these variabilities will persist to the differentiated cell state. Although initial reports were highly encouraging as to the similarities between newly derived
iPS cells and existing ESC lines, newer studies support thorough molecular characterization of clonal iPS cell lines and reveal that significant differences between iPS and ESCs remain at both the transcriptional (Chin et al, 2009) and epigenetic levels (Doi et al, 2009). Inappropriate establishment of epigenetic marks in the undifferentiated iPS cells will undoubtedly yield heterogeneous differentiated cells for subsequent applications. In fact, a brief comparison of even human ES cell lines demonstrated that differentiation preferences are highly inconsistent among different ES cell lines resulting in variable levels of lineage-specific gene activation (Osafune et al, 2008).

During these early phases of studying iPS cells, it is important to consider all possible obstacles to the safe and reliable use of iPS cells in these downstream applications (Belmonte et al, 2009). Leading stem cell researchers remain cautious about possible false promises regarding the application of induced pluripotent stem cells (Izpisua Belmonte et al, 2009; Soldner & Jaenisch, 2012). The quality of the starting iPS cell lines for in vitro differentiation experiments may not need to be pluripotent to the strictest criteria, as long as the differentiation of the cell type of interest could be accomplished accurately and efficiently. For this purpose, if one is interested in studying the cell types of the mesoderm lineage, the capacity of the iPS cell line to give rise to ectoderm/endoderm lineage may not be relevant. Even in these restricted directed differentiation experiments, appropriate epigenetic circuitry needs to be established for the lineage specific genes to be re-expressed upon receiving differentiation signals. For any clinical applications of iPS cells, it is essential to ensure that the differentiated cells derived from iPSCs have minimal risks of inducing cancer or teratoma development in patients. As we transition into large scale efforts to make patient-specific iPS cell lines and model human diseases in vitro, proper epigenetic remodeling may remain at the
forefront of stem cell biology and developing methods to stably establish ESC-like epigenetic circuitry during the reprogramming process will be of high priority.

5.3. **Future Directions**

5.3.1. **What are the initiating epigenetic events of pMX silencing?**

Traditionally retroviral silencing mechanisms in ES cells have been examined by performing infections followed by monitoring of retroviral expression by flow cytometry and chromatin immunoprecipitations to determine the epigenetic marks associated with the retroviral transgene. Although such approaches have been valuable in determining the silent epigenetic marks associated with retroviral transgenes the epigenetic events that initiate the silencing cascade have been difficult to pinpoint. This is due to two main limitations of performing bulk infections. Firstly, the virus integrates into a distinct integration site within each cell, some cells receive multiple integrations while some cells within the population are not infected at all. Secondly, retroviruses enter the nucleus during mitosis followed by integration into the host genome and this occurs at different time points in asynchronous population of cells. Given that retroviral silencing is established immediately after infection, it is easy to see how performing bulk infections would have a shortcoming in capturing these early events. Although a recent study has somewhat addressed these issues by monitoring retroviral expression in G9a$^{-/-}$, Suv39$^{-/-}$, Eset$^{-/-}$ and Dnmt3a/3b$^{-/-}$ ESCs over a timecourse to evaluate histone methyltransferase roles in establishment and maintenance phase of retroviral silencing (Leung et al, 2011) its analysis was limited to only that set of epigenetic modifying enzymes. Furthermore, even though G9a was suggested to regulate the initiation phase of silencing, G9a$^{-/-}$ ES cells still had a significant portion (~70%) cells that had silenced the retrovirus, implicating additional epigenetic pathways.
To address these issues, I propose that the partial EOS3F-24 iPS cell line combined with the 2i treatment can be utilized to screen a large set of epigenetic enzymes in their role for the initiation phase of retroviral silencing. We have obtained an endoribonuclease-prepared siRNA (esiRNA) library of 1008 epigenetic enzymes and chromatin remodeling proteins (Fazzio et al, 2008) which can be used to screen pathways involved in silencing. To perform this experiment a round of infection using pMX-RFP virus can be performed followed by expansion of infected cells and identification of subclones with single integration sites. The infected cells can then be grown in 96-well plates and transfected with the esiRNA library where each well would have a knockdown of a single epigenetic modifier. Cells would then be grown in 2i to initiate retroviral silencing and individual wells that would fail to turn off the red virus would be counted as hits in the screen (Figure 5.2).

![Diagram of 2i-Mediated conversion system for finding initiating epigenetic events in retroviral silencing.](image)

Figure 5.2. 2i-Mediated conversion system for finding initiating epigenetic events in retroviral silencing.

