Title
Epigenetic Regulation of Transcription and Cellular Potency during Mammalian Development

Permalink
https://escholarship.org/uc/item/44q7v8mn

Author
Sachs, Michael

Publication Date
2014

Peer reviewed|Thesis/dissertation
Epigenetic Regulation of Transcription and Cellular Potency during Mammalian Development

by

Michael Sachs

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the
Copyright 2013
by
Michael Sachs
This dissertation would not have been possible without Karen and Gerald, who have provided support and encouragement beyond the limits of expectation or rational parenting.

-thank you
ACKNOWLEDGEMENTS

I am severely indebted to the many persons who have supported, guided, and taught me throughout the course of my doctorate. Without their help, I would not have reached where I am now. First and foremost, my mentor Miguel Ramalho-Santos has been a tremendous guide and role model. His trust in me to explore novel avenues of research and strike out on my own is unparalleled and has allowed me to grow as a scientist and a person. I will continuously strive to emulate his manner of cautious optimism throughout my professional life.

My committee members, Robert Bleloch and Matthias Hebrok, are pillars of scientific knowledge that have aided me in scientific development and beyond the bench. From them I have learned numerous lessons in scientific planning, project management, and presentation that I will carry forward.

As nothing in life happens in a vacuum, I would like to thank the members of the Ramalho-Santos lab throughout the years. My first mentor in the lab was Marica Grskovic who was gracious enough to take on an overreaching, overbearing rotation student and accept me as part of the lab. Her energy and critical questions have served as a template for my approach to research. My time spent working with Rachel Nitta was invaluable and I am thankful for her hard work and tolerance as together we trudged through the cloudiest part of my research. I will always fondly remember the many outlandish scientific coffee sessions with Kathryn Blaschke and Kevin Ebata who have always balanced my craziest of ideas and ensured I look further afield when I wasn’t crazy enough. I am proud to share an authorship with them. The last couple of years working with Marcela Guzman have been an enjoyable collaboration and it is with mixed feelings I approach its conclusion. The whole of team Chd1 has been a great help in handling this pesky molecule; thank you Fong Ming Koh, Chih-Jen (Lance) Lin, and Priscilla Wong. In
addition I would like to thank Fanny Polesso for her boundless energy, and Laure Blouin for her ability to make order(s) from chaos.

Outside of the lab, the LegoScope has been a wonderfully rewarding endeavor that has contributed greatly to my ability to prioritize, manage projects, and meet deadlines in a manner scientific experimentation could not. For this I need to acknowledge Harrison Liu, Reid Williams, and our mentors; Terri O’Brien, Scott Paterson, and Judy Lee. In addition, my classmates and friends have taught me numerous lessons over the years from humility to perseverance, and everything in between. Without them, my grasp on sanity would be even more tenuous than it is now. Thank you Melissa Lezameta, Fadi Malak, JP Morgan, Robin Smith, Robert Judson, Karen Ring, Brandon Mahne, Julie VanderMeer, Rachel Nitta, Lucho Fuentealba, Ashley Mahne, Kasia Skrzypczynska, Eric Woo, Justin Chen, and The Killer Space Robots.

Most importantly, family has provided the encouragement and support that has led to this juncture and will allow me to pursue my goals into the future. My brother, Mathew has served as a sounding board and has always been willing to discuss and evaluate regardless of how eccentric the idea- thank you. Trips back home to see Lauren, Maya, and Brook have been invaluable in helping me to remember what is truly important in life. To my parents, Karen and Gerry, you have always encouraged pursuit of my goals wherever they led. I am forever grateful for the sense of curiosity you inspired in me to explore the world and the sense of wonder from this experiment called life. Thank you.
CONTRIBUTIONS

Chapter 2 is the result of a collaboration between multiple scientists across several years. While many experiments have not been mentioned here, they have all contributed to shaping and guiding the direction of this project. The people that have contributed are Rachel Nitta, Marcela Guzman, Fong Ming Koh, Chih-Jen (Lance) Lin, Priscilla Wong, Jun Song, Courtney Onodera, Aaron Diaz, Zhihai Ma, Bradley Stohr, Morgan Royce-Tolland, Nevan Krogan, John Stamatoyannopoulos, and Miguel Ramalho-Santos.

Chapter 3 is adapted from the publication Sachs et al., 2013, which was published in Cell Reports on June 27, 2013 (Cell Reports, 2013, 3(6): 1777-1784). Courtney Onodera was second author and was responsible for the vast majority of the bioinformatics and analysis of the ChIP-seq data with supervision from Jun Song. Kathryn Blaschke and Kevin T. Ebata assisted in isolation of PGCs. Kevin T. Ebata was also responsible for all FACS during PGC isolation.
Epigenetic Regulation
of Transcription and Cellular Potency
during Mammalian Development

DNA provides the blueprint for all cellular functions through the process of gene transcription, while epigenetics contributes to regulating which genes are transcribed and to what extent. Both this genetic and epigenetic information must be passed from one cell to its daughters to preserve cellular identity. The accessibility of DNA to the transcriptional machinery is regulated by a multitude of chromatin proteins and complexes. Chd1 is a highly conserved enzyme chromatin remodeler essential to mammalian development. Despite its strong association with transcriptional activity, its precise role remains unclear. In chapter 2 of this dissertation I present evidence showing that Chd1 is dispensable for in vitro pluripotency, but is required for optimal expansion of mouse ES cells, a defect that likely results from a wide-spread reduction of RNA polymerase activity, which is known to limit cell cycle progression. This is a novel chromatin level regulation of cell proliferation.

In addition to regulating transcriptional initiation and elongation, chromatin packages DNA into domains that are either available for transcriptional activation or not. Bivalent domains are a unique chromatin state associated with pluripotency that maintain gene promoters as poised, in an available but inactive state, for rapid activation upon lineage specification. In chapter 3, I describe the observation of bivalent domains in the unipotent mammalian germline during embryogenesis. These domains are enriched for developmental regulators of all germ layers in a manner remarkably similar to ES cells. The maintenance of the somatic program in a poised state in the germline throughout development raises the intriguing possibility this epigenetic information imparts a path for transgenerational epigenetic inheritance.
# TABLE OF CONTENTS

**CONTRIBUTIONS** vi

**ABSTRACT** vii

**LIST OF TABLES** ix

**LIST OF FIGURES** x

**CHAPTER 1:** Introduction 1

**CHAPTER 2:** Chd1 Maintains Transcription by Regulating Chromatin Accessibility to RNA Polymerase 28

**CHAPTER 3:** Bivalent Chromatin Marks Developmental Regulatory Genes in the Mouse Embryonic Germline in Vivo 56

**CHAPTER 4:** Conclusion and Discussion 87

**APPENDIX I:** Primers 98

**APPENDIX II:** Low Cell ChIP Protocol 103

**APPENDIX III:** qRT-PCR Protocol: From Cells to Excel 108
LIST OF TABLES

CHAPTER 2: CHD1 MAINTAINS TRANSCRIPTION BY REGULATING CHROMATIN ACCESSIBILITY TO RNA POLYMERASE

**Table 2.S1:** Mouse ES cell lines generated in this study

**Table 2.S2:** Genes with reduced H3K4me3 in Chd1 KO ES cells

CHAPTER 3: BIVALENT CHROMATIN MARKS DEVELOPMENTAL REGULATORY GENES IN THE MOUSE EMBRYONIC GERMLINE IN VIVO

**Table 3.S1:** ChIP-Seq library statistics.

**Table 3.S2:** Bivalent domains in PGCs and ESCs

**Table 3.S3:** CAGE tag expression by tissue type in PGC orphan bivalent domains

APPENDICES

**Table A.1:** ChIP-qPCR primers

**Table A.2:** qRT-PCR primers
# List of Figures

## Chapter 1: Introduction

| Figure 1.1 | Post translation histone modifications associated with transcription | 3 |
| Figure 1.2 | Nucleosome dynamics during transcriptional elongation | 4 |
| Figure 1.3 | The epigenetic landscape as depicted by C.H. Waddington | 11 |

## Chapter 2: CHD1 Maintains Transcription by Regulating Chromatin Accessibility to RNA Polymerase

| Figure 2.1 | Chd1 KO ES cells are viable and pluripotent | 31 |
| Figure 2.2 | Chd1 KO ES cells exhibit a mild growth defect | 33 |
| Figure 2.3 | H3K4me3 is reduced at a subset of ribosome related genes in Chd1 KO ES cells | 36 |
| Figure 2.4 | Chd1 KO ES cells have reduced RNA Pol II recruitment | 38 |
| Figure 2.51 | Generation of Chd1 KO ES cells | 44 |
| Figure 2.52 | Directed differentiation to EpiLSCs is not affected by loss of Chd1 | 45 |
| Figure 2.53 | Chd1 KO ES cells can self-renew in chemically defined medium | 46 |
| Figure 2.54 | H3.3 incorporation is not affected in Chd1 KO ES cells | 47 |
| Figure 2.55 | DNA damage response is not affected in Chd1 KO ES cells | 48 |
| Figure 2.56 | Validation of H3K4me3 ChIP-seq in Chd1 KO ES cells | 49 |
| Figure 2.57 | Cell number normalized expression in Chd1 KO ES cells | 50 |
CHAPTER 3: BIVALENT CHROMATIN MARKS DEVELOPMENTAL REGULATORY GENES IN THE MOUSE EMBRYONIC GERMLINE IN VIVO

**Figure 3.1:** Bivalent genes in E11.5 PGCs are enriched for developmental regulators 61

**Figure 3.2:** PGCs are enriched for bivalent developmental regulators compared to surrounding somatic cells at E11.5 63

**Figure 3.3:** Developmental regulators remain bivalent and transcriptionally repressed during sexual differentiation in the germline 65

**Figure 3.4:** Orphan bivalent domains overlap CpG Islands and show tissue specific expression 66

**Figure 3.5:** ChIP enrichment is specific and constant independent of starting cell number 73

**Figure 3.52:** Analysis of H3K4me3- and H3K27me3-marked genes in PGCs and ESCs 74

**Figure 3.53:** Germline specific genes are transcriptionally active and enriched for H3K4me3 75
CHAPTER 1

EPIGENETICS AS A REGULATOR OF THE TRANSCRIPTOME

Cellular Identity

Within a single multi-cellular organism, all cells share the same genetic makeup. However, a wide diversity of cell types arises with unique gene expression patterns. The cell type specific transcriptome is a key determinant of cellular function and behavior. While it has been well described how the underlying genetic sequence of transcription factors, promoters, and enhancers regulate gene expression, these elements have not been able to account for the diversity of transcriptomes observed. The varied utilization of the same genomic elements is thought to be largely regulated by epigenetics (Jaenisch and Bird, 2003; Niwa, 2007). The term epigenetics refers to a heritable change in phenotype of cells or organisms that does not involve a change in the underlying DNA sequence (Bártová et al., 2008).

Chromatin and Transcription

Epigentics regulates transcription in two ways: i) by regulating DNA accessibility for transcription factor binding and subsequent recruitment of the transcriptional machinery, and ii) the processivity of RNA polymerases along chromatin wrapped DNA. DNA within a cell is packaged in a nucleoprotein complex, known as chromatin that consists of DNA and DNA-binding proteins such as histones, which comprise a nucleosome core. The nucleosome core is a protein complex that consists of two H2A-H2B histone dimers and a H3-H4 histone tetramer (Santenard and Torres-Padilla, 2009) and is wrapped by 147bp of DNA (Clapier and Cairns, 2009). Chromatin can be categorized as either “loosely” packed, called euchromatin or “densely” packed called heterochromatin (Fedorova and Zink, 2008). Densely packed heterochromatic DNA is generally transcriptionally silent (Yasuhara and Wakimoto, 2006). In contrast, euchromatin is exposed and transcriptionally active (Fedorova and Zink, 2008). There are multiple regulatory mechanisms that contribute to the packaging of chromatin and its interactions with the
transcriptional machinery. These include post translational modifications to histones, histone protein variants, direct methylation of DNA, and nucleosome spacing, as summarized below.

The cytosine nucleotide of DNA can be covalently linked to a methyl group via a number of enzymatic pathways (Ooi et al., 2009). The occurrence of the dinucleotide C followed by a G (CpG) is often enriched in clusters at gene promoters, known as CpG islands, and has been implicated in regulating transcription (Koh and Rao, 2013). Methylation of CpG islands is associated with gene silencing and is found in heterochromatic chromatin. Significant cross-talk exists between the DNA methylation machinery and other chromatin remodeling pathways, but DNA methylation is generally considered the strongest level of epigenetic transcriptional repression (Meissner et al., 2008). Other chemical modifications of DNA have been described in additional contexts, but their relation to transcription in an area of open investigation. The reader is pointed to recent reviews on the topic (Ndlovu et al., 2011; Kohli and Zhang, 2013).

Post-translational modifications to the exposed N-terminal tail of histones regulate chromatin structure and the affinity of other DNA binding proteins, including the transcriptional machinery (Bártová et al., 2008). The histone code hypothesis speculates that combinations of these modifications determine the chromatin state of the DNA (Tordera et al., 1993; Turner, 1993; Strahl and Allis, 2000). In addition, histone modifications have been shown to modulate binding of many transcription factors by regulating chromatin density through a variety of histone modifications (Bártová et al., 2008).

Similar to the histone code is the histone barcode (Hake and Allis, 2006). This hypothesis recognizes the regular occurrence of histone variants at transcriptionally important regulatory regions of DNA and proposes an instructional role. A histone variant is a gene nearly identical to that of a core histone with the exclusion of a few amino acid residues. They can be incorporated into nucleosomes and are thought to recruit specific chromatin modifying enzymes, such as those responsible for post translation
modification of histones (see below). Through this mechanism the histone barcode has been proposed to define the histone code, which in turn is recognized by transcription factors and other transcriptional machinery. In support of this, depletion of histone variants can cause significant transcriptional and phenotypic defects (Creyghton et al., 2008; Tamura et al., 2009; Lin et al., 2013; Skene and Henikoff, 2013), although this is not always the case (Hodl et al., 2009; Hödl and Basler, 2012). Further investigation into the role of histone variants in regulation of transcription and chromatin compaction is necessary.

**Figure 1.1. Post translation histone modifications associated with transcription.** Adapted from Barth and Imhof, 2010 (Barth and Imhof, 2010).

Euchromatin is often associated with active gene transcription (Fedorova and Zink, 2008) (Figure 1.1). The majority of euchromatin is located in gene rich regions (Folle, 2008). There is minimal DNA methylation within these regions and several histone modifications associate with euchromatin and active transcription such as H3K4me3, H3K36me3, H3K79me2, H3K9Ac, and H3K14Ac (Krogan et al., 2003; Barski et al., 2007; Koch et al., 2007; Bártová et al., 2008; Ooga et al., 2008). Euchromatin is also characterized by a high level of histone acetylation mediated by histone acetyl transferases (HATs) (Bártová et al., 2008). Maintenance of these histone modifications is through the trithorax group proteins, which form several different complexes that establish the chromatin marks associated with
transcription (Glaser et al., 2006; Schuettengruber et al., 2007). Key components of these complexes include HMTs, HATs, and ATP-dependent chromatin remodelers (Schuettengruber et al., 2007). Recruitment of trithorax complexes to specific genetic loci is poorly understood, although correlative evidence suggests a role for the histone variants H3.3 and H2A.Z (McKittrick et al., 2004; Garcia and Thomas, 2008; Cui et al., 2009; Goldberg et al., 2010; Ku et al., 2012). In contrast to euchromatin, heterochromatin is compact DNA with little transcriptional activity and high levels of DNA methylation. Characteristic histone modifications in heterochromatin include H3K9me2, H3K9me3, H3K20me3, H3K27me3, and H3K79me3 (Bártová et al., 2008).

While chromatin is essential for the packaging of DNA into both euchromatin and heterochromatin, RNA polymerases must overcome the barrier imposed by nucleosomes at transcriptionally active genomic regions. The precise mechanism of nucleosome dynamics during transcription has yet to be described, but current evidence points to a system where DNA-nucleosome interactions are transiently disrupted during transcript elongation (Clapier and Cairns, 2009) (figure 1.2). Multiple chromatin remodeling complexes have been associated with this activity which can be segmented as three distinct molecular events: i) nucleosome disassembly, ii) reassembly, and iii) positioning.

Figure 1.2. Nucleosome dynamics during transcriptional elongation.
Transcription initiates with the progression of RNA Pol II from a paused state at a transcriptional start site. This required the transient displacement of nucleosomes mediated by a number of chromatin remodeling complexes. The multi-protein SAGA complex has been shown in yeast to associate with active RNA Pol II through recognition of H3K4me3 via Sgf29 and disrupts DNA-nucleosome contacts via acetylation of histone H4 via Gcn5 (Govind et al., 2007; Bian et al., 2011). H4 acetylation reduces the electromagnetic affinity of the nucleosome for DNA and may be a target for histone eviction by the SWI/SNF enzymes (Schwabish and Struhl, 2007). The transient nature of this dissociation in thought to be accomplished through a negative feedback loop. In addition to facilitating the acetylation of H4, a catalytic subunit of the SAGA complex, Ubp8, deubiquitinates H2BUb, which is a prerequisite histone modification required for H3K4 trimethylation; the mark by which SAGA is recruited to sites of active transcription in the first place (Wyce et al., 2007; Owen-Hughes and Gkikopoulos, 2012).

Reassembly of histones during transcription in yeast is thought to be mediated by the highly abundant FACT complex (Owen-Hughes and Gkikopoulos, 2012). While originally identified as a nucleosome evicting complex (Orphanides et al., 1998), FACT is now thought to participate in both nucleosome disassembly and reassembly. In the “global accessibility” model, FACT destabilizes the core histones of the nucleosome, but tethers all components in close proximity for nucleosome assembly or histone swapping (Formosa, 2012; Hsieh et al., 2013). Depletion of FACT in yeast results in loss of nucleosomes at sites of high transcriptional activity without an effect on the rate of transcription, indicating its role in nucleosome reassembly in the wake of RNA Pol II (Jamai et al., 2009). FACT is likely a key component in nucleosome reassembly during transcription, however if additional chaperones are required is an area of active investigation.

It is unclear how tightly coupled nucleosome positioning and transcription are, but binding of chromatin remodelers with known nucleosome positioning activity such as Chd1 and ISW1 are highly correlated
with transcriptional activity (Gaspar-Maia et al., 2009; Smolle et al., 2012). One hypothesis suggests that regularly positioned nucleosomes inhibit transcription and that transcription associated nucleosome positioning is meant to prevent antisense and cryptic transcription in the wake of Pol II induced chromatin remodeling. In other words, nucleosome positioning results as a response to the disruption of chromatin caused by active transcription. This is supported by studies in yeast that demonstrate increased intergenic and antisense transcripts in a number of strains mutant for chromatin remodelers with pronounced shifts of the +1 nucleosome (van Bakel et al., 2013) or global loss of nucleosome spacing (Whitehouse et al., 2007; Quan and Hartzog, 2010; Pointner et al., 2012).

**rRNA as Major Constituent of the Transcriptome**

The primary output of any transcriptional program is ribosomal RNA (rRNA), which is processed into the fundamental nucleoprotein complex; the ribosome. In humans and mice there are ~300 copies of the rRNA gene arrayed as direct repeats across 5 genomic locations, of which only a subset will be transcriptionally active in a single cell type (Hamperl et al., 2013). rRNA is transcribed primary by RNA Pol I and constitutes ~80% of total RNA in mammalian cells (Moss et al., 2007).

A deficiency in the rate of ribosome production will limit cell growth and proliferation. Therefore it is unsurprising that cell cycling and ribosomal biogenesis are closely coupled (Grummt, 2003). In many cases of impaired cellular metabolism, such as nutrient starvation, toxic lesion, or viral infection, the causative lever is unclear and the mechanism remains a chicken or the egg scenario. However, there are several instances that demonstrate rRNA transcription directly affecting the cell cycle. Stimulation of mouse and human cells by EGF, a promoter of proliferation, requires phosphorylation of UBF, a chromatin component that localizes to active rRNA genes (see below), by ERK (Stefanovsky et al., 2001). Similar works have also demonstrated that the FGF2 growth factor directly targets UBF (Bouche et al., 1987; Sheng et al., 2005). Interestingly, loss of genome-wide loss of DNA methylation in mammalian
cells is not lethal, but does slow cell proliferation. Paradoxically, the loss of this transcriptional inhibitory mechanism results in a reduction of rRNA due to aberrant antisense transcription by RNA Pol II competing with the RNA Pol I transcriptional machinery at rRNA genes, which disrupts rRNA processing (Gagnon-Kugler et al., 2009).

Regulation of rRNA transcription is accomplished using 3 levels of modulation: i) number of transcriptionally active rRNA gene loci, ii) RNA Pol I initiation, and iii) RNA Pol I rate of elongation. Artificial manipulation can alter the number of active rRNA gene loci in vitro, but there is no evidence to indicate this is a physiological cellular response under normal or stressed conditions, albeit further study is warranted (Moss et al., 2007). Unlike RNA Pol II, mammalian RNA Pol I initiation does not require phosphorylation of the polymerase (Gokal et al., 1990). Functional recruitment to the promoter is mediated by a phosphorylated Rrn3 protein (Zhao et al., 2003). Interestingly, phosphorylation of Rrn3 is due to a combination of RSK and ERK activities, and mutation of these sites in mouse Rrn3 suppresses rRNA transcription (Zhao et al., 2003). The ability of the mTor nutrient-sensing pathway and the Jnk stress response to affect Rrn3 activation is direct evidence the environment regulates Pol I initiation and rRNA transcription (Mayer et al., 2004, 2005). The gene body of an active rRNA gene is thought to be in an ‘open chromatin’ state characterized by reduced nucleosome occupancy. The phosphorylated HMG family protein UBF can form a DNA-binding dimer that creates a nucleosome-like structure, known as an ‘enhancesome,’ which is thought to regulate Pol I elongation via its reversible phosphorylation by ERK1/2 (Stefanovsky et al., 2001, 2006). UBF is also subject to acetylation by a number of enzymes, although the functional significance of this unclear (Moss et al., 2007). Taken together, there are a number of mechanisms in which rRNA transcription and ribosome biogenesis are regulated, of which the mostly likely to involve an epigenetic component are modulation of active rRNA genes and regulation of Pol I elongation.
Defects in ribosome biogenesis produce a number of developmental defects and diseases in mammals. Knockout mice for many of the ribosomal protein (RP) genes are embryonic lethal, have growth defects, or exhibit anemia and other hematopoietic defects (Terzian and Box, 2013). Mutant phenotypes often appear to be the results of global suppression of protein synthesis due to reduced ribosomal output. Several of these maladies are conserved in humans (Narla and Ebert, 2010). Of particular interest is CHARGE syndrome which is characterized by growth retardation amongst other clinical features. Two thirds of CHARGE cases are caused by a loss of function mutation in Chd7 (Zentner et al., 2010b). Mice heterozygous for Chd7 exhibit similar pathologies, while homozygous mutants die by E11, thus further demonstrating the importance of this chromatin remodeler during development (Bosman et al., 2005; Hurd et al., 2007). A recent mechanistic study revealed Chd7 directly regulates rRNA gene transcription (Zentner et al., 2010a). Mutant embryos and embryonic stem (ES) cells expressed reduced levels of pre-rRNA, reduced protein synthesis, and slower cell proliferation. Further evidence for a role of Chd family remodelers regulating ribosome biogenesis can be found in Chd4; as a subunit of the NuRD complex. Depletion of Chd4 in mouse cell culture reduces pre-rRNA synthesis and cell proliferation (Xie et al., 2012).

