2017

The Human Heterochromatin Landscape: Genomic Subtypes, Bound Proteins, And Contributions To Cell Identity

Justin S. Becker
University of Pennsylvania, justinsbecker@gmail.com

The Human Heterochromatin Landscape: Genomic Subtypes, Bound Proteins, And Contributions To Cell Identity

Abstract
Large portions of mammalian genomes are packaged into structurally compact heterochromatin, which protects genome integrity and suppresses transcription of lineage-inappropriate genes. Characterization of heterochromatic regions has relied on genomic mapping of associated histone modifications, such as H3K9me3 and H3K27me3, and purification of proteins interacting with these modifications. Heterochromatic regions marked by H3K9me3 have been shown to impede gene activation during reprogramming to pluripotency, and I find that H3K9me3 domains can similarly impede conversion of fibroblasts to hepatocytes. However, both H3K9me3 and H3K27me3 can be found in transcriptionally active chromatin, limiting the accuracy of histone marks alone for identifying heterochromatin domains or bound proteins that impede reprogramming. I developed a biophysical method to purify heterochromatic regions, using sucrose gradients to isolate chromatin fragments that are resistant to sonication. Sequencing of the purified material (Gradient-seq) revealed the genomic landscape of structural heterochromatin in human fibroblasts, which is transcribed at low levels and contains largely distinct H3K9me3 and H3K27me3 domains, as well as unmarked regions. Gradient-seq also uncovered subtypes of H3K9me3 and H3K27me3 domains that are structurally euchromatic, a distinction corroborated by increased gene transcription, hypomethylation at CpG islands, decreased association with the nuclear lamina, and increased activation during hepatic reprogramming. Using quantitative proteomics, we found 172 proteins associated with heterochromatin after gradient sedimentation and H3K9me3-directed IP. The identified proteins include known transcriptional repressors and are enriched for proteins shown to impede reprogramming to pluripotency. We show that the RNA-binding protein RBMX, one of the proteins most enriched by gradient sedimentation and H3K9me3 IP, is a functional regulator of heterochromatin. RBMX and the related protein RBMXL1 are required for silencing of select heterochromatinized genes, and depletion of these proteins in fibroblasts renders H3K9me3-marked hepatocyte genes more competent for activation during reprogramming to the hepatic lineage. Thus, our biophysical method for heterochromatin isolation has allowed us to create a genome-wide map of chromatin compaction in human cells, to identify chromatin domain subtypes that impede conversion between differentiated lineages, and to discover novel heterochromatin proteins that contribute to this reprogramming barrier.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Cell & Molecular Biology

First Advisor
Kenneth S. Zaret

Keywords
Gradient-seq, H3K27me3, H3K9me3, heterochromatin, proteomics, reprogramming

Subject Categories
Cell Biology | Genetics | Molecular Biology

THE HUMAN HETEROCHROMATIN LANDSCAPE: GENOMIC SUBTYPES, BOUND PROTEINS, AND CONTRIBUTIONS TO CELL IDENTITY

Justin S. Becker

A DISSERTATION

in

Cell and Molecular Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2017

Supervisor of Dissertation

_____

Kenneth S. Zaret, PhD

Joseph Leidy Professor of Cell and Developmental Biology

Graduate Group Chairperson

_____

Daniel S. Kessler, PhD, Associate Professor of Cell and Developmental Biology

Dissertation Committee

Christopher J. Lengner, PhD, Assistant Professor of Biomedical Sciences (Chair)

Gerd Blobel, MD, PhD, Frank E. Weise III Professor of Pediatrics

Eric J. Brown, PhD, Associate Professor of Cancer Biology

Ben Z. Stanger, MD, PhD, Associate Professor of Medicine

ACKNOWLEDGMENTS

In my PhD work, as in most things, I am most indebted to my mom, who recognized that I was a scientist before I even knew it myself. It was she who recommended I seek out research opportunities as a high school student and who is responsible for sparking a passion that has guided my life ever since. Research has given me an opportunity to emulate what I admired so much in my dad—the fact that, through his medical career, he could use his mind in a way that helped people. I am lucky to have parents, my sister Lauren and my grandma, Omi, who are so supportive of my scientific career. Science can have its ups and its many downs, but I am in the rare position of having a family who makes me feel like, whether the experiments work or not, I'm doing the most important thing in the world.

My girlfriend Sam has been with me for the duration of my PhD, and though she never picked up a pipette, she has earned her fair share of this degree. Certainly some of the burdens of scientific research—the long hours, the uncertainty, the second-hand stress—have been borne by her in equal measure. And yet I don't think it's possible for another human to be rooting for my success any more than she has. Thanks, Sam, for indulging my late-night updates when something exciting happens in lab, and for so reliably picking me back up when the science gods break my heart.

The opportunity to work with my mentor, Ken Zaret, for the past four years has been an absolute privilege. In addition to his expertise and creativity, Ken brings an excitement and enthusiasm to our work that makes doing science fun, which is a major reason why I still want to make research my life's calling. He has a skill for distilling the fundamental biological concepts that are in question and designing experiments to get at

their core, and so I've been fortunate to have the sort of project that I wake up excited to work on every day. I've learned a great deal from Ken about project management and how to prioritize experiments, and our back-and-forth dialogues have a way of bringing out new ideas that neither of us entered the conversation with. Most importantly, nothing Ken has ever done has made me question whether he supported me or "had my back," and his concern for my well-being and success as a person always came ahead of the science. I could go on, but Ken has also been trying to teach me the virtues of brevity in writing.

The Zaret lab has been a supportive and inspiring, and also very entertaining, home during my PhD. The people in the lab make for that rare kind of environment where people give each other the benefit of the doubt, and where you can argue at one minute and joke around at the next. I thank my labmates for looking past my unfortunate pre-coffee (or empty-stomach) disposition and being my friends anyway. I've particularly enjoyed the evening science chats over the years with my fellow graduate school travelers Kate, Jessica, and Meilin, which have helped me to stay inspired and made me realize that we all got into this for the same reasons. In a stroke of supreme luck, the other person in the lab focused on H3K9me3 heterochromatin was Dr. Dario Nicetto, one of the most rigorous scientists I've ever met and also a terrific guy, with whom I've had the honor of exchanging scientific ideas.

I'm greatly indebted to the other scientists who contributed to my thesis research. Kelsey, Ryan, and Zhiying in the Zaret lab are all extremely careful scientists who quickly earned my trust, who managed to tolerate my obsessive approach to experiments, and who most importantly brought an enthusiasm to the questions we were pursuing. I am incredibly grateful for Greg's mentorship and guidance on bioinformatics and genomics; without him, none of the computational analyses in my thesis would have been possible,

and I would have left my graduate studies without this important skillset. Finally, one of the most important aspects of my research depended on the expertise of Simone, Shu, and Dr. Garcia down the hall, who carried out the proteomics analysis, contributed valuable ideas for refining our approach, and made themselves repeatedly available to us on a moment's notice.

I also want to thank my previous scientific mentors, Dr. Thomas Hintze and Dr. David Pellman, for taking a chance on me and for allowing me to be intellectually involved in the management of projects as soon as I started working in their lab. It was working with them that first allowed me to experience the thrill of discovery, which is the reason I'm here.

Maggie and Skip of the Penn MD/PhD program have been incredible resources to me, particularly when my stress was at its peak. Their professional advice and interpersonal support is one reason among many why there is no better place to train as a physician-scientist.

My housemates Sarah and Megan got to know me during the final busiest stretches of my PhD, and yet they were so immediately supportive and rallied behind me when I was feeling overwhelmed. Thank you for creating a home that I'm always excited to come back to after lab.

Finally, to my colleagues Sarah, Andrew, Ted, Liz, Alice, and Ray – your example and your brilliance has inspired me throughout my PhD, and I am incredibly grateful that I get to call you friends.