5.3.2. What are the pluripotent-specific modes of heterochromatin decompaction?

Work presented in this thesis demonstrates that ES and full iPS cells maintain a decompacted heterochromatin fibre architecture that is indistinguishable from the surrounding euchromatin. The binding of Nanog and Sall1 to major satellite repeats within chromocentres
reveals novel functions of these pluripotency TFs in chromatin ultrastructure organization. Due to a wide range of Nanog- and Sall1- interacting proteins, it would be of interest to determine the identity of co-occupying epigenetic modifiers that participate in guiding heterochromatin decompaction observed in pluripotent cells. To this end, future experiments can assess the identity of protein complexes that are found at major satellite, and other repetitive regions contained within chromocentres, by performing sequential ChIP experiments. My finding that major satellite repeats remain methylated at CpG dinucleotides suggest that DNA demethylation is not required for heterochromatin decompaction but additional epigenetic modifications have not been assessed. For example, it has been demonstrated that global levels of H3K9me3 (Efroni et al, 2008) are elevated upon ES cell differentiation, therefore, one would suspect that H3K9 methyltransferases or demethylases may associate with Nanog and Sall1 at major satellite repeats. As an accompanying study, KO ES cell lines for various epigenetic modifying enzymes can be employed to reveal the epigenetic pathways that are required to condense the heterochromatin structure upon differentiation. Our publication demonstrates that heterochromatin compaction appears within 72 hours of LIF withdrawal from ES cell culture and if, for instance, histone methyltransferase Eset is involved in this process, differentiation of Eset−/− ES cells would not be accompanied by appearance of compact chromocentres. This approach can be expanded to use the esiRNA library (described in the previous section) during differentiation to determine the required pathways for compaction to occur.

Alternatively the heterochromatin fibre condensation found in the Nanog and Sall1 KO ES cells could be a downstream effect of misregulation of Sall1 or Nanog target genes. If so, that begs the question of what role do these TFs serve in the H3K9me3-enriched territory.
Given that both genes are known repressors of transcription (Karantzali et al, 2011; Liang et al, 2008; Liu et al, 2008; Loh et al, 2006) it would be of interest to determine whether their silenced genomic targets co-occupy the chromocentre regions by DNA-FISH analysis. If genomic targets that Nanog occupies are located within chromocentre domains, this could lead to the speculation that Sall1 serves as a protein anchor at transcriptionally silent H3K9me3 domains which Nanog could recognize and co-occupy with genes undergoing silencing in a histone H3K9 methylation-dependent manner.

5.3.3. MECP2e1 Neuronal Cell-Type Specific Function?

MECP2 associates with chromatin throughout the genome and in the mouse hypothalamus it regulates transcription of over 2000 genes as both a transcriptional activator and repressor (Chahrour et al, 2008). Activity-dependent phosphorylation of Mecp2 is instrumental in guiding the neuronal differentiation process through its transcriptional regulation activity (Cohen et al, 2011). If key phosphorylation sites such as S421 or T308 are mutated, MECP2 association with its transcriptional co-repressors is abolished, and genes with roles in promoting synapse formation and dendritic branching are abnormally expressed. Gene regulation is central to MECP2 function and targeted over-expression of its deregulated activity-dependent target \( Bdnf \) results in rescue of MECP2-dependent RTT neuronal phenotypes (Kline et al, 2010; Li et al, 2013). Despite extensive evidence of Mecp2 gene regulation in mouse models of RTT, the extent to which MECP2 may play similar roles in the human system are only currently being elucidated. For instance, recent report demonstrated that TALEN-induced \( MECP2 \) mutations lead to complete absence of activation of a few activity-dependent genes (Li et al, 2013). Consistent with the findings in mouse models I
demonstrate in this thesis that MECP2 is bound at most investigated loci, but I did not investigate the binding dynamics or gene expression changes upon neuronal KCl stimulation of MECP2e1-mutant neurons.