Chd1

Chromodomain helicase DNA binding protein 1 (Chd1) is a conserved gene that is highly expressed in multiple cells including ES cells (Ramalho-Santos et al., 2002; Efroni et al., 2008). It is associated with transcriptional initiation, elongation, and chromatin remodeling in vitro, in Drosophila, and whole cell chromatin rearrangement during silk moth development (Lusser et al., 2005; Srinivasan et al., 2005; Konev et al., 2007; Papantonis et al., 2008; Lin et al., 2011). In support of this, Chd1 has been shown to directly interact with Mediator as a component of the pre-initiation complex (PIC) during RNA Pol II recruitment and in vitro experiments suggest a role for Chd1 in transcriptional activation (Lin et al., 2011).
Interestingly, Chd1 has also been identified as a component of the SAGA and FACT complexes during transcriptional elongation (Kelley et al., 1999; Simic et al., 2003; Pray-Grant et al., 2005; Sims et al., 2007). Chd1 has a nearly identical DNA binding profile as RNA Pol II in *Drosophila* and mammalian cells (Mito et al., 2005; Srinivasan et al., 2005; Gaspar-Maia et al., 2009). Despite its RNA Pol II association and ability to enhance transcription, it is not required for initiation or elongation (Sims et al., 2007). It is possible that Chd1 has different roles in different genomic or cellular contexts.

Chd1 has a single homolog, *Hrp1*, in *Saccharomyces cerevisiae* and two homologs in *Saccharomyces pombe; Hrp1 and Hrp3*. Under normal conditions, Chd1 is not required in yeast and knockout has minimal affects on survival or transcription (Gkikopoulos et al., 2011; Hennig et al., 2012; Pointner et al., 2012). However, Chd1 null strains do experience a loss of regular nucleosome positioning at coding regions (Gkikopoulos et al., 2011; Hennig et al., 2012; Pointner et al., 2012; Shim et al., 2012). This nucleosome sliding enzymatic activity has been confirmed in vitro for *Drosophila* and yeast (Lusser et al., 2005; Nodelman and Bowman, 2013).

Maternal Chd1 has been shown to be essential for proper male pronucleus decondensation during *Drosophila* fertilization. Lack of H3.3 incorporation was associated with Chd1 by demonstrating a phenotype similar to what is seen in the H3.3 transportation chaperone HIRA1 mutant (Lusser et al., 2005; Konev et al., 2007). While this data shows Chd1 as essential for proper chromatin remodeling during *Drosophila* fertilization, it is still unknown if Chd1 acts directly through H3.3.

Recently, Chd1 was identified as a gene required for optimal self-renewal and pluripotency in mouse ES cells using an RNAi screen (Gaspar-Maia et al., 2009). Chd1 RNAi ES cells could be maintained in culture but displayed differentiation defects and a propensity to accumulate heterochromatin. While it has been suggested that this defect is due to a lost maintenance of open or relaxed chromatin, a precise role in mammalian biology remains undefined.
**EPIGENETIC INHERITANCE**

*Pluripotency and Open Chromatin*

Maintenance of the epigenetic landscape is an essential part of retaining cellular identity through cell division. Loss of enzymes required for propagation of epigenetic marks from mother to daughter cell results in cell identity and specification defects (Rajasekhar and Begemann, 2007; Schuettengruber et al., 2007). This is important during mammalian development at the pre-implantation stage where the pluripotent Inner Cell Mass (ICM) will eventually give rise to every cell type within the developing embryo. The small population of pluripotent ICM cells is a particularly intriguing cell type because they must propagate and expand in a manner that preserves their unspecified cell state, yet retain the plasticity to activate a specific transcriptional program later in development. This transcriptional plasticity is thought to be regulated by maintaining genes in a unique ‘open chromatin’ state. Embryonic Stem (ES) cells are a convenient in vitro model that retains the pluripotent potential of the ICM from which they are derived.

The predominant theory states that pluripotent cells early in development maintain an open chromatin state in which genes are potentially available for expression. The ES cell genome is transcriptionally hyperactive with widespread transcription in both coding and non-coding regions, including sporadic low level expression of tissue specific genes (Carter et al., 2005; Efroni et al., 2008). As development proceeds and cell lineage choices constrict, the chromatin becomes increasingly compact, thus restricting gene expression to a select few genes that constitute the transcriptome of a cell. This general concept was originally proposed by Conrad Waddington and the phenomena of epigenetic mediated lineage restriction is known as Waddington’s canal (Figure 1.3) (Waddington, 1942, 1957, 1968).
Bivalent Domains and Pluripotency

Pluripotency is dependent on the maintenance of a proper epigenetic landscape (Gaspar-Maia et al., 2011; Orkin and Hochedlinger, 2011; Young, 2011). Bivalent domains, which are defined by the paradoxical coexistence of a permissive histone mark (H3K4me3) and a repressive one (H3K27me3), are thought to play an important role in pluripotency by keeping developmental genes in a silenced state poised for activation upon differentiation (Azuara et al., 2006; Bernstein et al., 2006). According to this model, bivalent genes resolve to H3K4me3-only if they are activated upon differentiation into a particular lineage, and resolve to H3K27me3-only in the alternative lineages where they are not activated (Bernstein et al., 2006; Mikkelsen et al., 2007). In support of this model, bivalent domains have been shown to be present preferentially at developmental regulatory genes in undifferentiated ES cells and several adult tissues in vivo including sperm, testis, the cerebellum, and the hematopoietic compartment (Bernstein et al., 2006; Roh et al., 2006; Mikkelsen et al., 2007; Cui et al., 2009; Hammoud et al., 2009; Adli et al., 2010). In addition, the moderate de-repression of developmental regulators in Eed mutant ES cells, which have reduced levels of H3K27me3, further supports this model (Boyer et al., 2006; Chamberlain et al., 2008).

However, it has recently been proposed that the concomitant presence of H3K4me3 and H3K27me3 at developmental genes in human ES cells is an in vitro artifact that results from a heterogeneous mixture of multiple monovalent subpopulations (Hong et al., 2011) or from culture of mouse ES cells in suboptimal conditions (Marks et al., 2012). In addition, direct evidence for bivalency in the developing mammalian embryo is very limited and conflicting due to technical difficulties associated with the low amounts of material available. Two genes bivalent in ES cells were shown to be bivalent in the inner cell
mass of the mouse blastocyst (Alder et al., 2010), and similar results were found for four genes in the E5.5 epiblast (Rugg-Gunn et al., 2010). However, a separate genome-wide study reported that genes bivalent in ES cells are not bivalent in the inner cell mass of the blastocyst (Dahl et al., 2010). The universal nature of bivalency has been further questioned due to conflicting reports from non-mammalian species. While bivalent domains have been demonstrated to exist in zebrafish blastomeres (Vastenhouw et al., 2010), multiple studies in Xenopus and Drosophila embryos have failed to find evidence for bivalency (Akkers et al., 2009; Schuettengruber et al., 2009; Schneider et al., 2011). We therefore lack a clear understanding of whether bivalency exists in embryonic cells, and how it relates to pluripotency.

*Latent Pluripotency and the Germline*

Segregation of germ cells from somatic cells is a critical aspect of multi-cellular organism development that ensures perpetuation of all heritable information. In many species this is accomplished through “pre-formation,” which relies on the maternal deposition of determinants, known as the germ plasm, at the beginning of development that specify a unique cell type separate from the soma which will eventually form the germ line. In mice and likely all mammals, the process of “epigenesis” induces, through external signaling, the specification of Primordial Germ Cells (PGCs) from somatic cells to form the precursors to the mature oocytes and spermatogonia (Extavour and Akam, 2003). In both scenarios, the germ line exists as a cell type with latent pluripotent potential that will not be activated until the next generation.

*Primordial Germ Cell Specification*

In the mouse, expression of *Fragilis* in the most proximal epiblast cells at ~E6.25-6.5 marks a small population of PGC precursor cells (Saitou et al., 2002). A short time later at ~E7.0-E7.25 in the extra-embryonic mesoderm, a population of double positive *Fragilis* and *Dppa3* positive cells develop that
repress Homeobox genes such as *Hoxb1* and *Hoxa1*, which are widely expressed elsewhere in the embryo (Saitou et al., 2002). However, while important for maintaining genomic integrity of PGCs, knock-out studies have shown neither *Fragilis* nor *Dppa3* are required for PGC specification (Payer et al., 2003; Nakamura et al., 2007; Lange et al., 2008).

Specification of the PGC precursors is thought to happen in the most proximal epiblast cells at E6.25 with the expression of *Blimp1*; similar to the onset of *Fragilis* expression (Ohinata et al., 2005). Initially, the *Blimp1* positive population expresses somatic genes such as *Hoxb1* and *Hoxa1* similar to surrounding somatic mesodermal cells (Kurimoto et al., 2008), but later with the activation of *Prdm14* and *Tcfap2c* at E6.5 and E6.75 respectively, the somatic program, including expression of Homeobox genes, is silenced (Yamaji et al., 2008; Weber et al., 2010).

While *Blimp1* is widely expressed compared to *Fragilis*, knock-out of *Blimp1* causes PGC specification defects that include reduced numbers of *Dppa3* positive cells, continued expression of the somatic program, aberrant migration, and the eventual loss of all PGCs (Ohinata et al., 2005; Vincent et al., 2005; Kurimoto et al., 2008). Similarly, *Prdm14* deficient mice do not correctly specify a germline (Yamaji et al., 2008). *Prdm14* deficient cells do not reactive pluripotency marks such as Sox2 and no not reduce H3K9me3 and up-regulate H3K27me3 respectively during the subsequent epigenetic reprogramming (see below). However, *Prdm14* deficient cells do successfully silence the somatic program. *Tcfap2c* is ubiquitously expressed until the blastocyst stage at E3.5, but thereafter its expression is confined to the extra-embryonic ectoderm and E6.75-E13.5 PGCs (Development et al., 2001; Werling and Schorle, 2002; Weber et al., 2010). An epiblast specific knock-out of *Tcfap2c* results in a drastic loss in PGC number at specification, migration defects, and their eventual loss (Weber et al., 2010). Taken together, there have been a number of genes identified that are critical to PGC specification with overlapping but distinct functions; i) *Blimp1* silences the somatic program and activate PGC/pluripotency genes, ii) *Prdm14*
induces PGC/pluripotency genes and activates part of the epigenetic reprogramming, iii) Tcfap2c is required for PGC proliferation and migration.

**Epigenetic Reprogramming in the Germ Line**

Current evidence points to a phase of epigenetic erasure and reprogramming in mouse PGCs that is thought to prevent transgenerational transmission of epimutations via removal of both DNA methylation and histone modifications. Shortly after PGC specification, initiation of global DNA demethylation is evident by E8.5 (Seki et al., 2005), and complete demethylation by E13.5 includes X-chromosome reactivation in females, erasure of genomic imprinting, and demethylation of transposable/repetitive elements (Popp et al., 2010; Guibert et al., 2012; Saitou et al., 2012). Concurrent with the initiation of DNA demethylation, global changes to histone modifications are observed with the loss of H3K9me2 by E7.75 and an increase in H3K27me3 by E9.5 (Seki et al., 2005, 2007). However, due to the small number of PGCs at these time points, histone modification data is limited to immunofluorescent experiments and high resolution correlations to DNA methylation loss remain unknown.

A number of recent high resolution studies have confirmed that most genomic loci are hypo-methylated by E9.5 (Seisenberger et al., 2012; Kobayashi et al., 2013), but a subset of loci progressively, asynchronously demethylate until E13.5 (Seisenberger et al., 2012; Hackett et al., 2013). This subset of loci that undergo late epigenetic reprogramming include PCG specific genes, imprinted loci, and various transposable/repetitive elements. Notably, IAP elements and a couple hundred CpG islands not linked to IAPs do not completely demethylate in the germline and are potential vehicles for transgenerational epigenetic inheritance (Hajkova et al., 2002; Seisenberger et al., 2012; Hackett et al., 2013). Only in the last several months have high resolution genome-wide data describing histone modifications in PGCs become available (Lesch et al., 2013; Ng et al., 2013; Sachs et al., 2013). These data will be discussed in chapter 3 and the implications in chapter 4.
**Similarities between the Germ Line and ES Cells**

PGCs have long been known to spontaneously give rise to teratomas in vivo (Stevens, 1964). Interestingly, this is the same type of multi-lineage carcinoma formed by ES cells implanted in mice. Reflective of this similarity, PGCs can be isolated and cultured in vitro to form an immortal, pluripotent cell type called Embryonic Germ Cells (EGCs), which are phenotypically indistinguishable from ES cells derived from the ICM (Matsui et al., 1992; Resnick et al., 1992; Labosky et al., 1994). However, EGC derivation is possible only during a limited window of PGC development, E8.5-E12.5, which correlates with epigenetic reprogramming events (Matsui et al., 1992; Tada et al., 1998; Durcova-Hills et al., 2001). The significance of this is unknown.

In addition, ES cells and PGCs are transcriptionally similar. PGCs express most of the master transcription factors known to be essential for pluripotency such as Pou5f1, Sox2, Myc, and Nanog (Sabour et al., 2011). Ablation of many of these, results in the loss of PGCs to apoptosis or differentiation (Kehler et al., 2004; Chambers et al., 2007; Saitou et al., 2012). Interestingly, human PGCs do not express Sox2, suggesting a distinct method of regulating pluripotency (Perrett et al., 2008). As noted above, a distinguishing characteristic of ES cell chromatin is the presence of bivalent domains at developmental regulatory genes (Azuara et al., 2006; Bernstein et al., 2006). Until recently, high resolution investigation of PGC chromatin was not technically feasible. Chapter 2 presents for the first time high resolution, high quality data describing the chromatin state of PGCs during the suspected window of epigenetic reprogramming and a comparison to ES cells.

Despite the phenotypic, transcriptional, and epigenetic similarities between ES cells and PGCs, PGCs are unipotent; not pluripotent, and cannot contribute to chimeras as ES cells do (Leitch et al., 2013). Under physiological conditions, PGCs only retain a latent form of pluripotency to be activated in the zygote of
the next generation. The precise mechanism in which the germline represses pluripotency is unknown, and its elucidation will have impacts on the fields of development, inheritance, and evolution.

References


Labosky, P. a, Barlow, D.P., and Hogan, B.L. (1994). Mouse embryonic germ (EG) cell lines: transmission through the germline and differences in the methylation imprint of insulin-like growth factor 2 receptor (Igf2r) gene compared with embryonic stem (ES) cell lines. Development (Cambridge, England) 120, 3197–3204.


Sheng, Z., Liang, Y., Lin, C., Comai, L., and Chirico, W.J. (2005). Direct Regulation of rRNA Transcription by Fibroblast Growth Factor 2 Direct Regulation of rRNA Transcription by Fibroblast Growth Factor 2. 25,


CHAPTER 2

Chd1 Maintains Transcription by Regulating Chromatin Accessibility to RNA Polymerase

Summary

Cell survival, pluripotency, and cell fate determination are dependent on maintenance of a proper epigenetic landscape which in turn regulates both the accessibility of the transcriptional machinery and its efficiency. Chromodomain helicase DNA binding protein 1 (Chd1) is a chromatin remodeling enzyme essential for mammalian embryogenesis. However, the precise mechanism of action for Chd1 has yet to be identified. While it has been assigned multiple and often conflicting chromatin remodeling activities in different studies, Chd1 is consistently associated with the transcriptional initiation and elongation machinery. While Chd1 is essential for early mammalian development, in this study I find that it does not appear to be required for survival of ES cells or induction of the major germ layers during in vitro differentiation. However, Chd1 deficient ES cells exhibit a media dependent growth defect in culture that is likely due to slower cell cycling. A rigorous cell number normalized transcriptional analysis has revealed what is likely a global reduction in transcription evident at nearly every RNA Pol II dependent gene tested. This is supported by reduced Pol II binding. Concomitant with this transcriptional reduction, a genome-wide analysis showed a small, but highly transcribed set of genes reduced for H3K4me3 in the gene body. These genes are highly enriched for functions in ribosome biogenesis, suggesting a direct link between reduced ribosome production and slower cell grown mediated by regulation of the chromatin state.
Introduction

Proper regulation of chromatin state is essential for coordinating the myriad of transcriptional changes taking place in the developing mammalian embryo. Chromodomain helicase DNA binding protein 1 (Chd1) has been suggested as an essential regulator of chromatin in an in vitro model of mammalian pluripotency and development (Gaspar-Maia et al., 2009). In support of this, unpublished data from the Ramalho-Santos lab has revealed that Chd1 is required for the rapid expansion of the early post-implantation (E5.5-6.5) mouse epiblast. This is in contrast to findings in Drosophila which require Chd1 only for female fertility, and Saccharomyces cerevisiae which does not require Chd1 under normal conditions (Konev et al., 2007; Gkikopoulos et al., 2011; Hennig et al., 2012; Pointner et al., 2012). An explanation for this discrepancy remains open to study.

Common to all organisms studied, Chd1 associates with the transcriptional machinery and mammalian Chd1 directly binds H3K4me3 (Kelley et al., 1999; Simic et al., 2003; Flanagan et al., 2005; Pray-Grant et al., 2005; Sims et al., 2007; Gaspar-Maia et al., 2009; Lin et al., 2011; Shim et al., 2012; Smolle et al., 2012). In eukaryotes, RNA polymerase must pass through nucleosomes packaged as chromatin. To overcome this obstacle, a number of chromatin remodeling complexes exist as part or in conjunction with the transcriptional machinery. Essential to this process is the disassembly or displacement of nucleosomes, followed by nucleosome reassembly or incorporation in the wake of elongating RNA polymerase (Williams and Tyler, 2007). Chd1 has been proposed to participate in both aspects of this process, although in varying capacities depending on context.

RNA Pol II must first bind a gene at the 5′-end of the transcriptional start site (TSS) as part of the initiation complex before transcriptional elongation through the gene body. Growing evidence suggests Chd1 has opposing roles related to nucleosome assembly dependent on its association with either the initiating or elongating transcriptional machinery (Petty and Pillus, 2013). In support of a role for
nucleosome disassembly at promoters, Chd1 interacts with the mediator and SAGA complexes, which are involved with nucleosome disassembly during formation of the pre-initiation complex (PIC), RNA Pol II recruitment, and transcriptional initiation (Pray-Grant et al., 2005; Lin et al., 2011). In *Saccharomyces cerevisiae*, loss of Chd1 induces increased nucleosome occupancy specifically at promoters (Ehrensberger and Kornberg, 2011; Radman-Livaja et al., 2012). In contrast, Chd1 is required for incorporation of the histone variant H3.3 and CENP-A in *Drosophila* and HeLa cells respectively (Walfridsson et al., 2005; Konev et al., 2007; Okada et al., 2009). Loss of Chd1 is also associated with reduced nucleosome occupancy at highly transcribed gene bodies in *Drosophila* and *Saccharomyces cerevisiae* (Lee et al., 2012; Radman-Livaja et al., 2012). This data suggests Chd1 regulates nucleosome assembly during transcriptional elongation.

Despite the uncertainty of the molecular mechanism of Chd1, the cellular consequence of Chd1 loss in the mammalian system is still unknown. In this study I generated Chd1 knockout ES cells and show a depletion of H3K4me3 at a sub-set of highly expressed genes concurrent with a wider global reduction in transcription previous unreported in Chd1 null models. I suggest this is a direct result of reduced RNA polymerase binding and processivity.

**Results**

*Chd1 KO ES cells appear pluripotent in vitro but have a medium dependent self-renewal defect*

Chd1 KO mouse embryos are embryonic lethal at E5.5-6.5 and fail to initiate gastrulation (personal communications M. Guzman). In order to investigate the cellular function of Chd1, I generated a number of Chd1 depleted ES cells (Table 2.S1). Chd1 KO ES cells were created lacking an exon that codes for part of the helicase domain and introduces a frame-shift using a sequential targeting method (Figure 2. S1A, S1B) and tamoxifen inducible Chd1 KO ES cells were derived from E3.5 blastocyst outgrowths (gene
Figure 2. Chd1 KO ES cells are viable and pluripotent in vitro. (A) Pre and post tamoxifen treatment of a Chd1 inducible KO (iKO) ES cell colony. 10x objective. (B) Transcription was measured with qRT-PCR for the indicated genes during tamoxifen treatment of 3 independently derived Chd1 iKO ES cell lines. All qRT-PCR data represented as a percent of the housekeeping gene L7 ± standard deviation. (C) qRT-PCR for the indicated genes of 4 independently derived Chd1 sequential KO (seqKO) and control ES cell lines. (D) qRT-PCR for the indicated genes at day 4 of embryoid body formation of 4 Chd1 seqKO and a control ES cell lines. Genes are reporters for specific germ layers as indicated. (E) Self-renewal capacity of 2 chd1 seqKO and control ES cell lines assessed using a colony formation assay in FBS. Error bars represent standard deviation of 2 replicate wells.
targeting done by F. M. Koh and ES cell derivation by C. Lin). Both model systems were validated for loss of full length Chd1 at the RNA and protein level (Figure 2. 1A, 1B, S1C). Loss of Chd1 did not affect induction of markers for all germ layers upon differentiation to embryoid bodies (Figure 2. 1C). A directed differentiation toward epiblast-like cells (EpiLCs) showed that there was no effect on the kinetics of differentiation upon loss of Chd1 (Figure 2. S2A, S2B). While cultures of EpiLCs are difficult to maintain long-term, Chd1 mutant EpiLCs appeared to maintain expression of pluripotency and epiblast specific markers at levels similar to controls for up to 19 days (Figure 2. S2C, S2D). The pluripotency defects previously reported in Chd1 RNAi ES cells (Gaspar-Maia et al., 2009) may have been due to different culture conditions, and in particular the use of unpurified house-made LIF of unknown activity then versus commercial recombinant and purified LIF in this study (data not shown). Although the presence of additional growth factors in the house-made LIF cannot be ruled out. Of note, all of the above expression analyses were not cell-number normalized for RNA content. This common practice introduces caveats that I will come back to later. Nevertheless, they clearly indicate that Chd1 KO ES cells can induce differentiation markers of all major lineages.