ABSTRACT

THE HUMAN HETEROCHROMATIN LANDSCAPE: GENOMIC SUBTYPES, BOUND PROTEINS, AND CONTRIBUTIONS TO CELL IDENTITY

Justin S. Becker

Kenneth S. Zaret

Large portions of mammalian genomes are packaged into structurally compact heterochromatin, which protects genome integrity and suppresses transcription of lineage-inappropriate genes. Characterization of heterochromatic regions has relied on genomic mapping of associated histone modifications, such as H3K9me3 and H3K27me3, and purification of proteins interacting with these modifications. Heterochromatic regions marked by H3K9me3 have been shown to impede gene activation during reprogramming to pluripotency, and I find that H3K9me3 domains can similarly impede conversion of fibroblasts to hepatocytes. However, both H3K9me3 and H3K27me3 can be found in transcriptionally active chromatin, limiting the accuracy of histone marks alone for identifying heterochromatin domains or bound proteins that impede reprogramming. I developed a biophysical method to purify heterochromatic regions, using sucrose gradients to isolate chromatin fragments that are resistant to sonication. Sequencing of the purified material (Gradient-seq) revealed the genomic landscape of structural heterochromatin in human fibroblasts, which is transcribed at low levels and contains largely distinct H3K9me3 and H3K27me3 domains, as well as unmarked regions. Gradient-seq also uncovered subtypes of H3K9me3 and H3K27me3 domains that are structurally euchromatic, a distinction corroborated by increased gene transcription, hypomethylation at CpG islands, decreased association with the nuclear lamina, and

increased activation during hepatic reprogramming. Using quantitative proteomics, we found 172 proteins associated with heterochromatin after gradient sedimentation and H3K9me3-directed IP. The identified proteins include known transcriptional repressors and are enriched for proteins shown to impede reprogramming to pluripotency. We show that the RNA-binding protein RBMX, one of the proteins most enriched by gradient sedimentation and H3K9me3 IP, is a functional regulator of heterochromatin. RBMX and the related protein RBMXL1 are required for silencing of select heterochromatinized genes, and depletion of these proteins in fibroblasts renders H3K9me3-marked hepatocyte genes more competent for activation during reprogramming to the hepatic lineage. Thus, our biophysical method for heterochromatin isolation has allowed us to create a genome-wide map of chromatin compaction in human cells, to identify chromatin domain subtypes that impede conversion between differentiated lineages, and to discover novel heterochromatin proteins that contribute to this reprogramming barrier.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

PREFACE

This dissertation is a series of chapters organized around published and submitted manuscripts of which I am the primary author. Chapter 2 presents a review article that was published in *TRENDS in Genetics* in January 2016. Chapter 4 presents a research manuscript that has been submitted for publication. The experimental and analytical methods used during my thesis research, in particular for the manuscript in Chapter 4, are presented Chapter 3. These chapters all closely resemble the text of the submitted or published work, with alterations to reflect the format and organization of the thesis. The opening chapter discusses the relevant literature for the topics in covered in this dissertation, including a description of gaps in the field prior to my work, which goes beyond the more limited scope of the review in Chapter 2. Meanwhile, Chapter 5 presents the major conclusions from my thesis research and contextualizes them in relation to existing literature, while also presenting avenues for future investigation. Included as an Appendix at the end of the document is an additional manuscript to which I contributed as third author during my graduate work, which was published in *Genes and Development*.

CHAPTER 1: INTRODUCTION

In multicellular organisms, a fixed complement of chromosomes is able to instruct the development of the hundreds of different cell types, each with distinct morphologies and functions, which makes possible the complex physiologies of eukaryotes. As an organism develops, there is generally no change in the sequence of DNA or the number of genes, and thus the diversity of cell fates must be achieved through the selective activation and repression of different groups of genes, with each cell type having a different signature of gene activity. In humans, it is estimated that less than 60 percent of the genomic sequence with evidence for transcription is copied into RNA in a given specific cell type (Djebali et al., 2012). Thus, the mechanisms that selectively induce or suppress gene activity are central to understanding how diverse cell types are produced during the course of normal development or tissue regeneration (Zaret, 2008; Hemberger et al., 2009; Adam and Fuchs, 2016). Subsequent conversion of cells to a different cell type of interest, such as for therapeutic replacement of tissue destroyed by disease (Wu and Hochedlinger, 2011), requires a rewiring of the gene expression signatures that are established during development, to activate genes required for the function of the new cell type and extinguish expression of inappropriate genes (Vierbuchen and Wernig, 2012; Papp and Plath, 2013).

A gene's activity in any given cell type is closely related to the physical structure of the chromatin fiber, with active regions being associated with a more open and accessible conformation, while transcriptionally silent regions of DNA exhibit a denser packaging of nucleosomes. Physical compaction of chromatin is understood to play a causal role in gene silencing, by reducing the accessibility of the DNA template to binding

by transcription factors and other machinery necessary for transcription (Workman and Kingston, 1998; Li et al., 2007). On the spectrum of chromatin accessibility, regions with the highest level of compaction are termed “heterochromatin,” which classically have been identified cytologically by their enhanced uptake of DNA dyes and dark appearance in an electron microscope (Heitz, 1928; Brown, 1966; Underwood et al., 2016). By contrast, “euchromatin” refers to the portion of chromatin that has a more open or decondensed physical structure, appearing light in electron micrographs, and where the majority of transcriptional activity is confined.

Heterochromatin formation occurs and has been most frequently studied at highly repetitive, noncoding regions of the genome, such as regions flanking mammalian centromeres and telomeres (Pardue and Gall, 1970; Fanti et al., 1998; Guenatri et al., 2004; Probst and Almouzni, 2008), or at sequences with propensity for parasitic self-duplication, such as DNA- or RNA-based transposable elements (Pimpinelli et al., 1995; Kazazian, 2004). The packaging of such loci into dense heterochromatin structures prevents mutagenic recombination between repetitive sequences (Fanti et al., 1998; Peters et al., 2001) and suppresses the activity of transposons (Slotkin and Martienssen, 2007; Montoya-Durango et al., 2009). Thus, heterochromatin serves to protect genome integrity in addition to silencing inappropriate transcription (Eden et al., 2003; Peng and Karpen, 2008). In metazoans, such repeat-rich regions form heterochromatin in most or all cell types in the body, and thus are termed “constitutive heterochromatin” (Brown, 1966; Saksouk et al., 2015). By contrast, regions of “facultative heterochromatin” are dynamic during development, forming at particular stages or in a lineage-specific fashion (Brown, 1966; Trojer and Reinberg, 2007). One classic example of facultative heterochromatin is the inactivation of one of the two X chromosomes in females, which is silenced in early

development to maintain proper gene dosage and which appears as a compacted, electron dense Barr body (Heard, 2005; Peters et al., 2002; Rego et al., 2008). Facultative heterochromatin also plays an important role in cell fate control, by stably repressing lineage-inappropriate gene expression (Trojer and Reinberg, 2007; Allan et al., 2012; Xu et al., 2014a; Snitow et al., 2016). The hallmark chromatin staining pattern of particular cell types, such as the “clock-faced chromatin” of antibody-secreting plasma cells (Jinn-Fei and el-Labban, 1986), illustrates the dramatic and specific rearrangements in condensed heterochromatin that can accompany cell differentiation.

1.1 Properties of heterochromatin and role in cell identity

Defining features of heterochromatin include its physically dense structure and reduced transcriptional output. Early biophysical studies showed that the darkly staining heterochromatin masses in vertebrate nuclei had a structure that was more resistant to mechanical shearing with a sonifier or pressure cell, and that these regions could be physically isolated from the more heavily sheared euchromatic fibrils by differential centrifugation (Doenecke and McCarthy, 1975; Frenster et al., 1963). These studies showed that the shearing-resistant heterochromatin structures had minimal transcriptional activity, contained repetitive DNA satellites, had a higher ratio of histone protein to DNA, and were dependent on linker histone for their resistance to shearing (Doenecke and McCarthy, 1975; Duerksen and McCarthy, 1971; Frenster et al., 1963). The condensed nature of heterochromatic sequences is more typically assessed by resistance to cleavage by nucleases such as DNase I and II and micrococcal nuclease (MNase) (Gottesfeld et al., 1975; Wallrath and Elgin, 1995; Hamid et al., 1996), where the limited activity of the

nucleases is assumed to mirror inaccessibility to transcriptional machinery. Moreover, genetic studies of position-effect variegation (PEV) showed that compact heterochromatic structures, when placed adjacent to active genes via chromosomal rearrangement, can spread laterally along chromosomes into nearby loci and dominantly suppress transcription (Tartof et al., 1984; Wallrath and Elgin, 1995). This phenomenon was classically observed in the case of the *white* gene in *Drosophila* fruit flies, where juxtaposition of the gene near pericentromeric or telomeric sequences caused a mosaic pattern of gene expression and eye color in the developing fly (Tartof et al., 1984; Wallrath and Elgin, 1995). More recent work has recapitulated this phenomenon in mammalian systems using reporter transgenes that can be silenced by proximity to both repetitive and gene-containing forms of heterochromatin (Ayyanathan et al., 2003; Tchasovnikarova et al., 2015). Additional conserved properties of heterochromatin include spatial proximity to the nuclear lamina (Csink and Henikoff, 1996; Guelen et al., 2008; Poleshko et al., 2013) or nucleoli (Sadoni et al., 1999; van Koningsbruggen et al., 2010), as well as replication late in S phase (O'Keefe et al., 1992).