Due to the heterogeneous nature of all differentiation protocols, performing single-cell analysis is the ideal method to correlate gene expression of MECP2 and its target genes in human iPS-derived neuronal cultures. Through optimal design of primers, relationships between MECP2 and its speculated gene targets within individual cells can be assessed. To do this, RTTe1 neurons and MECP2-null neurons can be infected with the MECP2e1 lentiviral construct and induced to differentiate. The panel of genes used for the Fluidigm array would include most of the genes reported in chapter 3, with an addition of high priority MECP2 target genes, including activity-dependent genes. MECP2e1 lentiviral specific primers would be included to identify the cells that are expressing the rescue construct. By identifying appropriate neuronal subtypes that are expressing the construct and comparing expression of target genes to cells not expressing the construct, MECP2e1-regulated genes can be identified. As a more global assessment, single-cell RNA-Sequencing experiments would allow an unbiased approach to investigating MECP2e1-dependent gene expression changes. If the recent findings (Li et al, 2013) are validated, I suspect that MECP2 rescue construct-expressing cells would have a global increase in gene expression. Combined with KCl neuronal stimulation, such approaches can determine resting and activated neuronal gene expression changes in presence and absence of MECP2e1.

Given that MeCP2e1 isoform is the more dominant isoform in neuronal cell types, representing about 90% of the overall MeCP2 protein in mouse tissues, it is conceivable that the phenotypes associated with patients and cellular models of MECP2e1 mutant RTT are due
to the overall reduction of MECP2 rather than MECP2e1-specific dysfunctions. To address this issue, the MECP2 constructs generated in our lab (Rastegar et al, 2009) can be utilized to investigate the necessity of either isoform in iPSC-based models RTT. One of the cell lines presented in this thesis, line Δ3-4#20, completely lacks MECP2 protein and as such is an ideal platform for transduction experiments with the isoform-specific lentiviral rescue constructs. As such, experiments described above can be conducted in presence of either MECP2e1 or MECP2e2-expressing lentiviral constructs in Δ3-4#20 neurons. If MECP2e1 is the required isoform, abnormal gene expression changes would only be rescued in the presence of MECP2e1 and would remain unaffected in MECP2e2-expressing cells.

Figure 5.3. Assessing MECP2e1 roles in transcriptional regulation.
5.4. **Concluding Remarks**

The work in this thesis examined mouse iPS cell models of chromatin organization and MECP2e1 roles in human RTT iPS cell-derived neurons. I demonstrate through dissection of the reprogramming process into partial and full iPS cell states that chromatin reorganization events can be studied to reveal interesting links between pluripotency, retroviral silencing and heterochromatin decompaction. In the future, examination of epigenetic events underlying transitions from active to silent and compact to decompacted states in retroviral expression and chromocentre organization, respectively, will provide further insight into the multilayered epigenetic mechanisms involved in regulation of silent chromatin domains in pluripotent cells.

MECP2 is an alternatively spliced global modulator of gene expression in neuronal tissues. Here I show that a *MECP2e1*-specific mutation does not affect neuronal identity specification in iPS cell-derived neuronal model of RTT. However, *MECP2e1* mutation does alter its chromatin association, affecting gene expression of some target genes and induces neuronal soma size defects associated with global loss of MECP2 protein. Future experiments are described to assess MECP2e1 roles in regulating global gene expression changes in resting and activated human neurons. These approaches may uncover neuronal pathways that are perturbed in RTT neurons and may provide additional targets for future developments of RTT treatments.
6. References


Ananiev G, Williams EC, Li H, Chang Q (2011) Isogenic Pairs of Wild Type and Mutant Induced Pluripotent Stem Cell (iPSC) Lines from Rett Syndrome Patients as In Vitro Disease Model. *PloS one* 6: e25255


pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. *Science* **321**: 1218-1221


Feng B, Ng J, Heng J, Ng H (2009) Molecules that promote or enhance reprogramming of somatic cells to induced pluripotent stem cells. *Cell Stem Cell* **4:** 301-312


Fuks F (2003) The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Research* **31:** 2305-2312


Gurdon JB (1962a) Adult frogs derived from the nuclei of single somatic cells. Dev Biol 4: 256-273


Show Global Epigenetic Remodeling and Widespread Tissue Contribution. *Cell Stem Cell* **1**: 55-70


Okada M, Yoneda Y (2011) The timing of retroviral silencing correlates with the quality of induced pluripotent stem cell lines. *Biochim Biophys Acta* **1810**: 226-235


Sakaki-Yumoto M (2006) The murine homolog of SALL4, a causative gene in Okihiro syndrome, is essential for embryonic stem cell proliferation, and cooperates with Sall1 in anorectal, heart, brain and kidney development. Development 133: 3005-3013


Teich NM, Weiss RA, Martin GR, Lowy DR (1977) Virus infection of murine teratocarcinoma stem cell lines. *Cell* **12**: 973-982