While seemingly pluripotent in vitro, Chd1 KO ES cells do display a mild, medium dependent defect in their ability to self-renew consistent with previous findings (Gaspar-Maia et al., 2009) (Figure 2. 1E, S3A). To further explore this self-renewal defect, precise measurements of cell growth were recorded during routine culture maintenance. Chd1 KO ES cells exhibited a 23% reduction in cell number compared to either wild type or heterozygous controls when cultured in a serum based medium (Figure 2. 2A, 2B, S3B, S3C). It stands to reason these lower cell counts could be due to either increased cell death or decreased growth. Dual staining for the pre-apoptotic and apoptotic markers, AnnexinV and PI respectively, indicated low rates of apoptosis similar to control cell lines when grown in 2i media (Figure 2. 2C). An assessment of cell cycle phases using a DNA dye and flow cytometry showed no significant change in the ratio of cells in S-phase to either G0/G1 or G2/M when grown in serum-based media
Figure 2.2. Chd1 KO ES cells exhibit a medium dependent self-renewal defect. (A) Calculated doubling time of 2 seqKO and control ES cell lines grown in serum based media for 10 passages. (B) Absolute cell number of 3 Chd1 iKO ES cell lines grown in either a serum based or 2i media as indicated. Each line represents the cell number difference between a cell line previously treated with tamoxifen (+4OHT) and one without treatment (-4OHT) during routine passaging. All cell lines were counted and seeded at equal densities at each passage. (C) FACS plots of 2 Chd1 iKO lines with (+4OHT) or without (-4OHT) tamoxifen treatment for a marker of pre-apoptotic cells (Annexin V) and apoptotic cells (PI). Bar graph represents quantification of FACS plots using gates pictured.
(personal communication F. Ming Koh). The inability to detect changes to the cell cycle or rates of apoptosis may be a reflection of the sensitivities of these assays in view of the mild reduction in cell number, or sensitivity to specific culture conditions. Alternatively, a growth defect that slows all phases of the cell cycle would not be revealed by these assays.

Changes to chromatin state

I next examined the chromatin state of Chd1 deficient ES cells. Chd family proteins have been implicated in many cellular functions and I aimed to identify changes to the chromatin state that would indicate where Chd1 was functioning. It has been theorized, based on experiments in Drosophila, that Chd1 is required for incorporation of the histone variant H3.3 (Konev et al., 2007; Goldberg et al., 2010). Analysis of global and locus specific H3.3 incorporation using a myc-tagged H3.3 transgene in Chd1 KO ES cells showed normal H3.3 incorporation (Figure 2. S4A, S4B). Similar results were obtained using a Chd1 shRNA knock-down in a H3.3-HA knock-in ES cell line, ruling out any effect of super-physiological H3.3 levels (data not shown).

As a component of the NuRD complex, multiple members of the Chd family, including Chd3, Chd4, and Chd5, have been implicated as both positive and negative regulators of double strand break (DSB) DNA repair (Stanley et al., 2013). While Chd1, has been directly associated with the DSB response or the NurD complex, I nonetheless examined γH2A.X (H2A.XS139P) levels as markers for DNA damage repair, but saw no significant change (Figure 2. S5B). A Chd1 mutant S. cerevisiae strain was observed to have reduced levels of the transcription associated mark H2Bub (Lee et al., 2012), but this was not robustly replicated in murine ES cells (Figure 2. S5A).

Due to the association of Chd1 with H3K4me3, I sought to characterize the genome-wide distribution of this mark in Chd1 KO ES cells using ChIP-seq. Surprisingly there were few changes, with only 87 genes located proximal to a region of H3K4me3 depletion and 0 genes exhibiting a gain of H3K4me3 (Table
In contrast, ChIP-seq for a mark associated with transcriptional repression with no known interaction with Chd1, H3K27me3, detected zero genes with either a gain or loss. (Figure 2. S2A). Validation of the H3K4me3 depletion was confirmed using ChIP-qPCR in two independent ES cell lines (Figure 2. S6B). Whereas H3K4me3 is typically concentrated at the transcriptional start site (TSS) of actively transcribed genes, all regions of H3K4me3 depletion occurred in the gene body of highly expressed genes (Figure 2. 3A, 3B). In support of this being a direct effect of Chd1 depletion, genes with lower H3K4me3 are highly enriched (p=1.61x10^{-16}) for known Chd1 targets (Figure 2. 3C). I next performed a gene ontology analysis on genes with depleted H3K4me3 and discovered a strong enrichment for components of the ribosome and ribosome biogenesis (Figure 2. 3E). The rRNA gene occurs in multiple large tandem repeats and has thus not been mapped to the current mouse genome assembly. To further explore an effect on ribosome biogenesis I aligned the H3K4me3 ChIP-seq data directly to the 45.3kb rRNA gene consensus sequence, and noted a 23% reduction in signal (Figure 2. 3D). Taken together, these data demonstrate a reduction of H3K4me3 at ribosome related genes, which are some of the highest transcribed genes in a cell.

**Cell number normalized gene expression and RNA Pol II processivity**

Considering the association of H3K4me3 and Chd1 with transcription, I reasoned distinct transcriptional changes would be evident in the Chd1 KO ES cells. However, similar to Chd1 null S. cerevisiae and Pombe strains, a microarray analysis showed few genes changing either up or down (Figure 2. S7A) (Gkikopoulos et al., 2011; Hennig et al., 2012; Pointner et al., 2012). I then reasoned that Chd1 may be altering global transcription levels considering its ubiquitous association with RNA Pol II activity. In such a scenario, standard relative expression level techniques such as microarrays and qRT-PCR would not be able to detect a global shift in expression (Lin et al., 2012; Lovén et al., 2012; Nie et al., 2012). In order to overcome this obstacle, cell number normalized qRT-PCR was performed on a number of genes covering
Figure 2.3. H3K4me3 is reduced at a subset of ribosome related genes in Chd1 KO ES cells. (A) Representative tracks for H3K4me3 and H3K27me3 in a Chd1 het and KO line at the gene loci indicated. (B) Fraction of the 87 genes with reduced H3K4me3 in the indicated top percentile of expressed genes. Expression ranking based on wild type ES cells. (C) Genes with reduced H3K4me3 in Chd1 KO ES cells are enriched for known Chd1 targets. Chd1 target gene set from (Gaspar-Maia et al., 2009). (D) H3K4me3 ChIP-seq reads from control and Chd1 KO ES cells aligned the mouse rRNA gene consensus sequence. Read count is normalized for total reads per library. (E) Unfiltered gene ontology terms for the H3K4me3 reduced gene set in Chd1 KO ES cells.
a range of expression levels (Figure 2. 4A). Remarkably, every gene examined showed some degree of transcriptional deficiency. This included markers associated with pluripotency such as *Pou5f1, Nanog,* and *Klf2,* as well as a housekeeping gene, *Actg1,* and several genes associated with translation and ribosome function; *Rpl3, Rps9, Taf1d,* and *Elf4a2.* Remarkably, this down regulation was evident at mature and pre-splicing mRNAs, suggesting a reduction in RNA Pol II recruitment and/or processivity rather than splicing efficiency or mRNA stability. The reduction of pre-splicing mRNA was observed in 2/2 independent ES cell lines, and the reduction in mature mRNAs in 2/3 independent lines (Figure 2. 4A, S7B, S7C). To further explore deficiencies in RNA Pol II recruitment, ChIP-qPCR was performed for RNA Pol II in two different phosphorylation states (Figure 2. 4B, 4C). Both RNA Pol II S5p and S2p, which correspond to ‘paused’ and ‘elongating’ RNA Pol II respectively, showed reduced occupancy at the TSS and gene body of the two highly transcribed genes examined; *Rpl3* and *Rps9.* The effect was less evident at the *Klf2* locus which is transcribed at approximately 10 fold lower levels.

**Discussion**

Taken, together this study clarifies many outstanding questions in the field regarding the function of Chd1 in the role of pluripotency, cell survival, and development. Surprisingly, many of the roles assigned to Chd1 in lower organisms and knock-down models, such as in vitro pluripotency and H3.3 incorporation, appear to be unaffected by the total loss of Chd1 in a mammalian system (Konev et al., 2007; Gaspar-Maia et al., 2009; Goldberg et al., 2010). The findings from this study point to a role of Chd1 in regulating polymerase processivity, similar to Chd4 and Chd7 (Shimono et al., 2005; Zentner et al., 2010; Xie et al., 2012).
Figure 2.4. Chd1 KO ES cells have reduced RNA Pol II recruitment. (A) Cell number normalized qRT PCR for indicated genes. ‘Mature mRNA’ represents an intron-spanning primer set that only detected post-splicing mRNA. ‘Pre-mRNA’ represents a primer set amplifying an exon-intron boundary and only detects pre-splicing mRNA. Data represented as percent of indicated gene in control cell line ± standard deviation of 2 cell number normalized replicates. (B) ChIP-qPCR for S5p RNA Pol II at the indicated gene represented as the mean percent of input ± standard deviation of 2 technical replicates. (C) ChIP-qPCR for S2p RNA Pol II represented as in (A). Transcriptional start site (TSS); gene body (GB); pre-spliced rRNA (pre-rRNA); gene dessert (cntrl).
I have demonstrated a reduction in transcription and RNA Pol II binding at several RNA Pol II dependent genes when normalized for cell number and a mild reduction of H3K4me3 at a sub-set of these genes. These findings are in agreement with a previous investigation using a reporter/cell extract model that found Chd1 cooperates with the PIC to facilitate RNA Pol II mediated transcription (Lin et al., 2011). A genome-wide analysis using cell number normalized samples will be necessary to confirm the suspected global reduction in transcription at RNA Pol II dependent genes. It is important to note that standard genome-wide analytical techniques failed to detect any discernible change and that any truly global regulator will require cell number normalized techniques to unmask their function.

Why is the reduction in H3K4me3 not seen at all genes with reduced transcription? Mammalian Chd1 can bind H3K4me3, but has not been implicated in its incorporation (Flanagan et al., 2005; Pray-Grant et al., 2005; Sims et al., 2007). It is likely that the reduction of H3K4me3 is an indirect affect of Chd1 loss mediated by the lower RNA Pol II activity. This change may only be detectable at a sub-set of the most highly transcribed genes in a cell.

Intriguingly, preliminary qRT-PCR and ChIP data (Figure 3. S7B) point to a reduction in rRNA transcription, suggesting a role for Chd1 in RNA Pol I dependent transcription similar to Chd4 or Chd7 (Shimono et al., 2005; Zentner et al., 2010; Xie et al., 2012). However, antisense RNA Pol II mediated transcription has been shown to indirectly regulate RNA Pol I processivity at rRNA genes (Gagnon-Kugler et al., 2009). Further studies investigating changes in antisense transcript levels and RNA Pol I/II binding at rRNA genes will be necessarily to clarify the role of Chd1 in RNA Pol I mediated transcription.

Many of the mild self-renew defects seen in the Chd1 KO ESCs were only evident in a serum based medium and not chemically defined media such as 2i or KSR. This selectivity may be explained by the rapid cell cycling of ES cells in serum based media compared to defined media. Cross-talk between the cell cycle and global transcription has been well described, often with a concomitant shift of one when
the other is altered (Grummt, 2003). Therefore, the slower growth rate in defined media is likely indicative of lower rates of transcription and ribosome biogenesis. Loss of Chd1 may only become critical at the higher rates of transcription evident in a serum based medium. Data from in vivo mouse embryogenesis studies indicate that loss of Chd1 becomes critical in highly proliferative cell types, such as the E5.5 epiblast and fetal liver blood progenitors (personal communications M. Guzman & F. M. Koh). Future studies will be needed to clarify how a chromatin remodeler associated with transcriptional regulation can influence cell-cycle.

**Experimental Procedures**

**Cell Culture**

Mouse ES cells were routinely passaged in gelatin coated 6-well plates every other day. Upon passage cells counts were obtained using a ViaCell (Beckman Coulter) and either 200k or 600k cells were seeded per 6-well well for serum media or 2i media respectively. Serum media consisted of 15% FBS, 1x Penicillin/Streptomycin, 1x Non-essential amino acids, 0.057mM beta-mercaptoethanol, and 1:10,000 diluted Lif (ESGro, Millipore) in high glucose DMEM supplemented with GlutaMAX and pyruvate (Gibco, 10569-10). 2i media was formulated with 1x Penicillin/Streptomycin, 1x N2 supplement (Gibco, 17502048), 1x B27supplement (Gibco, 17504-044), 50µg/ml BSA Fraction V, 1:10,000 diluted Lif (ESGro, Millipore), 1µM Mek inhibitor (PD0325901), and 3µM Gsk3 inhibitor (CHIR99021) in a based media consisting of a 50%-50% mix of DMEM-F12 (Gibco, 10565-018) and Neurobasal TM (Gibco, 21103049). Doubling time calculations were done with the following formula: Doubling Time=24*(Tz-Ty)/LN(COUNTz/SEEDy); where Tz=date and time of harvest, Ty=date and time of seeding, COUNTz=total cell count at time of harvest, SEEDy=number of cells seeded.
Media for differentiation and maintenance of EpiLSCs was the same as the 2i media with the following changes: Lif, Mek, and Gsk3 inhibitors were removed. Activin-A and Fgf2 were added to 20ng/ml and 12ng/ml respectively. All EpiLSC differentiations were performed on irradiated mouse embryonic fibroblasts. Embryoid bodies were formed by plating ES cells on non-tissue culture treated plastic in the serum media without Lif.

**Derivation of seqKO ES cells**

In brief, the Chd1 targeting cassette plasmid was obtained from the KOMP consortium (clone PG134G06), linearized, and transfected into low passage E14 mES cells using Fugene (Promega). Cells were subjected to selection with neomycin, screened for LacZ expression, and clonal lines derived. Proper integration was verified by sequencing multiple long range genotyping PCR reactions. Cells were then infected with a Cre/GFP adenovirus, subjected to FACS for GFP positive cells, selected for sensitivity to neomycin, and clonal lines derived. Proper excision was verified by sequencing multiple genotyping reactions. Targeting and excision of the second allele was accomplished in a similar manner except without the added benefit of screening for LacZ expression.

**Immuno-fluorescence**

Staining was performed according to standard lab procedures. In brief, adherent cells were fixed with 4% PFA for 10min at room temperature and washed 3 times with PBS. Permeabization was achieved with 0.25% Triton-X100 treatment for 10min at room temperature followed by 3 PBS washes. Cells were blocked for 1hr at room temperate with 10% Donkey serum in PBS-T before exposure to a primary antibody (αOct4, Santa Cruz 9081; αChd1, R&D Systems AF619; αNanog, Abcam ab80892; αMyc-tag, Abcam ab9132; αH3K27me3, Millipore 07-449; αH2A.X, Millipore 05-636; αγH2A.X, Millipore 05-636) using a 1:200 dilution in blocking buffer at 4°C overnight. Cells were then washed with PBS and treated
with a secondary antibody (1:500 dilution) for 1hr at room temperature, washed with PBS, and mounted with a cover slide using VectorShield with DAPI.

Annexin-V and PI staining for FACS were done according to manufacturer’s protocol (BD Bioscience 556570).

Western

2 color westerns were performed according to manufacturer’s instructions (LiCor). Samples were prepared using standard RIPA buffer. The soluble fraction was isolated with centrifugation and loaded onto a denaturing gradient SDS acrylamide gel. Transfer to a membrane was accomplished at 30V overnight at 4⁰C to a PVDF membrane (Millipore). Primary antibodies used were αChd1, Santa Cruz 49813; αTopol, BioLegend 605801; αH2B, Abcam ab1790; αH2Bub, Cell Signaling 5S46. 2-color blots were false colored to black and white for densitometry analysis using manufacturer’s software (LiCor).

qRT-PCR

qRT-PCR was performed as described in Appendix III. Primers are listed in Appendix I.

Microarray analysis

Microarray expression analysis of the Chd1 seqKO ES cells was done using Affymatrix expression array GeneChip U133 Plus 2.0. Triplicate samples were processed for wild type, Chd1+/−, and Chd1−/− samples. Relative enrichments were determined using D-Chip software and expression plots generated using R.

ChIP-qPCR

ChIP-qPCR was performed as described in Appendix II using 200k cells per sample. Primers used are listed in Appendix I. The followed primary antibodies were used: H3K4me3, Diagenode pAb-003-050; panH3, Abcam ab1791; H3K27me3, Millipore 073-449; SSrRNA-Pol II, Covance MMS-134R; S2rRNA-Pol II, Abcam ab5095; rabbit IgG, Abcam ab46540.
ChIP-seq Analysis

ChIP for H3K4me3, H3K27me3, and input controls was performed on 200K cell samples of either tamoxifen or control treated Chd1 iKO ES cells grown in serum media according to the protocol provided in appendix II. Libraries were generated using the ThruPLEX-FD prep kit (Rubicon Genomics) according to manufacturer’s instructions and sequenced on 2 lanes of an Illumina Rapid Run flow cell. Reads from each ChIP-Seq library were filtered to retain only unique sequences and aligned to the mm9/NCBI build 37 mouse genome using bowtie. Reads mapping to more than 1 genomic location were assigned coordinates to one of those locations randomly. UCSC tracks were generated uses MACS2 and normalized for total reads. Regions of enrichment and depletion were determined by comparing the H3K4me3 KO library to the H3K4me3 control library and the peak calling function of MACS2 with a confidence threshold of q=0.05. The H3K27me3 data was processed identically.