Among the diverse post-translational modifications that are observed on histone tails, di- and tri-methylation at the specific residue histone 3 lysine 9 (H3K9me2 and H3K9me3) is highly enriched over repeat-rich heterochromatic sequences in organisms ranging from *Schizosaccharomyces pombe* to humans (Nakayama et al., 2001; Schotta et al., 2002; Peters et al., 2003; Martens et al., 2005). The gene *Su(var)3-9* was discovered as a powerful suppressor of PEV in *Drosophila* with dominance over most PEV enhancers (Tschiersch et al., 1994), and it encodes a SET domain-containing enzyme that catalyzes the demethylation and trimethylation of H3K9, with a particular role in pericentromeric H3K9me3 (Rea et al., 2000; Czermin et al., 2001; Schotta et al., 2002). There are two

mammalian homologues of this enzyme, SUV39H1 and SUV39H2, which are required for the integrity of pericentromeric heterochromatin, the silencing of satellite repeat transcription, and the accuracy of mitotic and meiotic chromosome segregation (Aagaard et al., 1999; Peters et al., 2001). *Suv39h1/h2* double-mutant mice show perinatal lethality after E12.5, reduced size, hypogonadism in males, and increased tumorigenesis (Peters et al., 2001). In mammals, in addition to the SUV39H proteins, the enzyme SETDB1/ESET also catalyze H3K9me2 and H3K9me3 deposition (Schultz et al., 2002), while G9a/EHMT2 and GLP/EHMT1 contribute to H3K9me2 only (Tachibana et al., 2001, 2002, 2005). SETDB1 plays in important role in pluripotent cells by repressing transcription factor genes for developmental lineages and trophectoderm (Bilodeau et al., 2009; Yeap et al., 2009; Yuan et al., 2009; Lohmann et al., 2010), while G9a and GLP are required for the stable silencing of pluripotency genes like *Oct3/4* during early differentiation (Feldman et al., 2006; Epsztejn-Litman et al., 2008; Liu et al., 2015). Consistent with these observations, deletion of *Setdb1* in mice causes lethality around the time of implantation (Dodge et al., 2004), while loss of either G9a or GLP in mice is lethal by day E12.5 (Tachibana et al., 2002, 2005).

The H3K9me2/3 marks are not known to directly interfere with RNA transcription or directly affect chromatin structure, but rather serve as a “landing platform” for complexes involved in various aspects of transcriptional repression (Beisel and Paro, 2011; Black et al., 2012). The methylated lysine 9 residue is directly bound by the chromodomain of Heterochromatin Protein 1 (HP1) (Bannister et al., 2001; Lachner et al., 2001), which is a suppressor of PEV in *Drosophila* (Eissenberg et al., 1990) and has three isoforms in mammals: HP1α/CBX5, HP1β/CBX1, and HP1γ/CBX3 (Fodor et al., 2010). HP1 has the ability to self-oligomerize and interact simultaneously with more than one

nucleosome, features that may help to directly drive chromatin compaction (Canzio et al., 2011). Moreover, HP1 proteins interact directly with a variety of repressive chromatin modifiers, such as DNA methyltransferases (DNMTs) and histone deacetylases (HDACs) (Eissenberg and Elgin, 2014). This is consistent with observations that H3K9 methylation are epistatic to DNA methylation at CpG nucleotides and histone deacetylation at pericentromeric repeats and embryonically silenced genes, sites where these three chromatin features coincide (Lehnertz et al., 2003; Epsztejn-Litman et al., 2008). HP1 further interacts with the co-repressor KAP1/TRIM28/TIF1β (Ryan et al., 1999); both proteins not only localize to H3K9me3-marked chromatin but also can in turn recruit H3K9 methyltransferases to act on nearby nucleosomes (Schultz et al., 2002; Eissenberg and Elgin, 2014), providing a mechanism for heterochromatin spread (Al-Sady et al., 2013). Importantly, forced tethering of the fission yeast methyltrasferase Clr4 to chromatin is sufficient to drive gene repression, and when an H3K9 demethylase is ablated, this repression is stable through multiple mitotic and meiotic generations even after removal of the tethered Clr4 (Audergon et al., 2015; Ragunathan et al., 2015). These findings establish the potential of H3K9me3 to function in epigenetic inheritance and its causal role in gene silencing. Similar results have been obtained in mammalian cell culture systems, where ectopic recruitment of HP1 or KAP1 can drive heritable H3K9me3 spreading and gene silencing over distances of multiple kilobases (Ayyanathan et al., 2003; Groner et al., 2010; Hathaway et al., 2012).

In addition to H3K9me3, trimethylation of a different histone 3 residue, lysine 27 (H3K27me3), also has important roles in transcriptional repression and chromatin compaction (Margueron and Reinberg, 2011). H3K27 methylation is exclusively catalyzed by a single protein complex, Polycomb Repressive Complex 2 (PRC2), which in mammals

can contain either of two methyltransferases EZH1 or EZH2 (Müller et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002; Ferrari et al., 2014). This complex was named 'Polycomb' because its genetic disruption in *Drosophila* led to improper body segmentation (Lewis, 1978), associated with aberrant de-repression of Hox transcription factor genes (Lewis, 1978; Duncan, 1982). Recent genome-wide mapping studies have revealed a global correlation between H3K27me3 and transcriptional repression (Barski et al., 2007), and genes marked by H3K27me3 are markedly enriched for transcription factors involved in cell fate transitions, a result corroborated in diverse organisms and developmental lineages (Lee et al., 2006; Boyer et al., 2006; Schwartz et al., 2006; Bracken et al., 2006; Xu et al., 2014a). H3K27me3 generated by PRC2 promotes binding by the chromodomain-containing Polycomb Repressive Complex 1 (PRC1) (Lee et al., 2007), which can directly compact nucleosome arrays *in vitro* (Francis et al., 2004). However, other studies have found that the chromatin structure of H3K27me3-marked promoters can remain accessible to binding by general transcription factors and RNA polymerase II (Breiling et al., 2001; Dellino et al., 2004). These and other studies have suggested that H3K27me3-associated repression is mediated at least partially through blocking the transcriptional initiation or elongation of an engaged RNA polymerase (Breiling et al., 2001; Dellino et al., 2004; Stock et al., 2007; Kanhere et al., 2010). Thus, it remains unclear to what extent the H3K27me3 mark is associated with the highly compact chromatin structure that is characteristic of heterochromatin.

Although the term "facultative heterochromatin" is more frequently associated in the literature with H3K27me3 (eg, (Trojer and Reinberg, 2007)), regions marked by H3K9me3 show substantial developmental rearrangement. Differentiated human cells have megabase-sized domains of H3K9me3 that are expanded in both size and number

in comparison to pluripotent embryonic stem (ES) cells (Hawkins et al., 2010), consistent with a large increase in total H3K9me3 levels by mass spectrometry (Sridharan et al., 2013). Differentiation-specific H3K9me3 domains encompass several genes required for pluripotency, as well as numerous gene family clusters (encoding zinc finger transcription factors, adhesion proteins, olfactory receptors, and mediaters of neurotransmission) with cell type-specific function (Hawkins et al., 2010; Soufi et al., 2012; Vogel et al., 2006). These large H3K9me3 domains are observed in diverse human cell lines and tissues, with substantial regions of overlap as well as divergence when comparing different tissues (Soufi et al., 2012; Becker et al., 2016). H3K9me3 is also strongly enriched with domains mapped by their association with the nuclear lamina (lamina-associated domains, or LADs), which also undergo developmental rearrangement (Guelen et al., 2008). Meanwhile, H3K27me3-enriched regions can span hundreds of kilobases and, similar to H3K9me3-marked regions, cover a larger fraction of the genome in differentiated cells compared to ES cells (Pauler et al., 2009; Hawkins et al., 2010). Comparison of H3K9me3 and H3K27me3 domains in somatic cells reveals some regions of overlap but largely distinct positioning in the genome (Hawkins et al., 2010; Chandra et al., 2012). This is similar to the observation that H3K9me3 and H3K27me3 both contribute to silencing on the inactive X chromosome but occupy distinct spatial domains (Chadwick and Willard, 2004; Heard, 2005; Nozawa et al., 2013). Comparable to findings for H3K9me3, earlier studies suggested that coverage for the dimethyl mark H3K9me2 increases after exit from pluripotency (Wen et al., 2009). Other groups have challenged this result, arguing that it is dependent on the threshold set for defining H3K9me2-enriched regions (Filion and van Steensel, 2010), though mass spectrometry data also reveals an increase in H3K9me2 levels in mouse fibroblasts compared to ES cells (Sridharan et al., 2013). Regardless,

enrichment of H3K9me2 can be found in common among H3K9me3- and H3K27me3-marked regions (Chandra et al., 2012), suggesting that the two trimethyl marks allow more specific delineation of two flavors of repressive chromatin.

1.2 Heterochromatin as a barrier to cell reprogramming

The epigenetic mechanisms that enforce cell type-specific transcription programs must be overcome when cells are "reprogrammed" or converted to other cell fates, such as pluripotency (Graf and Enver, 2009; Vierbuchen and Wernig, 2012). In 2006, it was shown that overexpression of four genes is sufficient to convert differentiated cells, such as skin fibroblasts, to induced pluripotent stem (iPS) cells, which closely resemble ES cells and have the ability to give rise to any cell type in the embryo (Takahashi and Yamanaka, 2006; Takahashi et al., 2007). The four genes overexpressed in the original iPS technique—*OCT4*, *SOX2*, *KLF4*, and *CMYC* (henceforth, "OSKM")— all encode master transcription factors expressed in pluripotent ES cells (Boyer et al., 2005; Takahashi and Yamanaka, 2006). Additional combinations of genes are capable of generating iPS cells (Buganim et al., 2012; Shu et al., 2013), with all such methods fundamentally relying on lineage-specifying transcription factors to rewire gene expression programs toward a desired fate (van Oevelen et al., 2013).

Among cellular conversion techniques, iPS cells represent a highly successful example of reprogramming, as the cells produced closely resemble non-reprogrammed ES cells on the functional, transcriptional, and epigenetic level (Maherali et al., 2007; Wernig et al., 2007; Cahan et al., 2014). The technique also has tremendous utility in biomedicine, since iPS cells derived from a patient can be differentiated to a cell type of

interest, enabling the production of patient-specific cells for therapeutic transplantation (Wu and Hochedlinger, 2011; Hanna et al., 2007; Wang et al., 2015) or for modeling of diseases affecting tissues that cannot be cultured *ex vivo* (Moretti et al., 2010; Toma et al., 2015). Nonetheless, transcription factor-driven iPS reprogramming remains a highly inefficient process, with only a small fraction of the starting cell population (generally <0.1%) reaching the iPS state, through a process that appears to be stochastic (Hanna et al., 2009), and the duration of the process lasts over one month for human cells (Vierbuchen and Wernig, 2012; Papp and Plath, 2013). Although expression of the iPS reprogramming factors for a few days is sufficient to cause widespread downregulation of the somatic transcriptional program in the majority of cells, activation of the pluripotency network is more delayed (Brambrink et al., 2008; Stadtfeld et al., 2008; Polo et al., 2012; Tanabe et al., 2013). Among genes repressed in the starting cell type, a particular subset of pluripotency markers—so-called "late-reprogramming genes"—are only expressed in the final stages of the reprogramming process and are highly predictive of successful iPS conversion (Buganim et al., 2012; Polo et al., 2012).

Our laboratory previously mapped the genome-wide binding of the OSKM factors after 48 hours of expression in human fibroblasts (Soufi et al., 2012). This work showed that the factors are able to engage their gene targets in a range of chromatin contexts. Yet there are megabase-sized stretches of chromatin that lack binding for all four factors, even though the factors were able to bind such sites in ES cells (Soufi et al., 2012). These regions, defined by their differentiation-specific impediment to OSKM binding, were termed "Differentially Bound Regions" (DBRs) (Soufi et al., 2012). The DBRs, which average ~2 MB in size, closely correspond to domains of H3K9me3 present in fibroblasts and absent in pluripotent ES cells. In addition to H3K9me3, the DBRs have several other

features of heterochromatin: resistance to DNase-I digestion, transcriptional repression, overlap with LADs, and an enrichment for repetitive elements (Soufi et al., 2012). The DBRs also contain genes essential for pluripotency, including *NANOG*, *SOX2*, *DPP2*, *DPPA4*, *GDF3*, and *ZFP42*. Strikingly all of these genes were previously shown to be delayed in activation until the late stages of reprogramming (Buganim et al., 2012; Polo et al., 2012), suggesting that a failure of the reprogramming factors to access the DNA within heterochromatic DBRs was a key underlying mechanism for the inefficiency of iPS conversion. Indeed, reducing H3K9me3 levels through knockdown of SUV39H1/H2 methyltransferase caused an increase in OCT4 and SOX2 binding within DBRs (Soufi et al., 2012), concomitant with an increase in the rate and number of human iPS colonies formed (Onder et al., 2012; Soufi et al., 2012). In the murine system, similar results have been shown for shRNA against SETDB1, which caused one of the largest increases in iPS colony formation in a screen of 615 chromatin-related genes (Cheloufi et al., 2015). Mouse iPS reprogramming is also enhanced by inhibition or knockdown of the H3K9me2-related enzymes G9a and GLP (Shi et al., 2008; Sridharan et al., 2013), as well as by knockdown of HP1 isoforms, with the strongest effect seen for HP1γ/CBX3 (Sridharan et al., 2013). Thus, perturbation of H3K9me3-related heterochromatin has emerged as a major strategy for enhancing conversion to pluripotency.

Heterochromatin structures over core pluripotency loci not only impede the rate of the reprogramming process, but also are a major cause for cells arresting in non-pluripotent states. In a plate of cells expressing the reprogramming factors, many go on to become “partially reprogrammed” cells, which express a subset of undifferentiated markers and can be stably propagated, but lack expression of many late reprogramming genes and have limited developmental potential (Mikkelsen et al., 2008; Fussner et al.,

2011). In partially reprogrammed cells, core pluripotency genes retain H3K9me3 (Chen et al., 2013), high levels of DNA methylation (Mikkelsen et al., 2008; Fussner et al., 2011), and a compact chromatin structure (Fussner et al., 2011). Erasure of heterochromatin marks by knockdown of SETDB1 or SUV39H1, overexpression of the H3K9 demethylase KDM4B, or inhibition of DNA methyltransferases are sufficient to drive partially reprogrammed cells to the iPS state (Chen et al., 2013; Mikkelsen et al., 2008). These findings reveal that failure to convert H3K9me3-marked heterochromatin to active euchromatin is a major restriction point on the path to the iPS state. Moreover, even in fully reprogrammed human iPS cells, there exist large domains with aberrant non-CpG methylation compared to compared to ES cells (Lister et al., 2011), and 21 out of 22 of these domains fall within heterochromatic DBRs (Soufi et al., 2012). Thus, heterochromatin contributes to the persistent epigenetic signatures that remain in iPS cells, highlighting the challenge of activating these regions of chromatin during transcription factor-mediated reprogramming.

Moreover, the barrier imposed by H3K9me3-marked heterochromatin on reprogramming is not specific to particular transcription factor combinations. Reprogramming to pluripotency can also be carried out by somatic cell nuclear transfer (SCNT), a long-established technique (Gurdon et al., 1958) involving transfer of a nucleus from a differentiated cell into an enucleated egg in order to restore the developmental potential of the donor nucleus. SCNT allows for the cloning of mice to produce viable offspring (Wakayama et al., 1998), but most of the resulting embryos arrest at an early developmental stage (Pasque et al., 2011a). By generating transcriptomic data for two-cell mouse embryos, Zhang and colleagues curated large genomic regions called “Reprogramming Resistant Regions” (RRRs), which become acquire transcriptional

activity in normal fertilized embryos but remain aberrantly silenced in SCNT-derived embryos (Matoba et al., 2014). The RRRs, which ranged from 100 to 800 kilobases in size, had high levels of H3K9me3 in the donor nuclei and were enriched for LINE and LTR repeat elements (Matoba et al., 2014), similar to the DBRs discovered for human iPS reprogramming (Soufi et al., 2012). Knockdown of SUV39H1/H2 or overexpression of the H3K9 demethylase KDM4D ameliorated the RRR expression defect and allowed as much as 80% of the embryos to reach the blastocyst stage, compared to less than 20% in controls (Matoba et al., 2014). This dramatic improvement in the viability and developmental competence of SCNT is consistent with previous results after deletion of G9a methyltransferase (Epsztejn-Litman et al., 2008). Moreover, these findings suggest that H3K9me2/3-marked regions pose an impediment not only to the activity of the OSKM factors but also to the coordinated action of the diverse components of the egg cytoplasm.