All gene set analysis is based on the mm9 RefSeq gene set. A gene was considered enriched or depleted for a histone mark if a region of enrichment/depletion was within 10k of the gene, defined as the coordinated starting at the TSS and ending at the TES. Analysis of enrichment for Gene Ontology terms was done using DAVID (Huang et al., 2009).
Figure 2.S1. Generation of Chd1 KO ES cells. (A) Cartoon schematic of known protein domains in the Chd1 amino acid sequence. Exon 16 is removed to create a catalytically inactive protein and introduce a frame-shift with abortive stop codons. (B) Schematic of targeting cassette used to generate Chd1 seqKO and iKO ES cells. (C) Western blot validating the loss of full length protein in Chd1 seqKO ES cells.
Figure 2.52. Directed Differentiation to EpLSCs is not affected by loss of Chd1. (A) Transcription was measured with qRT-PCR for the indicated genes in 2 Chd1 iKO ES cell lines during the first 24hrs of EpLSC differentiation. Data represented as a percent of the housekeeping gene L7 ± standard deviation. (B) qRT-PCR for the indicated genes in 2 Chd1 seqKO and control ES cell lines during the first 5 days of EpLSC differentiation. Data represented as percent of indicated gene at t=0 ± standard deviation. (C) Same as in (A) for the first 19 days of EpLSC differentiation. Cultures were passaged with collagenase as necessary. (D) Immuno-fluorescence staining for a pluripotency marker, Nanog, in Chd1 iKO EpLSCs 19 days post differentiation either treated with or without tamoxifen.
**Figure 2.53.** Chd1 KO ES cells can self-renew in chemically defined medium. (A) Colony formation assay of Chd1 seqKO ES cells in 2i and KSR media as indicated. (B) Doubling time of Chd1 seqKO ES cells grow in 2i. (C) Doubling time of Chd1 seqKO ES cells grow in KSR.
Figure 2.54. H3.3 incorporation is not affected in Chd1 KO ES cells. (A) Immuno-fluorescence staining for transgenic tagged H3.3 and H3K27me3 in control and Chd1 seqKO ES cells. 40x objective. (B) ChIP-qPCR for transgenic tagged H3.3 at indicated genes represented as mean percent of input ± standard deviation for 3 biological replicates.
Figure 2.55. DNA damage response is not affected in Chd1 KO ES cells. (A) Western blot for H2Bub in 2 Chd1 iKO ES cell lines. (B) Immuno-fluorescence staining for H2A.X and gammaH2A.X in Chd1 iKO ES cells. 63x objective.
Figure 2.6. Validation of H3K4me3 ChIP-seq in Chd1 KO ES cells. (A) Representative tracks for H3K4me3 and H3K27me3 at genes unaffected by loss of Chd1 for these two marks. (B) ChIP-qPCR for indicated genes represented as mean percent of input ± standard deviation for 2 biological replicates of Chd1 iKO ES cells. (B) Data in (B) normalized for enrichment of H3 (data not shown).
Figure 2.57. Cell number normalized expression in Chd1 KO ES cells. (A) Microarray depicting minimal changes in gene expression between Chd1 seqKO ES cells and Chd1 +/- control ES cells grown in serum based media. Data normalized for total RNA using standard methods. (B) Cell number normalized qRT-PCR for indicated genes for a biologically independent Chd1 iKO ES cell line. Data represented as percent gene in control sample ± standard deviation of 2 cell number normalized samples. (C) Same as in (C) for a third biologically independent Chd1 iKO cell line.
<table>
<thead>
<tr>
<th>description</th>
<th>genotype</th>
<th>number of lines derived and validated</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sequential Chd1 Knock-Out</strong></td>
<td>Chd1+/Δ</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Chd1Δ/Δ</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Chd1+/Δ ; Tg(H3.3/myc)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Chd1Δ/Δ ; Tg(H3.3/myc)</td>
<td>2</td>
</tr>
<tr>
<td><strong>Embryonic Chd1 Knock-Out</strong></td>
<td>Chd1+/Δ</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Chd1Δ/Δ</td>
<td>2</td>
</tr>
<tr>
<td><strong>Inducible Chd1 Knock-Out</strong></td>
<td>Chd1+/flox ; R26tm1(CreER)</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Chd1Δ/flox ; R26tm1(CreER)</td>
<td>3</td>
</tr>
<tr>
<td><strong>Endogenous tagged Chd1 Knock-In</strong></td>
<td>Chd1tm1(Chd1/Flag)</td>
<td>1</td>
</tr>
<tr>
<td><strong>Endogenous tagged Chd1 Knock-In with tagged H3.3 transgene</strong></td>
<td>Chd1tm1(Chd1/Flag) ; Tg(H3.1/myc)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Chd1tm1(Chd1/Flag) ; Tg(H3.3/myc)</td>
<td>2</td>
</tr>
<tr>
<td><strong>Chd1 Knock-Down in Endogenous tagged H3.3A</strong></td>
<td>Tg(RNAi:GFP) ; H3.3Atm1(H3.3A/HA)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Tg(RNAi:Chd1) ; H3.3Atm1(H3.3A/HA)</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 2.S1. Mouse ES cell lines generated in this study.
<table>
<thead>
<tr>
<th>q-value (-log10)</th>
<th>refseq accession</th>
<th>gene symbol</th>
<th>q-value (-log10)</th>
<th>refseq accession</th>
<th>gene symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>21.0</td>
<td>NM_001081105</td>
<td>1S00012F01Rik</td>
<td>4.0</td>
<td>NR_024067</td>
<td>Snhg6</td>
</tr>
<tr>
<td>16.7</td>
<td>NM_013762</td>
<td>Rpl3</td>
<td>3.9</td>
<td>NM_011508</td>
<td>Elf1</td>
</tr>
<tr>
<td>10.9</td>
<td>NM_029767</td>
<td>Rps9</td>
<td>3.8</td>
<td>NM_009391</td>
<td>Ran</td>
</tr>
<tr>
<td>10.6</td>
<td>NM_013506</td>
<td>Eif4a2</td>
<td>3.8</td>
<td>NM_007970</td>
<td>Calm1</td>
</tr>
<tr>
<td>10.3</td>
<td>NR_002896</td>
<td>Snhg1</td>
<td>3.8</td>
<td>NM_172881</td>
<td>Ugt2b35</td>
</tr>
<tr>
<td>10.2</td>
<td>NM_011562</td>
<td>Tdgf1</td>
<td>3.8</td>
<td>NM_133721</td>
<td>Itga9</td>
</tr>
<tr>
<td>9.8</td>
<td>NR_028401</td>
<td>Tafl1</td>
<td>3.7</td>
<td>NM_001163478</td>
<td>Rabggb</td>
</tr>
<tr>
<td>9.8</td>
<td>NM_007438</td>
<td>Aldoa</td>
<td>3.6</td>
<td>NR_003270</td>
<td>Snhg3</td>
</tr>
<tr>
<td>8.6</td>
<td>NR_038073</td>
<td>Snhg4</td>
<td>3.5</td>
<td>NM_011665</td>
<td>Tubb5</td>
</tr>
<tr>
<td>8.6</td>
<td>NM_008722</td>
<td>Npm1</td>
<td>3.5</td>
<td>NM_016809</td>
<td>Rbm3</td>
</tr>
<tr>
<td>8.6</td>
<td>NM_008084</td>
<td>Gapdh</td>
<td>3.5</td>
<td>NM_153530</td>
<td>Dis3l2</td>
</tr>
<tr>
<td>8.3</td>
<td>NR_002840</td>
<td>Gaa5</td>
<td>3.4</td>
<td>NR_029468</td>
<td>Snhg12</td>
</tr>
<tr>
<td>8.1</td>
<td>NR_045824</td>
<td>1700016D04Rik</td>
<td>3.2</td>
<td>NR_034038</td>
<td>Sa30416N02Rik</td>
</tr>
<tr>
<td>8.1</td>
<td>NM_016806</td>
<td>Hnnpa2b1</td>
<td>3.2</td>
<td>NM_001029837</td>
<td>Pik3cd</td>
</tr>
<tr>
<td>7.7</td>
<td>NM_015781</td>
<td>Nap1l1</td>
<td>3.1</td>
<td>NM_001130184</td>
<td>Phf17</td>
</tr>
<tr>
<td>7.5</td>
<td>NR_040296</td>
<td>Gm19434</td>
<td>3.1</td>
<td>NM_025579</td>
<td>Tafl2</td>
</tr>
<tr>
<td>6.8</td>
<td>NM_011029</td>
<td>Rpsa</td>
<td>3.0</td>
<td>NM_026030</td>
<td>Elf2s2</td>
</tr>
<tr>
<td>6.8</td>
<td>NR_045893</td>
<td>Gm11974</td>
<td>3.0</td>
<td>NR_000002</td>
<td>Snord32a</td>
</tr>
<tr>
<td>6.8</td>
<td>NM_009609</td>
<td>Actg1</td>
<td>3.0</td>
<td>NM_020600</td>
<td>Rps14</td>
</tr>
<tr>
<td>6.7</td>
<td>NM_011975</td>
<td>Rpl27a</td>
<td>2.9</td>
<td>NM_001100452</td>
<td>Gltscr1l</td>
</tr>
<tr>
<td>6.6</td>
<td>NM_001159483</td>
<td>Rpl19</td>
<td>2.9</td>
<td>NM_010106</td>
<td>Eef1a1</td>
</tr>
<tr>
<td>6.4</td>
<td>NR_029736</td>
<td>Mir18</td>
<td>2.8</td>
<td>NR_029643</td>
<td>Mir293</td>
</tr>
<tr>
<td>6.3</td>
<td>NM_009091</td>
<td>Rps15</td>
<td>2.8</td>
<td>NM_026055</td>
<td>Rpl39</td>
</tr>
<tr>
<td>6.1</td>
<td>NM_052835</td>
<td>Rpl10</td>
<td>2.7</td>
<td>NM_031165</td>
<td>Hspal8</td>
</tr>
<tr>
<td>6.1</td>
<td>NM_001025388</td>
<td>Gm5506</td>
<td>2.6</td>
<td>NM_025274</td>
<td>Dppa5a</td>
</tr>
<tr>
<td>6.1</td>
<td>NR_007078</td>
<td>2410006H16Rik</td>
<td>2.6</td>
<td>NM_025586</td>
<td>Rpl15</td>
</tr>
<tr>
<td>5.9</td>
<td>NM_001039129</td>
<td>Hnnpa1</td>
<td>2.4</td>
<td>NM_001030209</td>
<td>Gm13152</td>
</tr>
<tr>
<td>5.9</td>
<td>NM_176902</td>
<td>Uballd2</td>
<td>2.3</td>
<td>NM_016844</td>
<td>Rps28</td>
</tr>
<tr>
<td>5.9</td>
<td>NM_024212</td>
<td>Rpl4</td>
<td>2.3</td>
<td>NR_040322</td>
<td>4930500G02Rik</td>
</tr>
<tr>
<td>5.7</td>
<td>NM_026147</td>
<td>Rps20</td>
<td>2.3</td>
<td>NM_012053</td>
<td>Rpl8</td>
</tr>
<tr>
<td>5.6</td>
<td>NM_009098</td>
<td>Rps8</td>
<td>2.3</td>
<td>NM_008503</td>
<td>Rps2</td>
</tr>
<tr>
<td>5.6</td>
<td>NM_009448</td>
<td>Tuba1c</td>
<td>2.1</td>
<td>NM_174346</td>
<td>Scl25a12</td>
</tr>
<tr>
<td>5.1</td>
<td>NM_008774</td>
<td>Palp1c</td>
<td>2.0</td>
<td>NM_008302</td>
<td>Hsp90ab1</td>
</tr>
<tr>
<td>5.0</td>
<td>NR_040721</td>
<td>Snhg5</td>
<td>1.9</td>
<td>NM_025974</td>
<td>Rpl14</td>
</tr>
<tr>
<td>4.9</td>
<td>NM_011287</td>
<td>Rpl10a</td>
<td>1.9</td>
<td>NM_012052</td>
<td>Rps3</td>
</tr>
<tr>
<td>4.6</td>
<td>NM_011290</td>
<td>Rpl6</td>
<td>1.9</td>
<td>NM_023065</td>
<td>Ifi30</td>
</tr>
<tr>
<td>4.6</td>
<td>NM_008452</td>
<td>Klf2</td>
<td>1.6</td>
<td>NM_011099</td>
<td>Pkm</td>
</tr>
<tr>
<td>4.4</td>
<td>NM_011291</td>
<td>Rpl7</td>
<td>1.6</td>
<td>NM_172768</td>
<td>Gramd1b</td>
</tr>
<tr>
<td>4.4</td>
<td>NM_010655</td>
<td>Kpn2</td>
<td>1.5</td>
<td>NM_009077</td>
<td>Rpl18</td>
</tr>
<tr>
<td>4.4</td>
<td>NM_011295</td>
<td>Rps12</td>
<td>1.5</td>
<td>NM_013512</td>
<td>Ebp4.14a</td>
</tr>
<tr>
<td>4.3</td>
<td>NM_009076</td>
<td>Rpl12</td>
<td>1.4</td>
<td>NM_008143</td>
<td>Gnb2l1</td>
</tr>
<tr>
<td>4.2</td>
<td>NM_029751</td>
<td>Rpl18a</td>
<td>1.3</td>
<td>NR_030703</td>
<td>Snord104</td>
</tr>
<tr>
<td>4.2</td>
<td>NM_008972</td>
<td>Ptma</td>
<td>1.3</td>
<td>NM_026509</td>
<td>Murc</td>
</tr>
<tr>
<td>4.0</td>
<td>NM_007907</td>
<td>Eef2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.52. Genes with reduced H3K4me3 in Chd1 KO ES cells. Ranked by probability of reduction, which is calculated based on total number of reads and relative reduction of reads found in the chd1 null ES cells.
References


**Summary**

Developmental regulatory genes have both activating (H3K4me3) and repressive (H3K27me3) histone modifications in embryonic stem cells (ESCs). This bivalent configuration is thought to maintain lineage commitment programs in a poised state. However, establishing physiological relevance has been complicated by the high number of cells required for chromatin-immunoprecipitation (ChIP). We developed a low cell ChIP protocol to investigate the chromatin of mouse primordial germ cells (PGCs). Genome-wide analysis of E11.5 PGCs revealed H3K4me3/H3K27me3 bivalent domains highly enriched at developmental regulatory genes in a manner remarkably similar to ESCs. Developmental regulators remain bivalent and transcriptionally silent through the initiation of sexual differentiation at E13.5. We also identified >2500 “orphan” bivalent domains distal to known genes and expressed in a tissue-specific manner, but silent in PGCs. Our results demonstrate the existence of bivalent domains in the germline and raise the possibility that the somatic program is continuously maintained as bivalent, potentially imparting transgenerational epigenetic inheritance.
**Introduction**

Pluripotency is dependent on the maintenance of a proper epigenetic landscape (Gaspar-Maia et al., 2011; Orkin and Hochedlinger, 2011). Bivalent domains, which are defined by the paradoxical coexistence of a permissive histone mark (H3K4me3) and a repressive mark (H3K27me3), are thought to play an important role in pluripotency by keeping developmental genes in a silenced state poised for activation upon differentiation (Azuara et al., 2006; Bernstein et al., 2006). In support of this model, bivalent domains have been shown to be present preferentially at developmental regulatory genes in undifferentiated embryonic stem cells (ESCs) and several adult tissues in vivo including sperm, testis, the cerebellum, and the hematopoietic compartment (Mikkelsen et al., 2007; Cui et al., 2009, 2012; Hammoud et al., 2009). The moderate de-repression of developmental regulators in eed mutant ESCs, which have reduced levels of H3K27me3, further supports this model (Azuara et al., 2006; Boyer et al., 2006). However, it has recently been proposed that the concomitant presence of H3K4me3 and H3K27me3 at developmental genes in ESCs is an in vitro artifact resulting from suboptimal culture conditions (Hong et al., 2011; Marks et al., 2012). In addition, direct evidence for bivalency in the developing mammalian embryo is limited and conflicting due to technical difficulties associated with the low amounts of material available (Alder et al., 2010; Dahl et al., 2010; Rugg-Gunn et al., 2010). The universal nature of bivalency has been further questioned due to conflicting reports from non-mammalian species, with bivalent domains demonstrated to exist in zebrafish (Vastenhouw et al., 2010) but not detected in Xenopus or Drosophila embryos (Akkers et al., 2009; Schuettengruber et al., 2009). We therefore lack a clear understanding of whether bivalency exists in embryonic cells in vivo, and how it relates to pluripotency.

Of particular interest to pluripotency are primordial germ cells (PGCs), the embryonic precursors of the germline. PGCs have a transcriptional profile similar to ESCs (Qin et al., 2012), including the silencing of
developmental regulators, and can give rise to pluripotent stem cells when cultured in vitro. However, unlike ESCs, PGCs are unipotent and unable to contribute to chimeras. Additionally, PGCs are thought to undergo a period of ‘epigenetic erasure’ or reprogramming via global removal of both DNA methylation and histone modifications that may prevent transmission of epimutations (Greer and Shi, 2012; Seisenberger et al., 2013). Due to the low number of PGCs present during development, nearly all evidence for this model is from immunofluorescence (IF)-based methodologies that lack gene level resolution and have shown conflicting results (Seki et al., 2007; Hajkova et al., 2008; Kagiwada et al., 2012). To characterize the chromatin state of PGCs, we developed a low cell number Chromatin Immunoprecipitation (ChIP) for the analysis of histone marks using less than 10,000 cells per IP without the need for carrier chromatin or pre-amplification. Using this technique, we performed ChIP-Seq and ChIP-qPCR to show that bivalent domains are present at developmental regulatory genes at multiple stages of PGC development.

Results

Low cell number ChIP protocol

We developed a low cell number ChIP protocol using a magnetic bead-based capture with key improvements over existing approaches (O’Neill et al., 2006; Acevedo et al., 2007; Dahl and Collas, 2008; Adli and Bernstein, 2011; Shankaranarayanan et al., 2011), including reduced handling of material for cross-linking, improvements to sonication reproducibility, and no requirement for carrier DNA or DNA pre-amplification (see Methods). We validated the protocol using a transgenic mouse ESC line expressing an N-terminal myc-tagged H3.3 histone, due to technical simplicity and the availability of genome-wide data (Goldberg et al., 2010). ChIP was performed for myc-H3.3 and a control IgG with decreasing amounts of chromatin starting from 2.5 million to 12,500 cells per IP (Figure 3. S1A). qPCR
was used to interrogate the enrichment status of 15 different loci representing 6 categories of genomic regions. Importantly, 11 of the 12 loci expected to be bound by H3.3 (Goldberg et al., 2010) were at least 10 fold more enriched for myc-H3.3 compared to IgG, demonstrating the specificity and low background of the assay. Additionally, the relative enrichment between loci remains stable independent of starting cell number, demonstrating the robustness and sensitivity of the assay (Figure 3. S1B). We envision that this protocol will be broadly useful to characterize the chromatin state of rare stem and precursor cell populations in vivo.

**Genome-wide analysis of H3K4me3 and H3K27me3 in PGCs**

To investigate the presence of bivalent domains in PGCs, we used a transgenic *Oct4:GFP* mouse line to isolate PGCs (GFP+) and the surrounding somatic cells (GFP-) by FACS at different stages of development as previously reported (Yoshimizu et al., 1999). Using our low cell ChIP protocol, we performed ChIP-Seq for H3K4me3 and H3K27me3 in two biological replicates of E11.5 PGCs, obtaining a total of approximately 386 million reads and between 15 and 32 million uniquely mappable reads post filtering per ChIP library (Table S1). The two biological replicates in our ChIP-Seq libraries show very strong correlation, with Pearson correlation coefficients of 0.978 for H3K4me3 and 0.996 for H3K27me3 (Figure 3. 1A, S2A, Table S1), indicative of the high reproducibility of the protocol. Furthermore, all IPs are enriched for signal over input as assessed using CHANCE (Table S1) (Diaz et al., 2012).

We identified H3K4me3-only regions, H3K27me3-only regions, and bivalent domains in both E11.5 PGCs (this study) and ESCs (Mikkelsen et al., 2007) using ChromHMM (Ernst and Kellis, 2012). The overall distribution of these histone marks is similar between PGCs and ESCs, except for an increase in H3K27me3-marked DNA in PGCs (Figure 3. S2B), which is in agreement with IF data (Seki et al., 2005, 2007; Kagiwada et al., 2012). Genes marked only with H3K27me3 in PGCs are involved in cell motility (Figure 3. S2C), which may be a consequence of PGCs having just completed their migration to the gonads at E11.5.
Bivalent genes in PGCs are enriched for developmental regulators

We next focused on bivalent regions. Consistent with their association with gene promoters, bivalent domains largely occur at transcriptional start sites (TSSs) in both E11.5 PGCs and ESCs (Figure 3. S2D). E11.5 PGCs and ESCs have a similar distribution and a highly significant overlap of genes marked as bivalent \( p<2.2\times10^{-16}, \) Figure 3. 1B, 1C, S2E). Similarly to what has been described for ESCs (Bernstein et al., 2006), the 2715 genes marked as bivalent in PGCs and ESCs are strongly enriched for regulators of early development (Figure 3. 1D). We note that while this manuscript was under review, Ng et al. reported a ChIP-Seq dataset for several histone marks in PGCs, including H3K4me3 and H3K27me3 (Ng et al., 2013). However, the authors did not analyze bivalency in detail and the low signal for H3K4me3 in E11.5 PGCs in that study precludes a direct comparison with our work (Figure 3. S2A). Therefore, we sought to validate our data using ChIP-qPCR in independent biological samples of PGCs. ChIP-qPCR at 12 genes representing 8 developmental pathways confirmed that all are bivalently marked in PGCs, with bivalency defined as enrichment above a background of 10% for both H3K4me3 and H3K27me3 and a maximum 2-fold difference in enrichment between the two marks (Figure 3. 1E). These bivalent regulators include transcription factors that are important for the specification of the 3 somatic germ layers, such as Pax6, Nkx2.2, MyoD, brachyury \( (T) \), Gata6 and FoxA1, and Hox genes such as HoxA3 and HoxB9. This is notable because ESCs are pluripotent and expected to maintain these genes in a poised transcriptional state for activation upon differentiation, but PGCs are unipotent and will not express any of these developmental regulators until post fertilization, in the next generation. In fact, the Hox cluster has been shown to be silenced as part of the early specification of PGC precursors at E7.25 (Saitou et al., 2002). A comparison of the expression pattern of genes bivalent in PGCs and ESCs demonstrates that these genes are transcriptionally repressed in ESCs and PGCs, but subsets are active in select lineage-restricted cell types such as MEFs, bone marrow, or brain tissue (Figure 3. 1F). Taken together, our genome-wide analysis reveals that the mouse embryonic germline \textit{in vivo} is remarkably similar to
Figure 3.1. Bivalent genes in E11.5 PGCs are enriched for developmental regulators. (A) Sample UCSC Genome Browser view of ChIP-Seq signal for 2 biological replicates of E11.5 PGCs and for ESCs. (B) E11.5 PGCs and ESCs have a similar number of genes marked by only H3K4me3, only H3K27me3, or both. (C) Overlap of genes marked by H3K4me3 and H3K27me3 in E11.5 PGCs and ESCs, p<2.2x10^-16 using Fisher’s exact test. (D) Top 5 biological process gene ontology terms as determined using DAVID for genes marked as bivalent in both PGCs and ESCs. (E) Validation of ChIP-Seq using ChIP-qPCR enrichment for H3K4me3 and H3K27me3 in E11.5 PGCs for a subset of bivalent genes representing regulators of the three germ layers. Data are mean percent of input ± standard deviation of enrichment. The dotted line represents level of background enrichment. (F) Heat map showing expression of the subset of genes marked as bivalent in both PGCs and ESCs in indicated cell types. EBs: embryonic bodies; MEFs: mouse embryonic fibroblasts; L.V. Brain: lateral ventricles of adult brain. Blue indicates down-regulation, red indicates up-regulation.
cultured ESCs with regards to the presence of bivalent chromatin at silenced developmental regulatory genes.

**PGCs are enriched for bivalency at developmental genes compared to the surrounding soma**

We next compared the chromatin landscape of PGCs to the surrounding soma. Using ChIP-qPCR, we examined the 12 previously tested developmental regulators plus Pou5f1/Oct4 in ESCs and E11.5 soma (Figure 3. 2A, 2B). 12/12 of these developmental regulators are bivalent in both E11.5 PGCs and ESCs. However, in the soma most of these genes have stronger enrichment for either H3K4me3 or H3K27me3 and only four genes are bivalent (Figure 3. 2B, 2C). These data indicate that developmental regulators of all somatic lineages exist in a bivalent state in ESCs and PGCs but not in somatic cells in vivo (Figure 3. 2D), suggesting that this distribution of bivalent domains is a distinguishing characteristic of pluripotency-associated cells both in vitro and in vivo.

**Developmental regulators remain bivalent in PGCs during the initiation of sexual differentiation**

E11.5 PGCs are sexually indifferent and capable of giving rise to pluripotent stem cells when cultured in vitro. After E11.5, PGCs initiate the process of sexual differentiation and express sex-specific genes by E13.5. Furthermore, E13.5 PGCs can no longer give rise to pluripotent stem cells (Labosky et al., 1994). We therefore sought to investigate whether the process of sexual differentiation and concomitant loss of the ability to give rise to pluripotent stem cell cells affects the bivalent chromatin state of PGCs. Using low cell ChIP-qPCR, we examined 8 representative developmental genes through a time-course of PGC development from E11.5 to E13.5. In agreement with previously published data (Qin et al., 2012), somatic developmental genes, including HoxA11, HoxB9, Pax6, Nkx2.2, T, MyoD, Gata6, and FoxA1 are expressed at very low to undetectable levels from E11.5 through E13.5 (Figure 3. 3). Interestingly, these
Figure 3.2. PGCs are enriched for bivalent developmental regulators compared to surrounding somatic cells at E11.5. (A) ChIP enrichment for H3K4me3 and H3K27me3 in ESCs at the indicated genes. Data are mean percent of input ± standard deviation of enrichment. The dotted line represents level of background enrichment. (B) ChIP enrichment for H3K4me3 and H3K27me3 in E11.5 soma at the indicated genes as described in (A). (C) The ratio of H3K4me3 to H3K27me3 log2 transformed for the indicated gene for ESCs, E11.5 PGCs, and E11.5 soma. The gray area indicates a ratio less than 2 and greater than 0.5. Genes where enrichment for either H3K4me3 or H3K27me3 is not above background are lighter in color. (D) The log2 of the H3K4me3/H3K27me3 ratio was calculated as in (C) (log2 $R_{cell\,type}$) for all genes examined independent of enrichment over background. The differences between cell type ratios were calculated as follows: $\Delta_{PGC} = \log2 R_{PGC} - \log2 R_{ESC}$, and $\Delta_{soma} = \log2 R_{soma} - \log2 R_{ESC}$ for each gene. Box-plots represent the distribution of the differences. Statistical significance was assessed using the Wilcoxon matched-pairs signed rank test (1-sided), **$p=0.001$. ND= not determined.
developmental regulators remain bivalent in PGCs throughout this time-period (Figure 3.3). In contrast, germline-specific genes, such as Dazl, Dppa3, and Vasa are enriched for only H3K4me3, coinciding with their transcriptional activation (Figure 3. S3). Taken together, these data indicate that somatic developmental regulators remain bivalent in the mouse embryonic germline from E11.5 through initiation of sexual differentiation at E13.5.

**Orphan bivalent domains are silent in PGCs and transcribed in a tissue-specific manner**

Surprisingly, we identified a large number of ‘orphan’ bivalent domains that cannot be associated with any annotated promoter or gene in the RefSeq database, and are distal to annotated genes by a median distance of 13kb (Figure 3. 4A). From a total of 9132 bivalent domains in E11.5 PGCs, 2886 are orphan, and many are also bivalent in ESCs (Table S2). Bivalent domains are known to have a strong association with unmethylated CpG islands (CGIs) (Ku et al., 2008). We found that 1413 (49.0%, p-value < 1x10^{-5}) of the orphan bivalent domains in PGCs map to experimentally defined unmethylated CGIs (Illingworth et al., 2010). We next investigated the expression status of the regions containing orphan bivalent domains. If the orphan bivalent domains were similar to bivalent genes (Figure 3. 1F), they would be silent in PGCs and expressed in somatic tissues. Indeed, only 23 (0.80%) orphan bivalent domains show signs of transcription in an RNA-Seq dataset from E11.5 PGCs (Seisenberger et al., 2012). Analysis of CAGE tag clusters showed that 927 of the 1413 (65.6%, p-value = 5.78x10^{-168}) orphan bivalent domains with CGIs are expressed in at least 1 of 22 tissue types (Figure 3. 4B, Table S3) (Kawaji et al., 2009). Similarly, 461 of the 1473 (31.3%, p-value = 3.61x10^{-5}) orphan bivalent domains without CGIs are expressed in at least 1 of 22 tissue types (Figure 3. 4B, Table S3). Furthermore, both classes of orphan bivalent domains (with or without CGI) display highly tissue-specific expression (Figure 3. 4C), more so than bivalent developmental regulators (p-value = 1.1x10^{-141}, 2.3x10^{-72} respectively), with a preferential activation in brain regions (Figure 3. 4C, E). Interestingly, we found bivalent domains in PGCs at 300 non-
Figure 3.3. Developmental regulators remain bivalent and transcriptionally repressed during sexual differentiation in the germline. Developmental genes remain bivalent and transcriptionally silent from E11.5 to E13.5 in both the male and female germline. ChIP enrichment for H3K4me3 and H3K27me3 in 11.5-13.5 PGCs at the indicated genes is represented as the mean percent of input ± standard deviation for at least 2 biological replicates. Transcription was measured with qRT-PCR for the indicated genes and represented as a percent of the housekeeping gene L7 ± standard deviation. At all stages of PGC development analyzed, somatic developmental genes have a similar level of enrichment for H3K4me3 and H3K27me3 and very low expression.
Figure 3.4. Orphan bivalent domains overlap CpG Islands and show tissue specific expression. (A) Distances of genic (assigned to gene, dark grey bars) and orphan PGC bivalent domains (white bars) to nearest gene, with
transparent overlay. Distances are presented as the log base 10 of the genomic length in base pairs from bivalent domains to either their nearest gene, with bivalent domains upstream of their nearest gene assigned negative distances after logarithm. Bivalent domains occurring within a gene were assigned a genomic length of 1. TSSs were defined as the starting positions of RefSeq genes (B) Number of orphan bivalent domains in E11.5 PGCs with characteristics of transcription start site activity: occurring in unmethylated CpG islands (CGI) identified using CAP-Seq (Illingworth et al., 2010), with a CAGE expression tag in 1 or more of 22 tissue types (Kawai et al., 2009), or both. (C) Distributions of the tissue-specificity scores of CAGE tag cluster expression in orphan bivalent domains with CGI in PGC, orphan bivalent domains without CGI in PGC, bivalent RefSeq genes, and all RefSeq genes. The p-values are from the Wilcoxon rank sum test. (D) UCSC Genome Browser view of an orphan bivalent domain in PGCs and ESCs (Mikkelsen et al., 2007) that overlaps a CGI (Illingworth et al., 2010) and a putative non-coding RNA (Guttman et al., 2009). RNA-Seq data indicates this region is transcriptionally silent in PGCs (Seisenberger et al., 2012) and ESCs (Marks et al., 2012), while CAGE tag data indicates expression in the hippocampus. (E) Hierarchical clustering of tissues based on the binary (on or off) expression status of CAGE tag clusters in regions marked as orphan bivalent domains in PGC. Yellow means expression and blue mean no expression. See also Table S2, S3.

coding RefSeq genes and 179 putative non-coding RNAs inferred from chromatin state (Guttman et al., 2009) and expression data(Carninci et al., 2005) (Figure 3. 4D, S2F). These data suggest that, similar to protein-coding developmental regulators, some non-coding RNAs are bivalent in the germline and poised for activation upon lineage commitment.

Discussion

In this work we describe an optimized ChIP protocol suitable for low numbers of cells and its use to examine the genome-wide and gene-specific distribution of H3K4me3 and H3K27me3 in mouse PGCs. We report that developmental regulatory genes remain bivalent and transcriptionally silent in vivo in PGCs, but not in adjacent somatic cells, throughout E11.5-E13.5, in a manner highly similar to cultured ESCs. In addition to developmental genes, we identify some ~3000 orphan bivalent domains that are enriched for CGIs and are expressed in a tissue-specific manner. Some of the bivalent domains identified here correspond to non-coding RNAs that we speculate may have developmental functions, similar to bivalent genes, although this remains to be determined. The findings presented here represent strong evidence for the existence of bivalent domains in the embryonic germline, and raise a number of important questions.
Do bivalent promoters escape the epigenetic reprogramming reported to occur in PGCs from E8.5 to E13.5? Our data support the suggestion that H3K27me3 may act as a potential mechanism to compensate for the loss of DNA methylation or H3K9me2/3 in mid-gestation PGCs (Sasaki and Matsui, 2008). It is also possible that the loss of certain histone marks observed by IF may occur primarily at abundant repetitive sequences of the genome and mostly spare unique genes, such as developmental regulators. The application of the low cell ChIP protocol reported here should help answer this question for other histone marks and time-points.

Does the germline continuously maintain the somatic program in a bivalent state until activation in the next generation? It is unclear why such a large set of important developmental regulators would remain bivalent in the developing germline, but not in somatic cells. The hypothesis that bivalency maintains developmental genes in a transcriptionally poised state is intuitive in pluripotent cells that activate these genes upon differentiation, such as ESCs (Azuara et al., 2006; Bernstein et al., 2006) or the epiblast (Rugg-Gunn et al., 2010). The presence of bivalent domains in PGCs may contribute to preventing expression of the somatic program in the germline. Our observations suggest that developmental regulators may be kept in a repressed but accessible state in the germline for activation post-fertilization, in the next generation. The transmission of bivalency through the germline could provide a substratum for epigenetic inheritance. Surprisingly, recent work indicates that the sperm genome maintains a residual level of nucleosomes, and that these are enriched for H3K4me3/H3K27me3 bivalent marks at developmental regulators (Hammoud et al., 2009). It will be interesting to determine whether bivalency is detected at other stages of germline development and in oocytes. Functional studies of regulators of bivalency in PGCs should provide important insights into these questions.
Experimental Procedures

Isolation of PGCs

Male B6 mice homozygous for a transgenic Oct4ΔPE:GFP reporter were crossed with Swiss-Webster females. The gonadal regions from multiple embryos were isolated and pooled prior to enzymatic dissociation for FACS.

Low Cell ChIP-qPCR

Cells were pooled into batches of ~50,000 cells and cross-linked in 1% formaldehyde (Sigma F8775-25ml) in PBS for 5min at room temperature and quenched with 125mM glycine. For each IP, 11µl of protein A Dynabeads (Life Technologies 1001D) were pre-incubated with 2.4µg of either anti H3K4me3 (Diagenode pAb003-050), H3K27me3 (Millipore 07-449), or IgG (Abcam ab46540) in ChIP lysis buffer. Cross-linked cells were lysed in Lo-Bind microfuge tubes (Eppendorf 022431021) followed by a 40 minute sonication in a BioRuptor sonicator (Diagenode UCD-200) set to high power, 7” ON, 15” OFF, changing ice/water slurry every 10 minutes.

Lysate was diluted in lysis buffer and cleared of debris. Cleared lysate from ~50,000 cells was divided into four equal aliquots, and one aliquot used for Input. The remaining three aliquots (equivalent to ~12,500 cells each) were used for IP of H3K4me3, H3K27me3 and IgG. Lysates were incubated with preformed bead/antibody overnight at 4C with mixing. The chromatin/bead/antibody complexes were washed sequentially with lysis buffer three times, DOC buffer once, and TE buffer once, followed by a transfer to a fresh PCR tube. Chromatin was eluted using elution buffer (1% SDS, 0.1M NaHCO3). IP and Input samples were treated with RNaseA followed by Proteinase K treatment. Cross-linking was reversed by incubating overnight at 65C while shaking. DNA purification was done using a QiaQuick PCR Purification Kit (Qiagen 28104). E11.5 ChIP data are representative of 3-7 biological replicates expect for Dnmt3b which was technically replicated. Additional PGC data represent a minimum of 2 biological
replicates except for male E12.5 which were technically replicated. All primers are listed in Appendix I. A full protocol is located in the extended experimental procedures.

*Low Cell ChIP-Seq*

ChIP material was obtained and processed as described above with the following modifications. Cells were cross-linked in 0.25% formaldehyde (Thermo Scientific 28906) in PBS for 10 minutes at room temperature prior to quenching with 125mM glycine. Cross-linked material was sonicated using a Covaris sonicator for 12 min at Duty 5%, Intensity 3, and Bursts 200. Two aliquots of E11.5 PGCs, consisting of 104,000 and 97,000 cells, were each divided equally to perform ChIP for H3K4me3, H3K27me3, and an input control. IP-DNA and Inputs were purified using a MinElute PCR Purification Kit (Qiagen 28004) and libraries generated using the ThruPLEX-FDPrep Kit (Rubicon Genomics R40048) with 20 cycles of amplification for IP-DNA and 15 cycles for Input DNA.

*ChIP-Seq Analysis*

Reads from each ChIP-Seq library were filtered to retain only unique sequences (Table S1). Reads were aligned to the mm9/NCBI build 37 mouse genome using bowtie and reads mapping only once were retained. Data for ESCs were handled in the same manner, obtained from GEO accession GSE12241 (Mikkelsen et al., 2007). To call and compare bivalent domains between the PGC and ESC datasets, a 13-state segmentation of the genome was generated using ChromHMM (v1.06) (Ernst and Kellis, 2012). The four samples (PGC H3K4me3, PGC H3K27me3, ESC H3K4me3, ESC H3K27me3) were treated as distinct marks from a single cell type to allow a direct and unbiased comparison of the genome segmentation in PGCs and ESCs. K-means clustering was used to group the state emission parameters for each sample into 2 groups ("on" or "off"), and these were used to classify each of the 13 ChromHMM segmentation states as "bivalent", "H3K4me3 only", "H3K27me3 only", or "none" for each
cell type. Data from the replicate PGC libraries were pooled together for this analysis. H3 ChIP-Seq data from the Mikkelsen dataset was used as control for the ESC samples.

All gene set analysis is based on the mm9 RefSeq gene set. To ensure a conservative segregation of bivalent domains associated with regulation of annotated genes from orphan bivalent domains, genes were considered bivalent if an H3K4me3/H3K27me3 enriched region fell in the range 0.2kb upstream of the transcription start site (TSS) through the transcription end site (TES) of any isoform of the gene. H3K4me3/H3K27me3 enriched regions occurring outside the range from 0.2kb upstream of the TSS through the TES of any isoform of a gene were called orphans. Genes were considered H3K4me3-only if no isoforms were bivalent and if any isoform had H3K4me3-only regions falling in the window from -0.2 kb upstream to 0.2 kb downstream of the TSS. Similarly, genes were considered to be H3K27me3-only if they were neither bivalent nor H3K4me3-only and if any isoform had regions of H3K27me3-only occurring in the +/- 0.2kb window. Analysis of enrichment for Gene Ontology terms was done using DAVID (Huang et al., 2009). The Gene Expression Omnibus accession number for the ChIP-Seq data reported in this paper is GSE46396.

**CAGE Analysis**

CAGE tag cluster locations in mm9 and tags per million (tpm) values for 22 tissue types were obtained from FANTOM 4 (Kawaji et al., 2009). To assess the statistical significance of the accumulation of CAGE tags, the probability $p_i$ of observing a CAGE tag within 200bp of each orphan bivalent domain was first estimated based on CAGE tag representation in the region $X_i$ containing the bivalent domain extended 10kb in each direction. Let $Y_i$ be the indicator random variable that indicates whether the observed bivalent domain in $X_i$ had a detectable CAGE tag cluster within 200 bp. Assuming that $Y_i$ are independent Bernoulli random variables with probability $p_i$, the sum of $Y_i$ is approximately normal with mean $\Sigma p_i$, and variance $\Sigma p_i(1-p_i)$, via the Lyapunov central limit theorem (see extended experimental procedures).
Tissue Specificity Calculation

For each CAGE tag cluster, tissue specificity score was computed based on the Jensen-Shannon divergence between the relative abundance of tpm values across the tissue types and the extreme distribution of being expressed in only one tissue type where the tag cluster has the greatest expression value (Cabili et al., 2011).
Supplemental Material

A

Binding Enrichment as Percent Input Chromatin

B

Binding Enrichment as Percent Input Chromatin Normalized to Positive Control

Figure 3.3. ChIP enrichment is specific and constant independent of starting cell number. (A) The low cell ChIP protocol has a low background and is consistent across 3 orders of magnitude of cell numbers. ChIP efficiency for myc and an IgG control in an H3.3-myc transgenic ESC line is represented as a percent of input. Enrichment is shown for 2 biological replicates of 2.5x10⁶, 2.5x10⁵, 1.25x10⁴ cells/IP, and 1 replicate of 2.5 x10⁴ cells/IP. Error bars express standard deviation of enrichment. (B) The data shown in (A) normalized to locus 1 shows the stability of enrichment independent of the number of starting cells. Positive and negative control regions were chosen based in part on Goldberg et al., 2010. Locus 1, TES Actb; locus 2, TFBS TF5-4; locus 3, TFBS TF5-2; locus 4, TFBS Oct4-43; locus 5, TFBS Nanog-34; locus 6, TFBS Oct4-4; locus 7, TSS Nnat; locus 8, TSS Grb10; locus 9, TSS Vamp7; locus 10, GB Nnat; locus 11, GB Grb10; locus 12, GB polrmt; locus 13, major satellite repeat; locus 14, minor satellite repeat; locus 15, chr8 gene desert. TES, transcriptional end site; TFBS, transcription factor binding site; TSS, transcriptional start site; GB, gene body; RE, repeat element; and GD, gene desert. H3.3 is expected to be enriched at TFBS, TSS, TES and GB, but not RE or GD (Goldberg et al., 2010).
Figure 3.52. Analysis of H3K4me3- and H3K27me3-marked genes in PGCs and ESCs. (A) UCSC Genome Browser view of ChIP-Seq signal for 2 biological replicates of E11.5 PGCs from this study, ChIP-Seq signal for E11.5 PGCs from a recent study (Ng et al., 2013), and RNA-Seq signal for E11.5 PGCs (Seisenberger et al., 2012) at Prdm1(Blimp1), a germline marker expressed at E11.5. (B) Genome-wide comparison of chromatin states in E11.5 PGCs and ESCs, presented at the base level as a bipartite graph. The area of each node is proportional to its genomic space in base pairs, with the exception of the “None” nodes, which are half as big as their true size. Colored arrows indicate the relative basewise composition of the PGC chromatin states in terms of the ESC chromatin states. (C) Top biological process gene ontology terms (level 3) as determined using DAVID for genes marked by H3K27me3 only in PGCs only. (D) Distances of all bivalent domains to nearest TSS for ESCs (white bars) and PGCs (dark grey bars) with transparent overlay. Distances are presented as the log base 10 of the genomic length in base pairs from bivalent domains to their nearest TSS, with bivalent domains upstream of their nearest TSS assigned negative distances after logarithm. Bivalent domains overlapping their nearest TSS were assigned a genomic length of 1. TSSs were defined as the starting positions of RefSeq genes. Bins in the histogram include their lower endpoint but exclude their higher endpoint. (E) Venn diagram representing all possible overlaps of genes marked by H3K4me3 and H3K27me3 in E11.5 PGCs and ESCs. (F) Most bivalently marked RefSeq genes are protein-coding, but some non-coding genes are bivalently marked as well. Including longer, multi-exon genes as well as single-exon genes predominantly composed of microRNAs and snoRNAs.
Figure 3.53. Germline specific genes are transcriptionally active and enriched for H3K4me3. Genes expressed in the germline are enriched for H3K4me3 over H3K27me3. ChIP enrichment for H3K4me3 and H3K27me3 in E11.5-13.5 PGCs at the indicated genes. Data represents the mean percent of input ± standard deviation for at least 2 biological replicates. Transcription was measured with qRT-PCR for the indicated genes and represented as a percent of the housekeeping gene L7 ± standard deviation. ND; no data.
<table>
<thead>
<tr>
<th>Library</th>
<th>Total Reads</th>
<th>Duplicate reads removed</th>
<th>Non-dup reads mapped</th>
<th>Read uniquely mapped</th>
<th>Signal-to-Noise Ratio (p-value)</th>
<th>Pearson's Correlation RepA/RepB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input RepA</td>
<td>41038811</td>
<td>5625066</td>
<td>31536892</td>
<td>29500691</td>
<td>N/A</td>
<td>0.9940900</td>
</tr>
<tr>
<td>Input RepB</td>
<td>46267026</td>
<td>6922657</td>
<td>35020258</td>
<td>32778014</td>
<td>N/A</td>
<td>0.9780485</td>
</tr>
<tr>
<td>H3K4me3 RepA</td>
<td>78362811</td>
<td>51187035</td>
<td>20133350</td>
<td>19233029</td>
<td>15.0587 (0.075979)</td>
<td>0.9780485</td>
</tr>
<tr>
<td>H3K4me3 RepB</td>
<td>80676889</td>
<td>58527900</td>
<td>15626806</td>
<td>14981050</td>
<td>20.114 (0.38541)</td>
<td></td>
</tr>
<tr>
<td>H3K27me3 RepA</td>
<td>71385592</td>
<td>41140846</td>
<td>23215330</td>
<td>21279922</td>
<td>N/A</td>
<td>0.9963008</td>
</tr>
<tr>
<td>H3K27me3 RepB</td>
<td>67945629</td>
<td>40356533</td>
<td>21147219</td>
<td>19405248</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.51. ChIP-Seq library statistics. Mapping statistics for ChIP-Seq libraries. Only non-duplicate (non-dup) reads were aligned to the genome. Only reads that could be assigned uniquely to the genome were used in later analysis. The signal to noise ratio (SNR) was computed using CHANCE. P-values represent the probabilities of observing the given SNR or less in ENCODE data. Large p-values (>0.05) indicate the H3K4me3 libraries from this study do not differ greatly from ENCODE data. Note that CHANCE does not have a feature to compare mouse H3K27me3 data with ENCODE. To compute correlation, the mouse genome was first segmented into non-overlapping 500bp bins. For each bin and each ChIP-Seq library, the number of reads with starting positions occurring in the bin was determined. In each library comparison, all genomic bins containing read starting position counts of zero for both biological replicates were discarded, and Pearson’s correlation was calculated across the values of the remaining bins.

<table>
<thead>
<tr>
<th>Type of comparison</th>
<th>PGC</th>
<th>ESC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bivalent Domains</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assigned to gene</td>
<td>6246</td>
<td>4369</td>
</tr>
<tr>
<td>Orphan</td>
<td>2886</td>
<td>1702</td>
</tr>
<tr>
<td>Total</td>
<td>9132</td>
<td>6071</td>
</tr>
<tr>
<td>Overlap with ESC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bivalent Domains</td>
<td>PGC Total overlapping ESC Total</td>
<td>4046</td>
</tr>
<tr>
<td></td>
<td>PGC Orphan overlapping ESC Orphan</td>
<td>857</td>
</tr>
<tr>
<td>Overlap with PGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bivalent Domains</td>
<td>ESC Total overlapping PGC Total</td>
<td>4341</td>
</tr>
<tr>
<td></td>
<td>ESC Orphan overlapping PGC Orphan</td>
<td>884</td>
</tr>
</tbody>
</table>

Table 3.52 Bivalent domains in PGCs and ESCs. Bivalent domains occurring in the region from 0.2kb upstream of the transcription start site through the transcription end site of any isoform of a gene were assigned to that gene. Bivalent domains that could not be assigned to any gene were considered “orphan.” Because the bivalent domains in the two cell types do not necessarily have identical boundaries, overlaps between the sets are presented individually for PGCs and ESCs.
<table>
<thead>
<tr>
<th>Tissue Type</th>
<th>Number of expressed CAGE tag clusters</th>
<th>Tissue Type</th>
<th>Number of expressed CAGE tag clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>hippocampus</td>
<td>1534</td>
<td>hippocampus</td>
<td>218</td>
</tr>
<tr>
<td>lung</td>
<td>923</td>
<td>liver</td>
<td>174</td>
</tr>
<tr>
<td>liver</td>
<td>788</td>
<td>lung</td>
<td>173</td>
</tr>
<tr>
<td>macrophage</td>
<td>539</td>
<td>macrophage</td>
<td>109</td>
</tr>
<tr>
<td>visual_cortex</td>
<td>441</td>
<td>embryo</td>
<td>70</td>
</tr>
<tr>
<td>embryo</td>
<td>425</td>
<td>DFAT-D1</td>
<td>46</td>
</tr>
<tr>
<td>somatosensory_cortex</td>
<td>320</td>
<td>visual_cortex</td>
<td>38</td>
</tr>
<tr>
<td>DFAT-D1</td>
<td>309</td>
<td>brain</td>
<td>34</td>
</tr>
<tr>
<td>cerebral_cortex</td>
<td>261</td>
<td>somatosensory_cortex</td>
<td>31</td>
</tr>
<tr>
<td>brain</td>
<td>239</td>
<td>cerebral_cortex</td>
<td>27</td>
</tr>
<tr>
<td>testis</td>
<td>178</td>
<td>blood</td>
<td>24</td>
</tr>
<tr>
<td>prostate</td>
<td>131</td>
<td>testis</td>
<td>19</td>
</tr>
<tr>
<td>blood</td>
<td>131</td>
<td>prostate</td>
<td>16</td>
</tr>
<tr>
<td>diencephalon</td>
<td>115</td>
<td>corpus_striatum</td>
<td>13</td>
</tr>
<tr>
<td>heart</td>
<td>94</td>
<td>diencephalon</td>
<td>11</td>
</tr>
<tr>
<td>muscle</td>
<td>92</td>
<td>heart</td>
<td>10</td>
</tr>
<tr>
<td>corpus_striatum</td>
<td>76</td>
<td>muscle</td>
<td>7</td>
</tr>
<tr>
<td>medulla_oblongata</td>
<td>19</td>
<td>mammary_gland</td>
<td>6</td>
</tr>
<tr>
<td>eye</td>
<td>10</td>
<td>eye</td>
<td>5</td>
</tr>
<tr>
<td>mammary_gland</td>
<td>8</td>
<td>amnion</td>
<td>2</td>
</tr>
<tr>
<td>amnion</td>
<td>2</td>
<td>placenta</td>
<td>1</td>
</tr>
<tr>
<td>placenta</td>
<td>1</td>
<td>medulla_oblongata</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3.53. CAGE tag expression by tissue type in PGC orphan bivalent domains. Number of expressed CAGE tag clusters occurring within 200bp of PGC orphan bivalent domains, by tissue type. Orphan bivalent domains overlapping CpG islands have higher overall CAGE tag expression than orphan bivalent domains not overlapping CpG islands.
Extended Experimental Procedures

Isolation of PGCs

Male B6 mice homozygous for a transgenic Oct4ΔPE:GFP reporter were crossed with Swiss-Webster females. Morning of the vaginal plug was counted as E0.5. The gonadal regions including (for E11.5) or excluding (for E12.5 and E13.5) the dorsal aorta were dissected from embryos and sexed by gonad morphology for E12.5 and later embryos. Tissue from multiple embryos were pooled and dissociated in 0.5mg/ml trypsin and 2mg/ml DNase-I, filtered through 40µm cell strainers, and cell sorted using a FACS-Aria. GFP positive and negative cells were collected as PGCs and soma control respectively.