In addition to influencing the kinetics of the reprogramming process, dissolution of heterochromatin may be instrumental to the pluripotent state itself. Electron spectroscopic imaging (ESI) allows high-resolution analysis of chromatin structures, including detection of nucleosomes, calculation of inter-fiber distances, and distinguishing of chromatin from protein-based complexes (Ahmed et al., 2010). In pluripotent cells of the developing embryonic epiblast, ESI reveals a dispersed network of 10-nm fibers and an absence of blocks of highly compacted chromatin, which is seen at later stages of development in the primitive endoderm or trophectoderm cells (Ahmed et al., 2010). These densely packed heterochromatin fibers are also absent in cultured ES or iPS cells, while they remain in partially reprogrammed cells (Fussner et al., 2011). ES cells also lack dense heterochromatin structures detected by electron microscopy (Underwood et al., 2016) and, compared to differentiated cells, have fewer and less intense H3K9me3 foci

(Meshorer and Misteli, 2006). These observations, along with the finding that ES-cell chromatin has a higher rate of exchange of linker histone and HP1 (Meshorer and Misteli, 2006), has led to the conclusion that pluripotency represents a more globally accessible chromatin state (Gaspar-Maia et al., 2011). This model is supported by the elevated global transcriptional output of ES cells, leaky transcription of tissue-specific genes, and expression of satellite repeats and LINE and SINE retrotransposons (Efroni et al., 2008), which demonstrates that heterochromatin function—in addition to structure—is different between pluripotent and differentiated cells.

Given the reduction or absence of traditional hallmarks of heterochromatin in pluripotent cells, it is currently unclear whether the profound influence of H3K9me3 removal on reprogramming to pluripotency will apply generally to other cell-conversion techniques, or whether it is purely a byproduct of the unique features of the pluripotent state. A variety of methods exist for reprogramming somatic cells directly to another differentiated lineage, without going to through a pluripotent intermediate, by overexpressing tissue-specific transcription factors (Ladewig et al., 2013). Before the iPS era, Weintraub and colleagues showed that ectopic MyoD expression was sufficient to convert fibroblasts to muscle progenitors (Davis et al., 1987), and now diverse transcription factor combinations have been discovered for converting fibroblasts to cardiomyocytes (Ieda et al., 2010), neurons (Vierbuchen et al., 2010), hepatocytes (Sekiya and Suzuki, 2011; Huang et al., 2011), as well as for causing cells to switch lineages within the hematopoietic (Xie et al., 2004) or pancreatic (Zhou et al., 2008) compartments. Directly reprogrammed cells are typically shown to have undergone a cell fate change by their morphology, expression of characteristic marker genes, and rescue of tissue function *in vivo* in transplantation experiments (Zhou et al., 2008; Sekiya and Suzuki, 2011; Huang

et al., 2011). However, in nearly all cases, transcriptome-wide assessment reveals substantial differences between directly reprogrammed cells and the native cell types they mimic (Cahan et al., 2014; Morris et al., 2014).. Transcriptomes of engineered cell types reveal persistent expression of genes specific to the starting cell type, failure to completely induce the gene networks of the desired cell population, and, in some cases, aberrant induction of genes for additional developmental lineages that share some of the same transcription factors (Marro et al., 2011; Cahan et al., 2014; Morris et al., 2014). Thus, whereas iPS reprogramming is an inefficient but highly faithful conversion to the pluripotent state, directed conversion methods often achieve high yields of reprogrammed cells but are – as a class – less faithful to their intended lineage (Cahan et al., 2014).

There is great interest in developing techniques to generate large numbers of functional human hepatocytes in *vitro*, which could allow therapeutic transplantation of patient-derived cells for metabolic liver disease (Haridass et al., 2008; Asgari et al., 2013) and improved screening of pharmaceuticals for pharmacokinetics and toxicity (Sahi et al., 2010). Two recent reports describe the reprogramming of human fibroblasts to a hepatocyte-like state, which they termed human induced hepatocytes (hiHeps), using different combinations of endodermal transcription factors that both included HNF1A and HNF4A (Huang et al., 2014; Du et al., 2014). The resulting cells exhibit hepatocyte morphology, express marker genes like albumin and α-1-antitrypsin, and demonstrate aspects of normal liver metabolism like glycogen synthesis, LDL update, and detoxification of certain xenobiotic drugs (Huang et al., 2014; Du et al., 2014). However, analysis of hiHep transcriptomes reveal widespread differences compared to native hepatocytes in culture, including hepatocyte-specific transcription factor and metabolic genes that remain uninduced in the reprogrammed cells (Huang et al., 2014; Du et al., 2014). The hiHep

cells are capable of engrafting in the livers of mouse models of liver failure (Huang et al., 2014; Du et al., 2014), including fumarylacetoacetate hydrolase (FAH)-deficient animals (Azuma et al., 2007), which represents an improvement over previous work using cells differentiated from pluripotent stem cells (Ji et al., 2013; Du et al., 2014; Vallier, 2014). However, *in vivo* engraftment and expansion has only been demonstrated in hiHep cells after cell immortalization with to disable cell-cycle checkpoints (tumorigenic manipulations that preclude therapeutic use), and, even so, the engrafted hiHep cells secreted an order of magnitude less albumin (Du et al., 2014) or required an order of magnitude more transplanted cells to rescue mouse survival (Huang et al., 2014) compared to native fetal hepatocytes. A particular challenge in the differentiation or reprogramming of cells to the mature hepatic fate is the activation of a panel genes encoding enzymes of the cytochrome P450 family, which are critical for drug metabolism and excretion (Vallier, 2014). However, the role of H3K9me3 or heterochromatin in impeding activation of these gene classes has not been explored, and no manipulations of chromatin state have been shown to improve conversion to the hepatic lineage.

1.3 The regulation and composition of mammalian heterochromatin

Given the role of H3K9me3-dependent heterochromatin in impeding cell reprogramming, the mechanisms that control H3K9me3 deposition—and, in particular, the processes that establish H3K9me3 domains in a cell type-specific manner—might be targeted in order to influence reprogramming efficiency or allow access to specific cell fates. However, very little is known about the protein machinery that mediates the observed dynamics in H3K9me3 domains among developmental stages and different lineages (Allan et al., 2012;

Soufi et al., 2012; Becker et al., 2016). Since H3K9me3 methyltransferases are constitutively expressed and do not themselves recognize specific DNA sequences (Aagaard et al., 1999; Tachibana et al., 2001; Marmorstein, 2003), additional mechanisms, such as transcription factor- or RNA-mediated recruitment of methyltransferases, are required to explain the selectivity of H3K9me3 deposition.

Several transcription factors have been shown to play a role in heterochromatin establishment, though in many cases such mechanisms were described for constitutively heterochromatic repeats as opposed to gene-containing domains. Gfi1b, Sall1, Zeb1, and certain Pax family transcription factors all bind to pericentromeric satellites, which contain sequence motifs for each of these factors within the repeated unit (Vassen et al., 2006; Yamashita et al., 2007; Bulut-Karslioglu et al., 2012). Gfi1b is known to directly interact with SUV39H1 and G9a, suggesting a direct role in H3K9me3 deposition (Vassen et al., 2006). Meanwhile, Pax3 and Pax9 are required in mouse fibroblasts for the silencing of major satellite transcription and for the integrity of pericentromeric heterochromatin, but the mechanism by which their binding leads to repression rather than persistent activity is unclear and may involve a transient state of transcript generation (Bulut-Karslioglu et al., 2012). The roles of Gfi1b, Pax3, and Pax9 in constitutive heterochromatin is surprising given the cell type-specific expression of these factors (Vassen et al., 2006; Bulut-Karslioglu et al., 2012), suggesting that alternative mechanisms maintain pericentromeric heterochromatin depending on the lineage and the factors expressed.