Low Cell ChIP-qPCR

Cells were pooled into batches of ~50,000 cells and cross-linked in 1% formaldehyde (Sigma F8775-25ml) in PBS for 5min at room temperature, quenched with 125mM glycine and centrifuged in a swing bucket carrier at 490 g for 10 minutes at 4C. The supernatant was removed and pellets were frozen at -80C for up to 6 months. For each IP, 11µl of protein A Dynabeads (Life Technologies 1001D) were pre-incubated with 2.4µg of either anti H3K4me3 (Diagenode pAb003-050), H3K27me3 (Millipore 07-449), or IgG (Abcam ab46540) for 2 hours at 4C with end-over-end rotation in ChIP lysis buffer (50 mM HEPES/KOH pH 7.5, 300 mM NaCl, 1 mM EDTA, 1 % Triton X-100, 0.1 % Deoxycholate (DOC), 0.1% SDS, 1x fresh protease inhibitor (Roche 11697498001)). Pellets of cross-linked cells were thawed and lysed in a 1.5ml Lo-Bind microfuge tube (Eppendorf 022431021) with 120µl of lysis buffer on ice for 5 minutes followed by 40 minute sonication in a BioRuptor sonicator (Diagenode UCD-200) set to high power, 7” ON, 15” OFF, changing ice/water slurry every 10 minutes.

Lysate was cleared by 2 rounds of dilution in 400µl of lysis buffer, centrifugation at 12,000 g for 10 minutes at 4C and collection of the supernatant. Cleared lysate from ~50,000 cells was divided into four
equal aliquots and one aliquot used for Input which was stored at 4C. The remaining three aliquots (equivalent to ~12,500 cells each) were used for IP of H3K4me3, H3K27me3 and IgG. Lysates were incubated with preformed bead/antibody complexes in a PCR tube with ~200µl total volume of lysate buffer overnight at 4C with mixing. The chromatin/bead/antibody complexes were washed sequentially with 180µl ice-cold lysis buffer three times, once with 180µl ice-cold DOC buffer (10 mM Tris pH 8, 0.25 M LiCl, 0.5% NP-40 (Sigma NP40S-100ml), 0.5% DOC, 1 mM EDTA), and once with 180µl ice-cold TE buffer, followed by a transfer to a fresh PCR tube. Chromatin was eluted using 2 successive rounds of incubation in 150µl elution buffer (1% SDS, 0.1M NaHCO3) for 20 minutes at room temperature with mixing, and eluates pooled. IP and Input samples were treated with 0.2 mg/mL RNaseA in 200mM NaCl, 10mM EDTA, 40mM Tris pH 6.5 and incubated for 30 minutes at 37C. RNaseA was inactivated with 0.67µg/ul Proteinase K treatment at 55C for 3 hours on a shaker followed by reversal of the cross-linking by incubating overnight at 65C while shaking. DNA purification was done using an QiaQuick PCR Purification Kit (Qiagen 28104), and quantitative PCR performed using SYBR FAST qPCR kit Master Mix (Kapa KK4604) on an ABI 7900HT. E11.5 ChIP data are representative of 3-7 biological replicates expect for Dnmt3b which was technically replicated. Additional PGC data represent a minimum of 2 biological replicates except for male E12.5 which were technically replicated. All primers used are listed in Appendix I. Appendix II provides a step-by-step protocol.

**Low Cell ChIP-Seq**

ChIP material was obtained and processed as described above with the following modifications. Two aliquots of E11.5 PGCs, consisting of 104,000 and 97,000 cells, were each divided equally to perform ChIP for H3K4me3, H3K27me3, and an input control. Cells were cross-linked in 0.25% formaldehyde (Thermo Scientific 28906) in PBS for 5min at room temperature prior to quenching with 125mM glycine. Cross-linked material was sonicated using a Covaris sonicator for 12’ at Duty 5%, Intensity 3, and Bursts 200. IP-DNA and Inputs were purified using a MinElute PCR Purification Kit (Qiagen 28004) and libraries
generated using the ThruPLEX-FDPrep Kit (Rubicon Genomics R40048) with 20 cycles of amplification for IP-DNA and 15 cycles for Input DNA.

**ChIP-Seq Analysis**

To correct for potential amplification biases, reads from each ChIP-Seq library were first filtered to retain only reads possessing unique sequences (Table S1). Retained reads were aligned to the mm9/NCBI build 37 mouse genome using bowtie (v0.12.7), allowing up to 10 mappings per read and suppressing all mappings for any reads with more than 10 valid alignments by using the optional parameters: -k 10 -m10 --best --strata. Reads mapping only once to the genome were used in later analysis. Data for ESCs were handled in the same manner, obtained from GEO (accession GSE12241) (Mikkelsen et al., 2007).

To compute correlation between biological replicates, the mouse genome was first segmented into non-overlapping 500-bp bins. For each bin and each ChIP-Seq library, the number of reads with starting positions occurring in the bin was determined. In each library comparison, all genomic bins containing read starting position counts of zero for both biological replicates were discarded, and Pearson’s correlation (Table S1) was calculated across the values of the remaining bins. Signal-to-Noise (SNR) was calculated using CHANCE. P-values (Table S2) represent the probabilities of observing the given SNR or less in ENCODE data. The p-values indicate the H3K4me3 libraries do not differ greatly from ENCODE data. Note that CHANCE does not have a feature to compare mouse H3K27me3 data with ENCODE.

To call and compare bivalent domains between the PGC and ESC datasets, a 13-state segmentation of the genome was generated using ChromHMM (v1.06) (Ernst & Kellis, 2012). The four samples (PGC H3K4me3, PGC H3K27me3, ESC H3K4me3, ESC H3K27me3) were treated as distinct marks from a single cell type to allow a direct and unbiased comparison of the genome segmentation in PGCs and ESCs. K-means clustering was used to group the state emission parameters for each sample into 2 groups ("on" or "off"), and these were used to classify each of the 13 chromHMM segmentation states as "bivalent",...
"H3K4me3 only", "H3K27me3 only", or "none" for each cell type. Data from the replicate PGC libraries were pooled together for this analysis. H3 ChIP-Seq data from the Mikkelsen dataset was used as control for the ESC samples. A stringent threshold value of 10^-6 Poisson tail probability was used to binarize the histone modification data into background and robust ChIP-enriched regions.

All gene set analysis is based on the RefSeq gene set downloaded from the refGene table of the mm9 mouse assembly on the UCSC Genome Browser on January 24, 2013. Parent gene/isoform relationships were determined using the “name” and “name2” fields in the refGene table. To ensure a conservative segregation of bivalent domains associated with regulation of annotated genes from orphan bivalent domains, genes were considered bivalent if an H3K4me3/H3K27me3 enriched region fell in the range 0.2kb upstream of the transcription start site through the transcription end site of each gene’s longest isoform. Similar gene ontology terms and overlap between PGCs and ESCs were found when bivalent genes were defined by a bivalent domain occurring ±0.2kb from a RefSeq TSS (data not shown). H3K4me3/H3K27me3 enriched regions occurring outside the range from 0.2kb upstream of the transcription start site through the transcription end site of a gene’s longest isoform were called orphans, with a median distance of 13kb to the nearest gene. Genes were considered H3K4me3-only if no isoforms were bivalent and if any isoforms had H3K4me3-only regions falling in the window from -0.2 kb upstream to 0.2 kb downstream of the transcription start site. Similarly, genes were considered to be H3K27me3-only if they were neither bivalent nor H3K4me3-only and if any isoforms had regions of H3K27me3-only occurring in the +/- 0.2kb window. Analysis of enrichment for Gene Ontology terms was done using DAVID (Huang, Sherman, & Lempicki, 2009).

CAGE Analysis

CAGE tag cluster locations in mm9 and tags per million (tpm) values for 22 tissue types were obtained from FANTOM 4 (Kawaji et al., 2009). We found that 927 out of 1413 (65.6%) orphan bivalent domains
with CGI contained detectable CAGE tags within 200 bp in at least one tissue type. Similarly, 461 out of 1473 (31.3%) orphan bivalent domains without CGI also expressed CAGE tags within 200 bp in at least one tissue type. To assess the statistical significance of the accumulation of CAGE tags, the orphan bivalent domains were extended by 10kb in each direction, and a Bernoulli process with probability \( p_i \) was associated to each extended domain \( X_i \) as follows: \( X_i \) is scanned with a running window of width equal to the size of the original bivalent domain, and a binary random variable is assigned at each scan position, with value 1 if the window has a detectable CAGE tag within 200 bp in any tissue type or value 0 otherwise; the average of these binary random variables then provides an estimate of \( p_i \) for this region. Let \( Y_i \) be the indicator random variable that indicates whether the observed bivalent domain in \( X_i \) had a detectable CAGE tag cluster within 200 bp. Assuming that \( Y_i \) are independent Bernoulli random variables with probability \( p_i \), the sum of \( Y_i \) is approximately normal with mean \( \Sigma p_i \) and variance \( \Sigma p_i(1-p_i) \), via the Lyapunov central limit theorem. This computation yielded a \( z \)-score of 27.6 (\( p \)-value = 5.78 \( \times 10^{-168} \)) for orphan bivalent domains with CGI and a \( z \)-score of 12.3 (\( p \)-value = 3.61 \( \times 10^{-35} \)) for orphan bivalent domains without CGI.

**Tissue Specificity Calculation**

For each CAGE tag cluster, tissue specificity score was computed based on the Jensen-Shannon divergence between the relative abundance of tpm values across the tissue types and the extreme distribution of being expressed in only one tissue type where the tag cluster has the greatest expression value (Cabili et al., 2011).

**qRT-PCR Analysis**

3000 PGCs were isolated by FACS as above and sorted directly into RLT buffer (Qiagen). RNA was purified using the RNeasy kit (Qiagen 74104). cDNA was generated using a High Capacity cDNA Reverse Transcriptase kit (ABI 4368814) and relative quantitative PCR was performed using L7 as a reference
gene. Similar results were obtained using H2A or Ubb as a reference gene. All primers used are listed in Appendix I.

References


Labosky, P. a, Barlow, D.P., and Hogan, B.L. (1994). Mouse embryonic germ (EG) cell lines: transmission through the germline and differences in the methylation imprint of insulin-like growth factor 2 receptor (Igf2r) gene compared with embryonic stem (ES) cell lines. Development (Cambridge, England) 120, 3197–3204.


CHAPTER 4

CONCLUSION

The role of chromatin in regulation of the transcriptional machinery is a growing area of interest at both the level of basic research and public health. The decoding of the human genome has granted us identification of the components involved in development and disease. However, epigenetics presents the opportunity to understand how these components are used and regulated (or mis-regulated). The data presented here helps to elucidate some of the interactions between chromatin and transcription. It will aid in laying the groundwork for future studies to uncover critical aspects of this interaction and has the potential aid in discovering novel points of therapeutic intervention.

Chd1

Chd1 is known to associate with the transcriptional machinery and the work presented in chapter 2 implicates a ubiquitous role in its regulation. Substantial evidence supports a role in both transcriptional initiation and elongation, albeit in opposite capacities at each location. Does Chd1 function regulate RNA Pol II recruitment or processivity? Evidence in chapter 2 suggests a reduction in RNA Pol II recruitment in Chd1 null cells, but this can potentially be indirect. A transcriptional run-on assay to directly measure nascent transcription within a cell would be pivotal in determining if Chd1 regulates processivity.

While Chd1 has been identified in immuno-precipitation (IP) studies of various transcription related complexes, IP-mass spectrometry of Chd1 has failed to identify high confidence protein interactions (unpublished data, M.S.). One potential explanation is that Chd1 only experiences weak, transient interactions with its parent complex, and direct IP would be biased to enrich for the unbound fraction. A
second, more likely, explanation is that Chd1 functions as a monomer but shares the same or proximal substrate as various transcription complexes. The apparent interaction seen in IP of transcriptional components (Kelley et al., 1999; Simic et al., 2003; Pray-Grant et al., 2005; Sims et al., 2007; Lin et al., 2011) could be mediated by a shared chromatin or DNA substrate rather than a direct protein-protein contact. Nuclease treatment of the lysate, which has been occasionally implemented to detect this, would fail to achieve the desired effect since any shared DNA substrate would be largely insulated from digestion. Multi-component structural biology studies that include chromatin and DNA substrates will be necessary to definitively answer this question, but will be complicated by the large size and unknown interplay between the various structural domains. If Chd1 stands to be a general regulator of chromatin, a clear understanding of its enzymatic activity and binding partners will greatly facilitate the clarification of its apparent opposing functions at different stages of transcription.

Mammalian Chd1 is expressed at low to moderate levels nearly ubiquitously and is up-regulated in highly proliferative cell types such as the epiblast and adult immune system (Hruz et al., 2008). Unexpectedly, depletion of Chd1 is suspected of reducing RNA Pol II processivity at all gene loci and potentially RNA Pol I. A more thorough investigation using a cell number normalized genome-wide approach, such as RNA-seq, will be required to definitively determine this. If this proves to be accurate, does Chd1 serve as an endogenous regulator of global RNA polymerase processivity? Or in other words, does the cell modulate Chd1 activity to regulate global transcription and thus cell-cycle frequency? Related studies have found the transcription factor c-Myc, a known oncogene, to be a global regulator of transcription (Lin et al., 2012; Lovén et al., 2012; Nie et al., 2012), suggesting a potential functional or biochemical interaction between c-Myc and Chd1. Moreover, chromatin level regulation of transcription and cell-cycle has been observed via phosphorylation of UBF at rRNA genes. It will be interesting to determine if a similar mechanism exists for Chd1 which possesses a large number of potential phosphorylation and ubiquitination sites. While inactivating mutations to Chd1 have been implicated in
developmental defects in both mouse and human, Chd1 mutations correlated with cancer seem to be dose responsive (Huang et al., 2012; Liu et al., 2012; Burkhardt et al., 2013). Unlike most cell-cycle related cancer drug targets, Chd1 oblation does not result in complete cell-cycle arrest, but rather a reduction in cycling frequency. Inhibition of Chd1 and similar chromatin level regulators of transcription may serve as a novel class of oncogenic targets with broader efficacy and reduced toxicity. Further studies of endogenous regulation of transcription via Chd1 are warranted to determine its prospects as a drug target.

Trans-Generational Epigenetic Inheritance

Recently, the phenomenon of trans-generational epigenetic inheritance has seen a rise in popularity as an explanation for disease predisposition. In common usage, this mode of inheritance refers to two distinct concepts: i) inheritance of acquired traits and ii) transmission of epigenetic information across generations. Inheritance of acquired traits is a reemergence of an old theory popularized by Jean-Baptiste Lamarck in the early 1800’s. The underpinnings of which are, the environment can influence an individual and that this influence can be inherited by subsequent generations. The above stated theory does not specify a mechanism in which this inheritance may happen. The transmission of epigenetic information across generations would require the presence of a stable molecular entity, other than DNA, to be carried and propagated through the germline. Definitive evidence for this transmission in mammals is lacking, but recent works (see below) provide some clues as potential mechanisms begin to emerge.

Experimental evidence for transmission of an acquired trait was first proposed in the context of genetic theory by Waddington and Schmalhausen over 60 years ago (Waddington, 1942, 1953; Schmalhausen, 1949; Pigliucci et al., 2006). Genetic assimilation (GA) predicts that a phenotypic plasticity, if stimulated by the environment and selected for via natural selection, will eventually become stable and cease to
require the original stimulant for its presentation. In experiments performed by Waddington, certain phenotypes (i.e. cross-veinless in *Drosophila melanogaster*) could be observed at low frequencies in wild type populations when the animals were subjected to a stimulant (i.e. heat shock during embryogenesis). When these animals were selected for and bred with continued exposure to the stimulant, the frequency of the phenotype became increasingly common. More surprisingly, the stimulus no longer became necessary to observe the cross-veinless phenotype (Waddington, 1953). Importantly, the underlying source of the phenotype may arise from genetic mutation, cryptic genetic variation, or epigenetic adaptation. Regardless of the mechanism, in terms of Darwinian evolution this was a non-adaptive evolution to an environmental stress. Of note, not every phenotype need be plastic nor result in an irreversible assimilation. The prevalence of such situations is subject to empirical investigation, although there is growing evidence for GA in plant and animal species (Heil et al., 2004; Mery and Kawecki, 2004; Palmer, 2004; Keogh et al., 2005; Scoville and Pfrender, 2010; Rosas et al., 2013).

Darwinian evolution, as originally presented did not specify a mechanism of how traits were inherited, nor did it propose a source for how variation that led to speciation arose. The advent of the Modern Synthesis incorporated the discovery of DNA as a vehicle of trait transfer, and spontaneous DNA mutation as the source of variation. The reemergence of acquired traits should not be viewed as an alternative to natural selection as it was at the turn of the century, but as an extension to our understanding of evolution and inheritance; as the Modern Synthesis was an extension to neo-Darwinism, which itself was an extension to the idea of common decent.

A molecular mechanism to describe the stabilization of an assimilated trait is an area of active debate. The emergence of epigenetics as a regulatory mechanism, including DNA methylation, histones, and extra-chromosomal nucleic acids, hints at intriguing possibilities. A number of studies have implicated
heritable environmental influences in humans and mice, but a definitive determination of the mechanism has been elusive (Morgan et al., 1999; Rakyan et al., 2003; Weaver et al., 2004; Pembrey et al., 2006; Heijmans et al., 2008). Most of these studies have implicated DNA methylation at gene promoters as the causative agent, but proof of epigenetic transmission is technically difficult and alternative explanations such as culture, gene duplications, or trans-acting DNA mutations are difficult to rule out (Yi and Richards, 2009; Daxinger and Whitelaw, 2010).

Epigenetic marks are considered relatively stable during somatic development (i.e. transmission during mitosis). However their persistence across generations would require escape from epigenetic reprogramming in the zygote and developing PGCs. In support of this notion, mature mammalian spermatogonia retain histones and histone post translation modifications, including bivalent domains, at several hundred developmentally important regulatory genes including the Hox cluster (Gatewood et al., 1990; Hammoud et al., 2009; Brykczynska et al., 2010; Miller et al., 2010; Carrell, 2012; Hisano et al., 2013). Minimal information is available on the chromatin state of mature/maturing oocytes, but with technical advances like the low-cell ChIP protocol presented in chapter 3, investigations into oocyte epigenetics can be expected. It will be interesting to see if bivalent domains are present at similar loci in oocytes as in spermatogonia and if these marks remain static in Meiosis I arrested oocytes waiting for ovulation as a female ages.

Single-cell ChIP analysis of the fertilized zygote is not technically feasible at present and therefore definitive evidence of epigenetic retention through zygotic reprogramming is not possible. The recent development of nano-pore technologies promises to lead the way toward single-cell ChIP, but is likely several years from maturation (Murphy et al., 2013). However, immunofluorescent studies have observed targeted enrichment of the histone variant H3.3 during the earliest phases of zygotic reprogramming of the mouse male pronucleus (Santenard and Torres-Padilla, 2009; Santenard et al.,
suggesting epigenetic marks of spermatogonial origin play a role in preferential recruitment. Further on in development, examination of the developing PGCs reveals that many loci completely escape DNA demethylation, albeit these are largely confined to transposable elements (Hajkova et al., 2002; Seisenberger et al., 2012; Hackett et al., 2013). Works described in chapter 3 and elsewhere have extended this examination to include histone marks during epigenetic reprogramming in PGCs (Lesch et al., 2013; Ng et al., 2013; Sachs et al., 2013). Technical limitations have prevented examination of the earlier time points observed for DNA methylation, but PGCs retain multiple chromatin marks, including bivalent domains, as early as E10.5 (unpublished data, M.S.) at several hundred developmentally important genes. Future studies are needed to examine histone variants in PGCs at these time points as a potential instructive code for rebuilding the epigenome after reprogramming. Taken together, there is strong evidence that a sub-set of genomic loci retain chromatin marks during epigenetic reprogramming of the mammalian germline.

While the functional significance of retained chromatin marks remains unknown, recent studies have implicated DNA methylation and chromatin marks in the germline as the causative agent of genetic assimilation in mammals. The earliest study to obtain this level of molecular resolution observed a reduced fertility of F₄ male rats caused by an environmental toxic exposure to the gestating F₀ great-grandmother. Alterations to promoter CpG methylation of a few genes within the spermatagonia was suggested as the mechanism of information transfer (Anway et al., 2005). With technical advances, studies have become more sophisticated and two recent studies have been able to link acquired transgenerational disease resistance and olfaction sensitivity to alterations in histone variants and DNA methylation respectively at biologically relevant genes in spermatogonia (Zeybel et al., 2012; Dias and Ressler, 2013). One intriguing possibility that arises from these studies is that complex phenotypes specific to multiple organ systems can be predicted by the epigenetic state of the paternal
spermatogonia. Further technical advances and future studies will be need to clarify the role of spermatogonial chromatin in development and disease.

The simultaneously largest gap and largest impact of trans-generation epigenetic reprogramming lies with establishing a functionally significant role for inherited chromatin. The above studies are intriguing in their observed correlations, but fail to conclude decisively the necessary agent or mark. Future studies will have to experimentally alter the chromatin state in one generation and observe the phenotype in the next. While technically challenging, these studies are possible. A PGC specific ablation of bivalent domains can be achieved via an inducible Pou5f1<sup>CreER<sup>+/</sup></sup>; Ezh2<sup>flox/flox</sup> mouse. Ezh2 is the enzymatic subunit of the PRC2 complex which deposits the repressive H3K27me3 modification. Genome-wide ablation studies such as this will be useful in establishing a role of inherited chromatin during normal development. However, in order to study specific epimutations and the potential epigenetic origins of disease, locus specific chromatin alternations will be necessary. The advent and recent maturation of CRISPR technologies (Ran et al., 2013), which link enzymatic proteins to a targeting nucleic acid, promises to permit these investigations for the first time. Gene specific epigenetic modifications will allow for unprecedented study of epimutation in development and inheritance.

The impact of trans-generation epigenetic inheritance on evolution and development are unclear, but potentially large. While several studies have identified cases most probably subject to this phenomenon, its prevalence remains untested and its importance compared to genetic variables unknown. Complex studies of this nature necessitate a great understanding of the mechanisms responsible for any epigenetic transmission before a phenotype can be dissected into discrete genetic and epigenetic origins.
References


### Table A.1. ChIP Primers

<table>
<thead>
<tr>
<th>oligo name</th>
<th>sequence</th>
<th>cited from</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pax6-ChIP-1F</td>
<td>CTAATCTGCCGAGCTGAACC</td>
<td></td>
</tr>
<tr>
<td>Pax6-ChIP-1R</td>
<td>GCAGGCCGCTAACTTTTCTTTTA</td>
<td></td>
</tr>
<tr>
<td>Nkx2.2-ChIP-1F</td>
<td>ACCACGTGAAGACTGCTGAGT</td>
<td>Bernstein et al., 2006</td>
</tr>
<tr>
<td>Nkx2.2-ChIP-1R</td>
<td>GTTTGTATCCGGTGAGTTATGT</td>
<td></td>
</tr>
<tr>
<td>MyoD1-ChIP-1F</td>
<td>CTCCCACCTCCACTGACATT</td>
<td></td>
</tr>
<tr>
<td>MyoD1-ChIP-1R</td>
<td>ATGTCCAGGGGTAACGTG</td>
<td></td>
</tr>
<tr>
<td>T-ChIP-1F</td>
<td>CCCCCTGTTGTTTCTTTGTA</td>
<td></td>
</tr>
<tr>
<td>T-ChIP-1R</td>
<td>GGAGGAGAATGGCACAACACTGA</td>
<td></td>
</tr>
<tr>
<td>Gata6-ChIP-1F</td>
<td>GATCATGTCTGAGGCGGGTT</td>
<td></td>
</tr>
<tr>
<td>Gata6-ChIP-1R</td>
<td>TACCCATTTCCTCCCTTC</td>
<td></td>
</tr>
<tr>
<td>FoxA1-ChIP-1F</td>
<td>GCCATGAAAGAGGAGGAACAA</td>
<td></td>
</tr>
<tr>
<td>FoxA1-ChIP-1R</td>
<td>CACAGCGAAGCTCATGAC</td>
<td></td>
</tr>
<tr>
<td>HoxA3-ChIP-F</td>
<td>AATTACCTCCCTGACATCTCA</td>
<td></td>
</tr>
<tr>
<td>HoxA3-ChIP-R</td>
<td>TTACAGAGCAGACCCACAAATG</td>
<td></td>
</tr>
<tr>
<td>Hoxa11-ChIP-F1</td>
<td>AGAAGTGCCCTGCTGCTGTA</td>
<td>Xingbo Xu et al., 2011</td>
</tr>
<tr>
<td>Hoxa11-ChIP-R1</td>
<td>GATTTGCAGGAGTACTTGA</td>
<td></td>
</tr>
<tr>
<td>Hoxb2-ChIP-1F</td>
<td>CTCATGTATCCAGGTCGT</td>
<td></td>
</tr>
<tr>
<td>Hoxb2-ChIP-1R</td>
<td>CGAAATTGCGAGGAATTAC</td>
<td></td>
</tr>
<tr>
<td>Hoxb9-ChIP-1F</td>
<td>CCATCCGAGATAGGGAATGC</td>
<td></td>
</tr>
<tr>
<td>Hoxb9-ChIP-1R</td>
<td>GCCCTGAAAGCGTCAGCT</td>
<td></td>
</tr>
<tr>
<td>P16-ChIP-F</td>
<td>ACA CTC CTG GCC TAC CTG AA</td>
<td></td>
</tr>
<tr>
<td>P16-ChIP-R</td>
<td>CGA ACT CGA GGA GAG CCA TC</td>
<td></td>
</tr>
<tr>
<td>Dnmt3b-2F</td>
<td>AGCGGCAAGTAGAAAAGTGTG</td>
<td></td>
</tr>
<tr>
<td>Dnmt3b-2R</td>
<td>TGTGTCAGGGCTGGCTGTA</td>
<td></td>
</tr>
<tr>
<td>MVH-F1</td>
<td>GCGGCTTAAACCGGCGTCAC</td>
<td>Xingbo Xu et al., 2011</td>
</tr>
<tr>
<td>MVH-R1</td>
<td>GCCTCAAACCAAGTGAGGAA</td>
<td></td>
</tr>
<tr>
<td>Fkbp6-2F</td>
<td>ATCTTGCGCACAACCTGT</td>
<td></td>
</tr>
<tr>
<td>Fkbp6-2R</td>
<td>GATCAGTCCCTGTTTCTC</td>
<td></td>
</tr>
<tr>
<td>Nanos2-ChIP-1F</td>
<td>AGGAGCCCTTAAATGGAGGA</td>
<td></td>
</tr>
<tr>
<td>Nanos2-ChIP-1R</td>
<td>TGAACAGGGTGTGACTCTTG</td>
<td></td>
</tr>
<tr>
<td>stra8-chip1-F1</td>
<td>ACCACCTCAACGTCAGGGTTGTA</td>
<td></td>
</tr>
<tr>
<td>stra8-chip1-R1</td>
<td>CGTGAGATGATGAGTCAGGA</td>
<td></td>
</tr>
<tr>
<td>Oct4-4-ChIP-F</td>
<td>GGAACCTGGTGGTGAGGTTGTA</td>
<td>Chew et al., 2005</td>
</tr>
<tr>
<td>Oct4-4-ChIP-R</td>
<td>AGCAGATTAAGGAGGCTAGGACGAG</td>
<td></td>
</tr>
<tr>
<td>Dazl-chip2-F</td>
<td>TAAACCTCAGGCAAGGAA</td>
<td></td>
</tr>
<tr>
<td>Dazl-chip2-R</td>
<td>CCACCTCCGAGGTGTACCA</td>
<td></td>
</tr>
<tr>
<td>spo11-chip1-F</td>
<td>GCCACTGTGTTTCTGGT</td>
<td></td>
</tr>
<tr>
<td>spo11-chip1-R</td>
<td>GAAGCATGTGTGGTCGTG</td>
<td></td>
</tr>
<tr>
<td>Upp1-1F</td>
<td>ATCCCTGAGGAGGACGAG</td>
<td></td>
</tr>
<tr>
<td>Upp1-1R</td>
<td>GACGTGGTCAAGAATGCAG</td>
<td></td>
</tr>
<tr>
<td>Gene</td>
<td>Forward Sequence</td>
<td>Reverse Sequence</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Klf4-chip1-F</td>
<td>AAGGAAGGCGTCTCCAGATTT</td>
<td></td>
</tr>
<tr>
<td>Klf4-chip1-R</td>
<td>TTGAGATCCTGGGTTGAAGGG</td>
<td></td>
</tr>
<tr>
<td>Fragilis-F1</td>
<td>GGTTCCTCAGAAAGTTTCTTTTCC</td>
<td></td>
</tr>
<tr>
<td>Fragilis-R1</td>
<td>TCACTCTTAAACAATGGTGGTTTG</td>
<td></td>
</tr>
<tr>
<td>Fkbp6-2F</td>
<td>ATCTTGGCGCACAACTGCTT</td>
<td></td>
</tr>
<tr>
<td>Fkbp6-2R</td>
<td>GATCACGTGGCTTCTCTC</td>
<td></td>
</tr>
<tr>
<td>Actin-7-F</td>
<td>AGCTCCAGCCTTTGTCAGTCGTA</td>
<td></td>
</tr>
<tr>
<td>Actin-7-R</td>
<td>TTCTTTGGGTTGGTGAAAGGG</td>
<td></td>
</tr>
<tr>
<td>TF5-4-F</td>
<td>TGTGATGGAGAATGCTCAATCTG</td>
<td></td>
</tr>
<tr>
<td>TF5-4-R</td>
<td>GAGCCAAATGAACCTTTTCTCCC</td>
<td></td>
</tr>
<tr>
<td>TF5-2-F</td>
<td>ACTACCCCATTTGTGGCAGTCGATG</td>
<td></td>
</tr>
<tr>
<td>TF5-2-R</td>
<td>GATGGCACACAAAGGAAGCATGG</td>
<td></td>
</tr>
<tr>
<td>TF_Oct4-43-F</td>
<td>GCAATGTTAAAGCGAGTTACAGGAACACA</td>
<td></td>
</tr>
<tr>
<td>TF_Oct4-43-R</td>
<td>AACCCTAAGGCCAGGATGGTCAGTAG</td>
<td></td>
</tr>
<tr>
<td>TF_Nanog-34-F</td>
<td>GGGGAAATACAGTAATTGACCAATCAGC</td>
<td></td>
</tr>
<tr>
<td>TF_Nanog-34-R</td>
<td>CCAAGACCACAAAGTTCCACCTACCTCA</td>
<td></td>
</tr>
<tr>
<td>Nnat- TSS F</td>
<td>ACCCCTCCCTTCACCATCC</td>
<td></td>
</tr>
<tr>
<td>Nnat- TSS R</td>
<td>CGCCAGGTCTACTGGTCT</td>
<td></td>
</tr>
<tr>
<td>grb10- TSS-1F</td>
<td>GAGGAGGAGCAGGAAAGC</td>
<td></td>
</tr>
<tr>
<td>grb10- TSS-1R</td>
<td>CCCCCTGGATGCTTTAATT</td>
<td></td>
</tr>
<tr>
<td>Vamp7- TSS F</td>
<td>TGGAGGAGAATCGAGTGTC</td>
<td></td>
</tr>
<tr>
<td>Vamp7- TSS R</td>
<td>CTATTGACTGAGGGGGAAG</td>
<td></td>
</tr>
<tr>
<td>Nnat- GB F</td>
<td>TTTCCCTAACGTTCCCCTG</td>
<td></td>
</tr>
<tr>
<td>Nnat- GB R</td>
<td>TCCAACTGTCTGAGAGCTG</td>
<td></td>
</tr>
<tr>
<td>grb10-GB F</td>
<td>CCAGCAAAATGGTGGAGATT</td>
<td></td>
</tr>
<tr>
<td>grb10-GB R</td>
<td>CCAGATTCAGGTGTTGCA</td>
<td></td>
</tr>
<tr>
<td>Polrmt-GB 1F</td>
<td>CAGGTAACGTGGAGTGACT</td>
<td></td>
</tr>
<tr>
<td>Polrmt-GB 1R</td>
<td>AGTGGGTTGAAGTGAACCTG</td>
<td></td>
</tr>
<tr>
<td>MajSat-F</td>
<td>GACGACTTTGAAAAATGACGAAATC</td>
<td></td>
</tr>
<tr>
<td>MajSat-R</td>
<td>CATATTCAGGTCCCTACGTGTC</td>
<td></td>
</tr>
<tr>
<td>MinSat-F</td>
<td>CATGGAATAGTAAAAAAC</td>
<td></td>
</tr>
<tr>
<td>MinSat-R</td>
<td>CATCTAATATGTGCTACAGTGTTG</td>
<td></td>
</tr>
<tr>
<td>intergenic_Ch8-F</td>
<td>AAAGGGCCCTCGCCTTTAAAAA</td>
<td></td>
</tr>
<tr>
<td>intergenic_Ch8-R</td>
<td>AGAGCTCCATGGCGAGTAGA</td>
<td></td>
</tr>
<tr>
<td>Rpl3TSSchipF</td>
<td>CGCAGATAGAGAGCGACCA</td>
<td></td>
</tr>
<tr>
<td>Rpl3TSSchipR</td>
<td>GGAGGACATCCTTGCTTTA</td>
<td></td>
</tr>
<tr>
<td>Rpl3TSSPLUSchipF</td>
<td>GCCATCCCTACAGACTGAC</td>
<td></td>
</tr>
<tr>
<td>Rpl3TSSPLUSchipR</td>
<td>GAATGGGCTGTGCTTTGCTT</td>
<td></td>
</tr>
<tr>
<td>Rpl3GBchipF</td>
<td>CTGGGGGCACATCTCATAGT</td>
<td></td>
</tr>
<tr>
<td>Rpl3GBchipR</td>
<td>AGATAGTGCCTCCCTGCGCTC</td>
<td></td>
</tr>
<tr>
<td>NanogTSSchipF</td>
<td>AAATCTATCGCCTTGAGCCGT</td>
<td></td>
</tr>
<tr>
<td>NanogTSSchipR</td>
<td>CACCAACAAATCAGCCTATCTG</td>
<td></td>
</tr>
<tr>
<td>oligo name</td>
<td>sequence</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>---------------------------</td>
<td></td>
</tr>
<tr>
<td>HoxA3-F1</td>
<td>AACTGGAGACCATTCCTGGTT</td>
<td></td>
</tr>
<tr>
<td>HoxA3-R1</td>
<td>CGATCGTGATCACACTCTCTG</td>
<td></td>
</tr>
<tr>
<td>HoxA11-F1</td>
<td>GCCACACTGAGGACAAGG</td>
<td></td>
</tr>
<tr>
<td>HoxA11-R1</td>
<td>TTGAGACGCTTCTTGTGTT</td>
<td></td>
</tr>
<tr>
<td>HoxB2-F1</td>
<td>GAGAAGAATCCACCACACAGAACC</td>
<td></td>
</tr>
<tr>
<td>HoxB2-R1</td>
<td>AGCAGTTGCGTGTTGTGTGAG</td>
<td></td>
</tr>
<tr>
<td>HoxB9-F1</td>
<td>AGGAAGCGAGGACAAGAGAG</td>
<td></td>
</tr>
<tr>
<td>HoxB9-R1</td>
<td>CCTTCTTAGTCCACGCTCTCT</td>
<td></td>
</tr>
<tr>
<td>Pax6-F1</td>
<td>TTATTATCCGAGGGGTCTGT</td>
<td></td>
</tr>
<tr>
<td>Pax6-R1</td>
<td>CAGGGTGCAAGAGCTGTGTT</td>
<td></td>
</tr>
<tr>
<td>Gata6-F1</td>
<td>TTAACACTGATTCGCTGCAACG</td>
<td></td>
</tr>
<tr>
<td>Gata6-R1</td>
<td>GTTCATCGTAAACGTGCTGAG</td>
<td></td>
</tr>
<tr>
<td>T-F</td>
<td>CTGGGAGCTCTTCCTTTCTG</td>
<td></td>
</tr>
<tr>
<td>T-R</td>
<td>GAGGACGTGCAGCTGAGA</td>
<td></td>
</tr>
<tr>
<td>FoxA1-F1</td>
<td>GAAAGGCTAGCCAGCTAGAG</td>
<td></td>
</tr>
<tr>
<td>Gene</td>
<td>Primer Forward</td>
<td>Primer Reverse</td>
</tr>
<tr>
<td>-------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>FoxA1</td>
<td>AGATGCAGCTGAGATTCGTG</td>
<td></td>
</tr>
<tr>
<td>MyoD</td>
<td>GCCGCCTGACCAAAGTGAAT</td>
<td></td>
</tr>
<tr>
<td>MyoD</td>
<td>CAGCGGTCCAGGTGCTAAGAAG</td>
<td></td>
</tr>
<tr>
<td>Nkx2.2.1</td>
<td>AAGAGCCCTTTCTACGACAGC</td>
<td></td>
</tr>
<tr>
<td>Nkx2.2</td>
<td>GATTGGAACGTCAGTCTTGG</td>
<td></td>
</tr>
<tr>
<td>P16</td>
<td>GTCTTTGTGTACCGCTGGAAC</td>
<td></td>
</tr>
<tr>
<td>P16</td>
<td>TTAGCTCTGCTCTTGGGATT</td>
<td></td>
</tr>
<tr>
<td>Dazl</td>
<td>ATGGGCTCAGTAAAGAAGTGAAGATAA</td>
<td></td>
</tr>
<tr>
<td>Dazl</td>
<td>AGTACATAAAATTTGTCTTCTGATTGC</td>
<td></td>
</tr>
<tr>
<td>Spo11</td>
<td>ATCAGAAGCGATTCAACAAAA</td>
<td></td>
</tr>
<tr>
<td>Spo11</td>
<td>GGTTGCAGTTGCTGCTCTTGG</td>
<td></td>
</tr>
<tr>
<td>Nanos2</td>
<td>CCATCCTGAGGCACTATGTGT</td>
<td></td>
</tr>
<tr>
<td>Nanos2</td>
<td>CTGACTGCTGGAGGAGTGGAACA</td>
<td></td>
</tr>
<tr>
<td>Stra8</td>
<td>GCATGGTTCACCAGTGTTGGCCTT</td>
<td></td>
</tr>
<tr>
<td>Stra8</td>
<td>GCAGCTGCACCAATGGCTGG</td>
<td></td>
</tr>
<tr>
<td>Vasa</td>
<td>TGTGCTCTCCCACCTCAGTA</td>
<td></td>
</tr>
<tr>
<td>Vasa</td>
<td>TATTTCACTGTGGTTGGCCTC</td>
<td></td>
</tr>
<tr>
<td>Upp1</td>
<td>TGAAGCAAGGGACCTTGGAAAA</td>
<td></td>
</tr>
<tr>
<td>Upp1</td>
<td>GGTAGAGCAGCCTCCTCCTCACA</td>
<td></td>
</tr>
<tr>
<td>Klf4</td>
<td>CAG GCT GTG GCA AAA CCT AT</td>
<td></td>
</tr>
<tr>
<td>Klf4</td>
<td>CCT GTG TGT TGG CGG TAG TG</td>
<td></td>
</tr>
<tr>
<td>Ifitm3</td>
<td>AGCAGTTCTCATAAGGACCAA</td>
<td></td>
</tr>
<tr>
<td>Ifitm3</td>
<td>CGGTTCCTCAGAAGGTTGTCTT</td>
<td></td>
</tr>
<tr>
<td>Fkbp6</td>
<td>CGG CTG ATG AAA CTT GGA GA</td>
<td></td>
</tr>
<tr>
<td>Fkbp6</td>
<td>AGG CTG GCT TGA ACA GGA AC</td>
<td></td>
</tr>
<tr>
<td>Rpl3RTf</td>
<td>GATGAGTGTAAGAGGGCGCTTC</td>
<td></td>
</tr>
<tr>
<td>Rpl3Tr</td>
<td>CTTGGTGAAAGCCTTCTTCTTCTT</td>
<td></td>
</tr>
<tr>
<td>Rpl3RtintronR</td>
<td>TTTCAAAGGAAAACCAAGGAGG</td>
<td></td>
</tr>
<tr>
<td>Rps9RTf</td>
<td>CGTCTCGACGGAGCTAAA</td>
<td></td>
</tr>
<tr>
<td>Rps9Tr</td>
<td>CTTGGACCTCAGAAGTGCCTTGC</td>
<td></td>
</tr>
<tr>
<td>Rps9RtintronR</td>
<td>GATCGCAGAAATTCGAAATTCAC</td>
<td></td>
</tr>
<tr>
<td>Taf1dRTf</td>
<td>TGGATGATGATGATGCTTTC</td>
<td></td>
</tr>
<tr>
<td>Taf1dRTTr</td>
<td>GCCGTAGGATTGTGTTGCTTC</td>
<td></td>
</tr>
<tr>
<td>Taf1dRtintronR</td>
<td>AACCCCATTAAGGGCCCTCCTC</td>
<td></td>
</tr>
<tr>
<td>Actg1RTf</td>
<td>CCTGAACCCCAAGCTAACA</td>
<td></td>
</tr>
<tr>
<td>Actg1Tr</td>
<td>ACATGGGCTGGGTATGGGAAG</td>
<td></td>
</tr>
<tr>
<td>Actg1RtintronR</td>
<td>GTCGGGGCTCAAGCATAC</td>
<td></td>
</tr>
<tr>
<td>Eif4a2RTf</td>
<td>GAATCCATCAGGAGTCAAG</td>
<td></td>
</tr>
<tr>
<td>Eif4a2RTTr</td>
<td>CACTGTGTCACGTCAATCC</td>
<td></td>
</tr>
<tr>
<td>Eif4a2RTtintronR</td>
<td>AAAATACAGGAGGAGACACTC</td>
<td></td>
</tr>
<tr>
<td>prerDNAF</td>
<td>TGTCGTTTACACCTGTC</td>
<td></td>
</tr>
<tr>
<td>prerDNAR</td>
<td>AAATAAGGTGGCCCTCAACC</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>----------------------</td>
<td></td>
</tr>
<tr>
<td>L7-F</td>
<td>AGCGGATTGCCTTGACAGA</td>
<td></td>
</tr>
<tr>
<td>L7-R</td>
<td>AACTTGAAAGGCCACAGGAA</td>
<td></td>
</tr>
<tr>
<td>Ubb-F</td>
<td>GCGTTTGTGCTTCATCAC</td>
<td></td>
</tr>
<tr>
<td>Ubb-R</td>
<td>GGCACAGCTCTCTGCT</td>
<td></td>
</tr>
<tr>
<td>H2A-F</td>
<td>ACATGCGGCGCTGGAGTA</td>
<td></td>
</tr>
<tr>
<td>H2A-R</td>
<td>CGGGATGATGCGCGTCTTGTT</td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX II. PROTOCOL FOR LOW-CELL CHROMATIN IMMUNOPRECIPITATION (CHIP) OF HISTONES

Crosslinking in Solution from FACS

*Note: do not process more than a few tubes at once to ensure accurate incubation times*

**FACS** 10k-200k cells into 1.5ml DNA-LO Bind Eppendorf tube pre-filled with 400μl PBS

Determine total volume of collected cells using a micro-pipette

Bring volume to 1000μl with PBS

Cross-Link

Add 15.9μl of 16% FA stock to each sample (final concentration of 0.25%)

Immediately mix by inverting tube- do not mix via pipetting

Incubate for 10 min at room temp

Add 53.5μl of 2.5M Glycine stock to each sample (final concentration of 125mM)

Immediately mix by inverting tube- do not mix via pipetting

Incubate for 5 min at room temp

Centrifuge @ 470 x g for 10 min @ 4C in a swing bucket centrifuge (most of the sample will be lost with a fixed angle rotor)

Use soft deceleration

Carefully aspirate all but ~20μl of supernatant using P1000 and P100 pipettes (a pale pellet will be visible down to ~20k cells)

Wash with PBS

Add 500μl PBS and mix by inverting tube

Centrifuge @ 470 x g for 10 min @ 4C in a swing bucket centrifuge (most of the sample will be lost with a fixed angle rotor)

Use soft deceleration

Carefully aspirate all but ~20μl of supernatant using P1000 and P200 pipettes

Flash Freeze in liquid nitrogen and store at -80C

Samples are good for 3-6 months, but signal is better when used immediately or next day, followed by a gradual loss of signal-to-noise ratio.