An important regulator of H3K9me3 recruitment is Krüppel-associated box (KRAB)-associated protein 1 (KAP1), also known as TRIM28 or TIF1β (Iyengar and Farnham, 2011). KAP1 is the obligate co-repressor for KRAB domain-containing zinc finger transcription factors (KRAB-ZNFs), a family of hundreds of recently evolved

transcription factors that are specific to tetrapods (Friedman et al., 1996; Kapopoulou et al., 2016). KAP1 interacts directly with both HP1 and SETDB1, and thus it acts as a molecular link between the sequence-specific binding of KRAB-ZNFs and the subsequent deposition and propagation of H3K9me3 (Ryan et al., 1999; Schultz et al., 2002; Iyengar and Farnham, 2011). Phosphorylation of KAP1 at serine 824, which occurs in response to DNA damage, causes a global reduction in chromatin compaction (Ziv et al., 2006), and mutations mimicking constitutive phosphorylation of this residue promote iPS reprogramming (Seki et al., 2010) Ectopic tethering of KAP1 or a KRAB domain is sufficient to induce spreading of H3K9me3 over distances of tens of kilobases and mitotically heritable gene silencing (Ayyanathan et al., 2003; Groner et al., 2010; Gilbert et al., 2014). However, recent findings suggest that KRAB-ZNF mediated recruitment of KAP1 may not have a widespread role in heterochromatin domain formation. Deletion of KAP1's RBCC domain, which is required for KAP1-KRAB interaction, predominantly affects KAP1 binding at the 3' end of ZNF genes (Iyengar et al., 2011). These genes represent a small subset of KAP1's chromatin binding, do not significantly change expression upon loss of KAP1 (Iyengar et al., 2011), and may not be heterochromatic given evidence of transcriptional elongation (Blahnik et al., 2011). While KAP1 is required for the silencing of endogenous retroviruses (Rowe et al., 2010; Ecco et al., 2016), no study has reported the elimination of a large (>10 kb) H3K9me3 domain upon loss of a specific KRAB-ZNF or even KAP1 itself.

Indeed, a challenge for transcription factor-dependent models of H3K9me3 domain formation is the sheer size of these domains. While lateral propagation of H3K9me3 from a single nucleation site, such as a transcription factor binding motif, is compatible with experimental evidence (Groner et al., 2010; Hathaway et al., 2012) and

mathematical modeling (Hodges and Crabtree, 2012) for distances of tens of kilobases, such mechanisms to not account for the creation of H3K9me3 domains of multiple megabases, as is observed in differentiated human cells (Hawkins et al., 2010; Chandra et al., 2012; Soufi et al., 2012). The formation of such domains may involve the coordinated action of many reiterated transcription factor binding events, or may alternatively involve mechanisms unrelated to local sequence motifs.

In addition to transcription factor-dependent mechanisms, it has been shown that nucleation and maintenance of heterochromatin domains in some cases depends upon noncoding RNA (ncRNA). Such RNA-based mechanisms have been best elucidated in the fission yeast *S. pombe*, where maintenance of pericentromeric heterochromatin requires the RNA interference (RNAi) pathway (Volpe et al., 2002), which is activated by low-level transcription of the pericentromeric repeats themselves (Djupedal et al., 2005; Kato et al., 2005). Repeat-derived ncRNAs are processed into small interfering RNAs (siRNAs) and stay tethered to the site of nascent heterochromatin transcription by RNA-RNA base pairing; this creates a binding platform for the recruitment of additional machinery involved in transcriptional silencing, including an H3K9me3 methyltransferase (Bühler et al., 2006; Zhang et al., 2008; Bayne et al., 2010). More recently, additional ncRNA-based mechanisms that are distinct from RNAi, such as those involving RNA processing factors and the RNA exosome, have been implicated in heterochromatin establishment in fission yeast (Bühler et al., 2007; Reyes-Turcu et al., 2011).

Whether RNAi contributes to heterochromatin in mammals remains controversial, in part owing to inconsistent evidence for heterochromatin-derived siRNAs (Saksouk et al., 2015). However, accumulating data suggest that other forms of ncRNA may have a substantial role in regulating mammalian heterochromatin. Several groups have observed

that, at some heterochromatin regions, the presence of H3K9me3 is compatible with simultaneous low-level transcription (Martens et al., 2005; Terranova et al., 2005; Lu and Gilbert, 2007). HP1α binds directly to RNA, including transcripts derived from satellite repeats, and this interactions is required for its localization to and function at pericentromeric heterochromatin (Maison et al., 2002; Muchardt et al., 2002; Maison et al., 2011). In normal mouse embryogenesis, pericentromeric heterochromatin formation requires an early burst of transcription from these regions (Probst et al., 2010; Casanova et al., 2013), and ectopic addition of satellite-derived dsRNAs is sufficient to rescue defective heterochromatin formation in an H3.3 mutant (Santenard et al., 2010). Away from the centromere, LINE-1 retrotransposons undergo a similar early developmental burst in transcription that precedes their silencing (Fadloun et al., 2013), and regions of persistent LINE-1 transcription on the X chromosome are found in close proximity to genes silenced by X inactivation, suggesting a possible direct role in silencing (Chow et al., 2010). Finally, a ribosomal-derived ncRNA has been shown not only to participate in silencing of ribosomal DNA, but also to initiate widespread heterochromatin formation and loss of pluripotency when ectopically expressed in human ES cells (Savić et al., 2014). The mechanisms and protein machinery that link the transcription of such ncRNAs to H3K9me3 deposition and chromatin compaction are largely undefined. Moreover, it remains to be determined what are the relative contributions of RNA-based versus transcription factor-based mechanisms in the formation of the megabase-scale domains of facultative heterochromatin that emerge during mammalian development (Hawkins et al., 2010).

To identify new proteins involved in the regulation of H3K9me3, some groups have performed chromatin immunoprecipitation (ChIP) to isolate H3K9me3-marked chromatin

fragments, followed by mass spectrometry (MS) to define the protein content of those fragments (Soldi and Bonaldi, 2013; Engelen et al., 2015; Ji et al., 2015). An alternative approach is the use of MS-based proteomics after pulldown of soluble complexes with an H3K9me3 peptide bait, which has been used to identify “reader” complexes interacting directly with the mark (Vermeulen et al., 2010; Eberl et al., 2013). Each of these studies profiled interacting proteins for multiple histone modifications, and the H3K9me3-associated proteins were not further studied to investigate a functional role in H3K9me3 or gene repression, nor were their genomic locations mapped (Vermeulen et al., 2010; Eberl et al., 2013; Soldi and Bonaldi, 2013; Engelen et al., 2015; Ji et al., 2015). Nonetheless, large numbers of H3K9me3-associated proteins were identified with no previously characterized role in heterochromatin. The identified H3K9me3 readers included several transcription factors with zinc finger domains (non-KRAB), including POGZ, CHAMP1/ZNF828, ZNF282D, ANDP, and ZMYM4, which were also shown to interact directly or indirectly with HP1 isoforms (Eberl et al., 2013; Vermeulen et al., 2010). Other groups have shown, by ChIP-MS, that H3K9me3-marked chromatin fragments are enriched for components of the nuclear scaffold/matrix, including Scaffold Attachment Factor (SAF)-A/HNRNPU, SAFB1, SAFB2, and SAFB-like transcriptional modulator, as well as Matrin3 (MATR3) (Soldi and Bonaldi, 2013), consistent with a tethering of heterochromatin to nuclear structures. Surprisingly, simple inspection of these H3K9me3-related proteomic datasets reveals large number of RNA-binding proteins. A published list of H3K9me3 readers (Vermeulen et al., 2010) includes several heterogeneous ribonucleoproteins (HNRNPs) (HNRNPA1, HNRNPA2, HNRNPA2B1, HNRNPK, and HNRNPL) as well as the RNA-binding proteins NONO and SFPQ, which are involved in splicing but have also been shown to act as corepressors for hormone receptors (Mathur

et al., 2001; Dong et al., 2009). A H3K9me3-directed ChIP-MS study in the same cell type identified all of these same proteins as well as scores of additional RNA-processing factors (Transformer-2 protein homologs A and B; serine/arginine-rich splicing factors (SRSFs) 1, 2, 3, 6, and 7; RNA binding motif protein, X-linked (RBMX); etc) (Soldi and Bonaldi, 2013). The specific RNA-binding protein HNRNPK has recently been found to interact with SETDB1 and KAP1, promote H3K9me3 deposition at gene promoters and endogenous retroviruses, and impede iPS reprogramming (Bao et al., 2015; Thompson et al., 2015). However, for the larger set of H3K9me3-associated RNA-binding proteins, their role in heterochromatic silencing is unexplored, as is whether RNA interaction is required for the interaction with H3K9me3-marked chromatin. These studies suggest that the composition and regulation of mammalian heterochromatin is likely to involve a wider diversity of factors than the canonical chromodomain proteins and repressive histone modifiers.

1.4 H3K9me3 outside of heterochromatin and avenues for investigation

Although H3K9me3 is enriched at both repetitive and gene-containing regions of repressed heterochromatin, care must be taken when inferring the relationship between H3K9me3-associated proteins and heterochromatin per se. First, two of the most comprehensive ChIP-MS datasets (Engelen et al., 2015; Ji et al., 2015) were prepared using murine ES cells, which appear to completely lack the highly compacted chromatin structures present in differentiated cells (Meshorer and Misteli, 2006; Fussner et al., 2011; Underwood et al., 2016). These findings illustrate that H3K9me3 deposition can in the absence of the high levels of physical compaction that are typically thought to define heterochromatic regions. Second, even in differentiated cells, the H3K9me3 mark is

present in actively transcribed regions of chromatin. Work in both *Drosophila* and mammalian cells has shown an increase in H3K9me3 and HP1 within in gene bodies upon transcriptional induction (Piacentini et al., 2003; Vakoc et al., 2005), and indeed HP1 appears to promote transcriptional elongation at least some genes (Piacentini et al., 2003, 2009; Riddle et al., 2012), with certain phosphorylated forms of mammalian HP1γ localizing preferentially to active chromatin (Lomberk et al., 2006).

In general, levels of H3K9me3 within transcribed gene bodies are quantitatively lower than regions of constitutive heterochromatin (Vakoc et al., 2005). However, specific regions have a seemingly paradoxical chromatin configuration with high levels of H3K9me3 coincident with gene transcription (Blahnik et al., 2011; Riddle et al., 2012). In human cells, this has been observed at clusters of KRAB-ZNF genes on chromosome 19, which feature large domains of H3K9me3 and HP1 binding (Vogel et al., 2006) overlapping with H3K36me3 (Blahnik et al., 2011), a mark that is associated with transcription elongation (Vakoc et al., 2006). Analysis of SNPs in RNA-seq data confirms that both alleles of the H3K9me3-marked ZNF genes are expressed (Blahnik et al., 2011). These clusters of genes are the same sites where KAP1 binding was observed to be most dependent upon interaction with KRAB domains (Iyengar et al., 2011), suggesting that KAP1 recruitment by KRAB-ZNF proteins to their own genes, previously proposed to be a negative feedback mechanism (O'Green et al., 2007), is compatible with gene expression (Blahnik et al., 2011; Valle-García et al., 2016) and therefore does not necessitate formation of repressive heterochromatin. The KRAB-ZNF clusters on chromosome 19 were also shown to have a distinct pattern of three-dimensional genomic contacts and, compared to other H3K9me3-marked regions, were classified as forming a distinct architectural subcompartment (Rao et al., 2014). Of note, this finding required

targeted analysis of chromosome 19 alone, and the approach lacked sensitivity to detect similar regions genome-wide (Rao et al., 2014). A similar chromatin state, with strong enrichment for H3K9me3 and HP1 coinciding with H3K36me3-marked, transcribed genes, has been observed on *Drosophila* chromosome 4 (Riddle et al., 2012). At present, no study has comprehensively defined regions of H3K9me3 that are permissive to transcription or have an accessible chromatin structure, and it is unknown whether such domains exist on human chromosomes other than chr19 or affect non-ZNF human genes.

The finding of the H3K9me3 mark in the context of transcriptionally active chromatin suggests a limitation of approaches dependent upon this modification, such has H3K9me3 ChIP-MS (Soldi and Bonaldi, 2013; Engelen et al., 2015; Ji et al., 2015), to define with confidence the protein composition of repressive heterochromatin. Further work is necessary to determine whether, for example, the large number of RNA-binding proteins found to bind or co-precipitate with H3K9me3 (Vermeulen et al., 2010; Soldi and Bonaldi, 2013) are “contaminants” from euchromatic forms of H3K9me3 domains, or alternatively whether such proteins indeed contribute to highly compacted heterochromatin in mammalian cells, similar to RNA-based mechanisms known to regulate yeast or fly heterochromatin (Grewal and Elgin, 2007; Bühler et al., 2007; Reyes-Turcu et al., 2011). A second limitation of H3K9me3-dependent approaches to define the heterochromatin proteome is the exclusion of heterochromatic regions marked by H3K27me3, as exemplified by regions on the inactive X chromosome (Plath et al., 2003; Rego et al., 2008), or possibly regions of heterochromatin that might exist despite the absence of either mark (as suggested by computational clustering of *Drosophila* and human chromatin states (Filion et al., 2010; Zhu et al., 2013)). Although some evidence suggests that H3K27me3-marked chromatin is more accessible to transcription factor

binding and hence may be less heterochromatic than H3K9me3-marked chromatin (Breiling et al., 2001; Dellino et al., 2004), this model would be best tested by a method that defined heterochromatic regions genome-wide based on chromatin structure, independent form histone mark, which could then be compared to H3K9me3 and H3K27me3 ChIP-seq profiles. Physical isolation of structurally compact regions as intact chromatin fragments would allow simultaneous analysis of their genomic localization and proteomic composition. Finally, to the extent that regions with heterochromatic structure diverge from H3K9me3 or H3K27me3 domains, it is important to assess the combinations of chromatin structures and histone marks that impart resistance to gene activation during cellular reprogramming, to gain insight into the barriers to cell fate conversion.

CHAPTER 2. H3K9ME3-DEPENDENT HETEROCHROMATIN: BARRIER TO CELL FATE CHANGES (REVIEW ARTICLE)

The following chapter contains a review article that was published in *TRENDS in Genetics* in January 2016 (Vol. 32(1), pages 29-41). The content of the paper focuses on the connection between H3K9me3-marked heterochromatin and cell fate control, with particular emphasis on H3K9me3 as a barrier to cell reprogramming. The paper was co-written by myself and Dario Nicetto, Ph.D. (co-first authors), with guidance from Dr. Zaret (senior author).

2.1 Modes of developmental gene silencing

The diverse repertoire of cell types in multicellular organisms is achieved by the differential regulation of gene expression. While much effort has been expended to study how genes are activated, less is known about mechanisms by which cell type-inappropriate genes are repressed, even though this is a crucial aspect of cell fate control (Fisher and Merkenschlager, 2002; Hemberger et al., 2009; Meister et al., 2011). It has been long been appreciated that genetic material in the nucleus can be partitioned into two general categories: open 'euchromatin,' which has a relatively low density of DNA and high rates of gene transcription, and 'heterochromatin,' which has a relatively high density of DNA and low rates of gene transcription. Heterochromatin was originally discerned cytologically by the intensity of dark staining with DNA dyes (Heitz, 1928). The physically condensed state of these regions, reflected by their increased resistance to nucleases (Wallrath and Elgin, 1995) and their compact properties in biophysical assays (Frenster et al., 1963; Gilbert and Allan, 2001), is mechanistically linked to gene silencing, since compaction renders the DNA template less accessible to binding by the transcriptional machinery. Heterochromatin also has the property of spreading along chromosomes, which is illustrated by the compaction and silencing of transgenes integrated proximal to endogenous heterochromatin regions (Wallrath and Elgin, 1995; Elgin and Reuter, 2013).

A large fraction of mammalian genomes is taken up by repeat-rich sequences—including tandem-repeat satellites near centromeres and telomeres, retrotransposons, and endogenous retroviruses—which pose a risk to genome integrity through their potential for illicit recombination and self-duplication. Thus, in all cell types, there is utility in keeping such regions physically inaccessible and, consequently, transcriptionally silent,

by packaging them in condensed heterochromatin. Such repeat-rich regions are classified as 'constitutive' heterochromatin, as their silencing is universal across developmental lineages (Saksouk et al., 2015). By contrast, 'facultative' heterochromatin refers to regions whose compaction and silencing is dynamic in development, such as at cell type-specific genes and enhancers (Trojer and Reinberg, 2007).