Preparation/Sonication Of Chromatin

*Note: Preparation of the beads should be started before this (see below)*

*Note: this protocol uses the Covaris. However similar results are possible with the Bioruptor, but more difficult to make reproducible*

**Sonication**

Resuspend in 125μl shearing buffer (10mM Tris pH 7.6, 1mM EDTA, 0.1% SDS)

Transfer 130μl to Covaris snap cap glass sonication tube

Sonicate for 12 minutes total
Duty: 5%
Intensity: 3
Burst: 200
This should produce 6-7W of power stably for the 12 min
Clear Lysate
Transfer sample to a 1.5ml Lo-DNA Bind Eppendorf tube
Rinse ‘empty’ Covaris tube with 120µl lysis buffer and pool rinse with sample, discard Covaris tube
Add 200µl lysis buffer to sample
Centrifuge @ 12,000 x g for 10 min @ 4C
Transfer ~400µl supernatant to a fresh 1.5 Lo-DNA Bind tube
Add 320µl lysis buffer to ‘debris pellet’ and vortex
Centrifuge @ 12,000 x g for 10 min @ 4C
Collect ~335µl of supernatant and pool with previous supernatant
Should have ~735µl of cleared lysate- keep on ice
Split this lysate evenly between: INPUT, IP-IGG, IP-1, IP-2
Do not need an IP-IGG for ChIPseq

**Immunoprecipitation:** (for 3 IPs, plus pipetting overhead)

*NOTE: USE PROTEIN A DYNABEAD FOR ANTIBODIES RAISED IN RABBIT*

Preparation of magnetic beads
Add 500µl of lysis buffer to a 1.5ml DNA LoBind
Transfer 44µl of WELL RESUSPENDED beads to the 1.5ml tube and mix via inversion
Collect beads using a magnet
Resuspend beads in 500µl lysis buffer to wash
Repeat wash
Collect beads using a magnet
Resuspend in 400µl lysis buffer
Preparation of Bead/Ab complex
Add 100µl lysis buffer to 3x 0.2ml PCR tubes (1 tube per IP)
Transfer 100µl of resuspended beads to each PCR tube
Add 2.4µg of Antibody-1 to PCR tube 1, Antibody 2 to PCR tube 2 and, Antibody 3 to PCR tube 3
Incubate @ 4C for ~2 hours with end-over-end rotation
Formation of Bead/Ab/histone:DNA complex
Brief low speed spin of PCR tubes to remove liquid from caps
Pellet beads using magnetic rack
Working a few samples at a time (do not let beads dry out!), remove supernatant
Add 180µl of lysate to each PCR tube
Incubate PCR tubes overnight @ 4C with end-over-end rotation
Transfer 180µl of lysate to a 1.5ml DNA LoBind tube and store @ 4C as INPUT
**Washing Beads**

- Brief low speed spin to remove liquid from caps
- Pellet beads using magnetic rack
- Working a few samples at a time (do not let beads dry out!), remove supernatant
- Lysis buffer Wash
- Resuspend the beads in 200µl ice cold Lysis buffer
- Incubate PCR tubes @ 4C for 10 min with end-over-end rotation
- Brief low speed spin to remove liquid from caps
- Pellet beads using magnetic rack
- Aspirate supernatant
- Repeat two more times (3 lysis buffer washes total)
- DOC Buffer Wash
- Resuspend the beads in 200µl DOC Buffer
- Incubate PCR tubes @ 4C for 10 min with end-over-end rotation
- Spin/pellet as above
- Pellet beads using magnetic rack
- Aspirate supernatant
- TE Wash
- Resuspend the beads in 200µl TE
- Transfer samples to fresh PCR tubes, discard old tubes
- Incubate PCR tubes @ 4C for 5 min with end-over-end rotation
- Spin/pellet as above
- Pellet beads using magnetic rack
- Aspirate supernatant *with pipette*

**Elution**

- Add 150µL fresh **Elution Buffer**
- Incubate PCR tubes @ room temp for 20 min with end-over-end rotation
- Spin/pellet as above
- Pellet beads using magnetic rack
- Transfer supernatant to fresh 1.5ml DNA LoBind tube
- Repeat elution with another 150µl elution buffer for 20 min
- Pool 2nd elution with 1st
- 300µl total elution volume

**Reversal of Crosslinking**

- Bring inputs to 300uL with elution buffer
Add the following to inputs and IP eluates:
Final concentrations: 200mM NaCl, 0.2 mg/mL RNase A, 10mM EDTA, 40mM Tris pH 6.5
For 300ul of IP/INPUT:
12uL of 5M NaCl
3uL of RNaseA 20mg/mL
7.5uL 0.4M EDTA
12uL Tris pH 6.5 1M
Incubate 30 min at 37C
Add 1µl of Proteinase K 20mg/mL stock
To reverse crosslink: incubate in programmed eppendorf shaker for 3h at 55 degrees and then overnight at 65C

Purification of DNA:
Use Qiagen Min-Elute PCR Purification tubes as per manufacturer’s instructions
Diluted each sample in 1.6ml of PB buffer
Elute with 12µl H2O
Quantify INPUT sample using Qubit fluorometer
IPs usually cannot be detected with under 100k cells per IP
### Solution Recipes:

**Lysis Buffer (300mM NaCl):** 500ml total  
- 50 mM HEPES/KOH pH 7.5  
- 300 mM NaCl  
- 1 mM EDTA  
- 1 % Triton X-100  
- 0.1 % DOC  
- 0.1 % SDS  
- 384ml H2O  

Add fresh: protease inhibitors

**DOC buffer:**  
- 10 mM Tris pH 8  
- 0.25 M LiCl  
- 0.5 % NP-40  
- 0.5% DOC  
- 1 mM EDTA

**Elution buffer (make fresh in ultra-clean H2O for every elution):**  
- 1% SDS  
- 0.1 M NaHCO3
**APPENDIX III. QRT-PCR: FROM CELLS TO EXCEL**

- This protocol is a guide for determining relative fold changes in RNA levels for a gene between different conditions. This is NOT an absolute quantification and does not give copy number information or allow for comparison of expression levels across experiments.
- In brief, cDNA is generated from RNA isolated from a culture and SYBR-green based quantification measures the RNA level of a gene of interest relative to a reference housekeeping gene. Fold change of the gene of interest between different conditions is calculated by normalizing to the reference gene expression levels.
- The major steps are:
  - RNA isolation
  - cDNA generation
  - qPCR plate setup
  - Thermo cycling and data analysis
- Usually this is a two day process, although it is possible to perform in a single day.
RNA Isolation

**Materials**

- RNeasy Mini RNA isolation kit (Qiagen)
- QIAshredder (Qiagen)
- RNase-Free DNase Set (Qiagen)
- DNasel, Amplification Grade (Invitrogen)
- 1 confluent well of a 12 well plate or equivalent (approx. 1x10^6 cells)
  - Fewer cells can be used if examining a highly expressed gene. If studying a repeat element or other species with a high genetic copy number, 1x10^6 cells should be used
- Fresh RLT buffer
  - Prepare by adding 10µl BME to 1ml RLT buffer (Qiagen)
  - Good for 1 month
- On-Column DNase solution (Qiagen)
  - For each sample mix 70µl RDD buffer with 10µl DNasel
  - Prepare fresh and store on ice

**Protocol**

- **Harvest Cells**
  - Aspirate media from cells
  - Add 350µl of RLT buffer to well and disrupt cells by scraping with pipette tip and pipeting up and down for 30”. Check no clumps are visible
  - Vortex plate for 1’
  - Flash freeze on dry ice
    - Can store samples @ -80°C at this point
- **RNA extraction**
  - Thaw the sample on ice
  - Transfer to QIAshredder spin column and centrifuge @ 13k xg for 2’
    - Collect in sample in supplied 2ml tube
  - Add 1 volume (350µl) 70% ethanol (RNase free) and mix by pipeting
  - Transfer the sample to an RNeasy column and centrifuge @ 13k xg for 30.” Discard flow through in “BME waste tube”
- **1st DNase Treatment (On-Column)**
  - Add 350µl RW1 buffer and centrifuge @ 13k xg for 30”. Discard flow through
  - Add 80µl ‘On-Column DNase Solultion’ and incubate at room temp. for 15’
  - Add 350µl RW1 buffer and centrifuge @ 13k xg for 30”. Discard flow through
- **RNA Wash**
  - Transfer the RNeasy column to a new collection tube
  - Add 500µl RPE buffer and centrifuge @ 13k xg for 30”. Discard flow through
    - Repeat above RPE buffer wash
  - Transfer RNeasy column to a new collection tube
- Centrifuge @ max speed for 2’ to remove residual ethanol

- **RNA elution**
  - Transfer the RNeasy column to a RNase free 1.5ml microfuge tube
  - Add 30µl ddH₂O (RNase free) directly to column membrane and incubate at room temp for 1’
  - Centrifuge @ 13 xg for 1’ and keep flow through. Discard column
  - Measure concentration using NanoDrop
    - Expected yield of 5-10µg RNA total per 1x10⁶ cells

- **2nd DNase Treatment (In-Solution)**
  - *This step is not necessary if studying species with low genomic copy number (i.e. most genes)*
  - Sample size can be scaled up or down linearly
  - For each sample mix in a PCR tube:
    - 3µg RNA (max of 24µl)
    - 3µl 10x DNase I Reaction buffer
    - 3µl DNase I, Amp Grade (1U/µl)
    - ddH₂O (RNase free) up to 30µl
  - Incubate at room temp. for exactly 15’
    - *Do NOT exceed 15’ incubation*
  - Add 3µl 25mM EDTA and mix to stop reaction
  - Incubate @ 65°C for 10’ to heat inactivate
  - RNA is ready for use in cDNA synthesis reaction

- Store all RNA samples @ -80°C, avoid repeated freeze thaws

### cDNA Synthesis

**Materials/ Equipment**

- High-Capacity cDNA Reverse Transcription Kit (Applied BioSystems)
- Thermo cycler

**Protocol**

- Determine the number of reactions to be performed
- Prepare 2x Master Mix for reactions
  - Per reaction:
    - 2µl 10x RT buffer
    - 0.8µl 25x dNTP mix (100mM)
    - 2µl 10x RT random primers
    - 1µl MultiScribe Reverse Transcriptase
    - 4.2µl nuclease free H₂O
  - Note: include 10% overhead to compensate for loss during pipeting (i.e. if 8 reactions are needed, prepare a master mix for 8.8 reactions)
  - Mix gently and store on ice
- Determine the number of –RT reactions to be performed
Note: This is essential when genomic DNA contamination is a concern (i.e. for repeat elements). In general, every qRT-PCR primer set should be tested against at least one –RT cDNA from cDNAs sets generated from similar cell types.

- Prepare 2x Master Mix for all –RT reactions
  - Per –RT reaction:
    - 2µl 10x RT buffer
    - 0.8µl 25x dNTP mix (100mM)
    - 2µl 10x RT random primers
    - 5.2µl nuclease free H2O
  - Note: include 10% overhead to compensate for loss during pipeting (i.e. if 8 reactions are needed, prepare a master mix for 8.8 reactions)
  - Mix gently and store on ice

- Dispense 10µl of each Master Mix into a 200µl PCR tube for each reaction and/or –RT reaction
  - Store on ice

- Prepare RNA
  - Dilute 1µg of DNase treated RNA into 10µl nuclease free water for each reaction
  - Add 10µl of RNA to Master Mix for each reaction and mix by pipetting twice
  - Seal the tubes and briefly centrifuge to remove bubbles
  - Store reactions on ice until loading into thermo cycler

- Program for Thermo cycler
  - Step 1: 25°C 10’
  - Step 2: 37°C 120’
  - Step 3: 85°C 5’
  - Step 4: 4°C forever

- After cycling, dilute each reaction with 180µl nuclease free H2O
- cDNA is now ready for qPCR
- Store @ -20°C

qPCR Plate Setup
- Here choice of primers, layout of reactions in a 384 well plate, and qPCR reaction setup will be addressed

Primer Set Choice

- All primer sets must work at the same annealing temperature to be run on a single plate
  - Ideally, a primer set should yield a single sharp peak in its melting curve (see analysis section below for more detail)
- At least one, but preferentially two housekeeping genes should be used as reference primer sets
  - L7 works well and is robust from 50°C to 65°C annealing temperatures
  - Ubb work well from 50°C to ~58°C. A secondary amplicon becomes prevalent at higher temperatures

Plate Layout
• Each reaction (i.e. primer set/cDNA pair) should be performed at least in duplicate and preferentially in triplicate. This includes reactions for reference genes
• It is helpful to map out all reactions in a color coded table or Excel file for complex setups

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>...</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Oct4</td>
<td>Oct4</td>
<td>Oct4</td>
<td>L7</td>
<td>L7</td>
<td>L7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>Oct4</td>
<td>Oct4</td>
<td>Oct4</td>
<td>L7</td>
<td>L7</td>
<td>L7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Oct4</td>
<td>Oct4</td>
<td>Oct4</td>
<td>L7</td>
<td>L7</td>
<td>L7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>Oct4</td>
<td>Oct4</td>
<td>Oct4</td>
<td>L7</td>
<td>L7</td>
<td>L7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>Oct4</td>
<td>Oct4</td>
<td>Oct4</td>
<td>L7</td>
<td>L7</td>
<td>L7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>Oct4</td>
<td>Oct4</td>
<td>Oct4</td>
<td>L7</td>
<td>L7</td>
<td>L7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>Oct4</td>
<td>Oct4</td>
<td>Oct4</td>
<td>L7</td>
<td>L7</td>
<td>L7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>Oct4</td>
<td>Oct4</td>
<td>Oct4</td>
<td>L7</td>
<td>L7</td>
<td>L7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Materials/Equipment

• ABI 7900HT Sequence Detection System
• 384 well plate centrifuge capable of 300x g
• SYBR Green PCR Master Mix (Applied BioSystems)
• MicroAmp Optical 384 well Reaction Plate (Applied BioSystems)
• MicroAmp Optical Adhesive Film (Applied BioSystems)
• Plate Sealing tool (Applied BioSystems)

Protocol

• Prepare 5µM stocks of all primer sets to be used
  o A single solution with both the forward and reverse primers at 5µM
• Calculate number of reactions for each cDNA
  o In the example above, this would be 6
  o Prepare a ‘cDNA master mix’ for each cDNA
    ▪ For each reaction:
      • 0.4µl cDNA
      • 4.2µl nuclease free H₂O
  o Note: include 10% overhead to compensate for loss during pipetting (i.e. if 6 reactions are needed, prepare a master mix for 6.6 reactions)
  o Store cDNA master mixes on ice

• Calculate number of reactions for each primer set
  o In the example above, this would be 24
  o Prepare a ‘primer master mix’ for each primer set
    ▪ For each reaction:
      • 0.4µl 5µM primer stock (200nM final)
      • 5µl 2x SYBR Green PCR Master Mix (Applied BioSystems)
  o Note: include 10% overhead to compensate for loss during pipetting (i.e. if 24 reactions are needed, prepare a master mix for 26.4 reactions)
  o Store primer master mixes on ice

• Dispense 4.6µl of the cDNA master mixes into appropriate wells of a MicroAmp Optical 384 well Reaction Plate
  o A multichannel pipette and 8 tube strips are helpful in expediting the process

• Dispense 5.4µl of the primer master mixes into appropriate wells of the 384 well plate
  o Use a fresh tip for each well and pipette twice to mix with cDNA
  o A multichannel pipette and 8 tube strips are helpful in expediting the process

• Seal the plate using MicroAmp Optical Adhesive Film (Applied BioSystems)
  o Apply with the plate sealing tool working from center to each corner. Pay close attention to regions on edge of plate

• Centrifuge plate @ 2000x g @ 4°C for 2’ to remove bubbles

• Store plate on ice until loading into thermo cycler
  o Plate can be stored for several hours without significant effects on PCR efficiency
Thermo cycling and Data Analysis

- This describes how to navigate and program the ABI 7900HT and ABI’s ‘SDS’ software
  - The machine we use is located on the 16th floor in the Blelloch lab

Plate Setup

- To turn on the machine press the large button on the lower right of its front side
- Start the ‘SDS v2.3’ program
  - The user name is ‘administrator’
  - There is no password
- To create a new plate, select new document under the File menu
  - Select the following:
    - Assay: Standard Curve (AQ) *(Ct values are easiest to export in this mode)*
    - Container: 384 Wells Clear Plate
    - Template: Blank template
  - After clicking OK, you will be taken to the Workspace
To instruct the machine to detect SYBR green, click on ‘Add detector…’

On the pop up screen find and select ‘SYBR1.’ Then click ‘Copy to Plate’ followed by ‘Done’
To instruct the machine to detect SYBR green in all wells:

- Select all wells by clicking on the grey box above ‘A’ and left of ‘1’ in the graphic plate layout
- Then check the box to use ‘SYBR1’
- Notice all wells will have a green bar under them that corresponds to the color of ‘SYBR1’
To omit wells that are empty, select unused wells and click the ‘Omit Well(s)’ box.

- The machine will not read data from these wells.

Optional: To assign names to the wells, select a well(s), then type a name in the box ‘Sample Name’.

The plate is now setup.
Programming the Thermo Cycler

- Click the ‘instrument’ tab to display the thermo cycling program
• Change the annealing temperature to match the melting temperature of your primers by typing in the text box under stage 3
  o Typical melting temperatures range between 55°C and 65°C
  o *A proper annealing temperature may have to be experimentally determined*
• Add a dissociation stage by clicking on ‘Add dissociation Stage’
  o This will add a ‘Stage 4’ at the end of the program which very slowly heats the reaction from 60°C to 95°C

• The Thermo cycler is now programmed
• To save the plate setup and thermo cycling conditions, select ‘Save As’ under the File menu and save the file in the ‘User’ directory with an appropriate name
Loading the PCR Plate and Starting the Program

- To connect the software to the machine, click on the ‘Real-Time’ tab and then click on ‘Connect to Instrument’
• To load the experimental plate and start the program, click ‘Open/Close’
  o The machine will slowly open and swing out a plate cradle
  o Remove the space holder plate (*this prevents dust from accumulating on the heating block*)
  o Place the experimental plate in the cradle with well A1 orientated to the top left
  o Click ‘Start Run’
    ▪ The plate cradle will swing back in and the door will slowly close

• The qPCR reaction is now running
  o After a few minutes the software will display an estimated completion time
    ▪ Under standard conditions, the machine will run for approx. 2 hours
• At end of run click ‘Open/Close’
  o Once the plate cradle swings out, remove experimental plate
  o Replace space holder plate and click ‘Open/Close’ again
• After plate cradle swings back in and door closes, click ‘Disconnect from instrument’
• To turn off the machine press the large button on the lower right of its front side
Analysis of Melting Curves for each Primer Set

- The data obtained from qPCR is only valid if the melting curves are sharp, single peaks
  - Melting curves with secondary peaks and/or broad peaks may indicate the presence of non-specific amplicons, therefore obscuring levels of the desired amplicon
- To view the melting curves after the run completes:
  - click the ‘analyze data’ button
Then select the 'Dissociation Curve' tab

To view the melting curve for a particular primer set, select the corresponding wells in the plate layout panel.
• A good melting curve (here for L7) looks like this:

• Examples of problematic melting curves:

  • ‘Bad’ melting curves can sometimes still be used as long as the curve is the same across all cDNAs
  • A major concern is if different cDNAs have different shaped melting curves
    o This indicated different amplicon formations and complicates comparison
  • The melting curve for each primer set should be checked
    o It is expected that –RT and water cDNAs will have abnormal melting curves
**Exporting Ct values**

- Relative quantification using Ct values must be accomplished using other software
  - REST is an Excel macro shareware that is convenient for analyzing experiments with multiple cDNAs (up to 6 cDNAs plus a reference cDNA)
- To export the raw Ct values from the SDS software
  - select ‘Export’ from the ‘File’ menu
Choose a file name and location to save the exported data

Check the box labeled ‘Current Table Only’

- Click ‘Export’ to complete the process
- The file format of the exported data is a .txt file although it is easier to manipulate the data in Excel
- There are several programs and statistical formulas that can be used to analyze qPCR data
  - REST is convenient for experiments with several cDNAs, but less than 9 primer sets
  - There are variations of REST available for analysis of large numbers of primer sets, but they can only handle 2 cDNAs
  - Analysis of a large number of cDNAs and primer sets may require specialized software or custom Excel macros
  - A standard formula for calculating fold change of a target gene relative to a reference condition is:

\[
R = \left( \frac{E_{\text{target}}}{E_{\text{ref}}} \right)_{\text{MEAN control}} \left( \frac{E_{\text{target}}}{E_{\text{ref}}} \right)_{\text{MEAN sample}}
\]

Where:

- \( E \) = efficiency of primer set amplification (assume ‘2’)
- \( CP \) = Ct value
- Target = target gene
- Ref = Reference gene (or housekeeping primer set)
- Control = Reference condition (or control cDNA)
PUBLISHING AGREEMENT

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

[Signature]
Author Signature

[Date]
12.30.2013