In organisms ranging from the fission yeast *Schizosaccharomyces pombe* to humans, repeat-rich constitutive heterochromatin is marked by dimethylation and trimethylation of histone 3 lysine 9 (H3K9me2 and H3K9me3) (Rea et al., 2000; Nakayama et al., 2001; Martens et al., 2005). In mammals, these covalent modifications are catalyzed by five members of the SET-domain containing family of methyltransferases. SETDB1 and the related enzymes SUV39H1 and SUV39H2 contribute to both H3K9me2 and H3K9me3 (Rea et al., 2000; Schultz et al., 2002), while GLP and G9a (also called EHMT1 and EHMT2, respectively) catalyze H3K9me1 and H3K9me2 (Tachibana et al., 2001, 2002, 2005). H3K9me2/me3 are bound by the chromodomain of Heterochromatin Protein 1 (HP1, three isoforms in mammals), which can self-oligomerize and recruit repressive histone modifiers, contributing to heterochromatin compaction and spread (Bannister et al., 2001; Lachner et al., 2001; Canzio et al., 2011). The methyltransferases that deposit H3K9me2 and H3K9me3 are required to establish high levels of DNA methylation at CpG dinucleotides and low levels of histone acetylation, two other hallmarks of heterochromatin (Epsztejn-Litman et al., 2008; Lehnertz et al., 2003). By contrast, cell type-specific repression of many genes requires trimethylation of a different H3 residue, lysine 27 (H3K27me3), which is catalyzed by the Polycomb repressive complex 2 (PRC2) (Boyer et al., 2006; Ezhkova et al., 2009; Lee et al., 2006; Xu et al., 2014a). This mode of "facultative" silencing is particularly prominent at many lineage-specifying transcription

factor genes, such as the homeobox (HOX) family (for detailed review of the role of PRC2 and H3K27me3 in development, see: (Margueron and Reinberg, 2011; Beisel and Paro, 2011)).

The presence of H3K27me3 over gene promoters is highly correlated with gene repression (Hawkins et al., 2010; Margueron and Reinberg, 2011), yet it has been shown that H3K27me3-marked promoters remain accessible to binding by general transcription factors and a paused RNA polymerase (Breiling et al., 2001; Dellino et al., 2004). This contrasts with chromatin marked by H3K9me3, which occludes the DNA from binding by transcription factors with diverse DNA-binding domains (Soufi et al., 2012). Thus, H3K9me3-dependent and H3K27me3-dependent silencing appear to be mechanistically different based on the extent to which the chromatin is accessible to other factors (see Table 2-1).

Although the H3K9me3 modification has been most often studied in the context of constitutive heterochromatin, genome-wide mapping studies have made clear its role in cell type-specific regulation of facultative heterochromatin (Hawkins et al., 2010; Vogel et al., 2006; O'Geen et al., 2007; Zhu et al., 2013). In differentiated human cells, H3K9me3 forms large contiguous domains ranging in size from the kilobase to the megabase scale (Hawkins et al., 2010; Soufi et al., 2012; Vogel et al., 2006) (Figure 2-1). These domains or 'patches' expand in both number and size during differentiation from pluripotency, and they span numerous genes repressed in a cell type-specific manner (Hawkins et al., 2010). In particular, there is enrichment for H3K9me3 over gene family clusters, such as those for zinc finger transcription factors, olfactory receptors, and neurotransmitter-related genes (in non-neuronal cell types) (Hawkins et al., 2010; Vogel et al., 2006; O'Geen et al., 2007), raising the possibility that H3K9me3 protects repetitive gene clusters from illicit

recombination similar to noncoding repeats, while also suppressing transcription. Such H3K9me3 domains are largely exclusive of the H3K27me3 domains that also expand during development (Hawkins et al., 2010; Pauler et al., 2009; Chandra et al., 2012), highlighting the different functions of these marks, although some developmental transcription factor genes are decorated by both modifications (Hawkins et al., 2010).

The related repressive modification, H3K9me2, similarly forms megabase-scale domains that include genes; the domains have been called Large Organized Chromatin K9 modifications (LOCKs) (Wen et al., 2009). Interestingly, binding sites for the insulator protein CTCF were detected at the boundaries of these large domains, suggesting that presence of such H3K9me2-decorated patches might be intimately connected to higher-order chromatin structures maintained by CTCF (Wen et al., 2009). Whether the boundaries of these H3K9me2 domains expand during differentiation from pluripotency has been a matter of dispute (Wen et al., 2009; Filion and van Steensel, 2010; Lienert et al., 2011), with some groups favoring a model of mostly invariant domains during development but local gain of H3K9me2 over select genes (Lienert et al., 2011). Nevertheless, the dimethyl mark is important for the silencing of lineage-inappropriate genes during differentiation (Wen et al., 2009; Lienert et al., 2011; Liu et al., 2015b), and mass spectrometry-based quantification of histone marks reveals an increase in both H3K9me3 and H3K9me2 in mouse fibroblasts compared with pluripotent cells (Sridharan et al., 2013). Taken together, the findings in this section indicate that H3K9me2/3 deposition is patterned according to cell identity and must be reset to specify new fates.

2.2 Heterochromatin: a barrier to cell reprogramming and cell fate plasticity

The hallmarks of cell identity are erased when differentiated cells are reprogrammed to induced pluripotent stem (iPS) cells (see Box 1 - Methods of Cellular Reprogramming) (Takahashi and Yamanaka, 2006). This conversion process requires that reprogramming transcription factors bind to their targets in DNA and reactivate pluripotency genes that were silenced in development, suggesting that accessing heterochromatic regions is a necessary step to fully reprogram cells. However, only a minor fraction of the starting cells (<0.1%) successfully complete this process (Papp and Plath, 2013; Vierbuchen and Wernig, 2012), raising the question of what chromatin features contribute to such inefficiency.

H3K9me3 heterochromatin impedes iPS reprogramming

Insights into chromatin impediments to reprogramming emerged from determining where the canonical iPS reprogramming factors (Oct4, Sox2, Klf4, and cMyc; henceforth OSKM) first bind the genome shortly after they are expressed in human fibroblasts (Soufi et al., 2012). All four factors target open chromatin sites, but only OSK, while not M, also target sites containing nucleosomes and lacking evident histone marks, making them pioneer factors (Iwafuchi-Doi and Zaret, 2014; Soufi et al., 2012, 2015). However, there are megabase-scale chromatin regions in which none of the four factors can target DNA in fibroblasts, even though these same domains have binding sites for the factors in pluripotent cells (Soufi et al., 2012). Thus, the domains were called Differentially Bound Regions (DBRs). The DBRs overlap with domains enriched for H3K9me3 in fibroblasts but not in embryonic stem (ES) cells (see Figure 2-1), suggesting that H3K9me3

heterochromatin may mediate the impediment to OSKM binding. Indeed, knockdown of the SUV39H1/H2 methyltransferases increases Oct4 and Sox2 binding in these regions (Soufi et al., 2012).

The DBRs encode diverse genes and repeat elements, including transcription factor genes essential for pluripotency, such as NANOG, SOX2, DPPA2, DPPA4, GDF3, and ZFP42 (Boyer et al., 2005; Soufi et al., 2012). Strikingly, all of these genes were independently shown to be delayed in activation until the late phases of reprogramming (Buganim et al., 2012; Polo et al., 2012), with endogenous SOX2 and NANOG highly restricted to cells that successfully reprogram (Buganim et al., 2012). The discovery that all four OSKM factors fail to bind within large patches of H3K9me3 heterochromatin (Soufi et al., 2012) that include key pluripotency genes provides mechanistic insight into the observation that these genes are more refractory to activation than others (Polo et al., 2012; Buganim et al., 2012). The DBRs also encompass 21 out of 22 of the domains found to have aberrant non-CpG methylation in human iPS cells, compared with ES cells (Soufi et al., 2012; Lister et al., 2011). This indicates that some H3K9me3 domains, in addition to impeding the rate or efficiency of reprogramming, have a persistent effect in the final reprogrammed state of iPS cells, rendering the conversion to the ES state incomplete.

These findings suggested that H3K9me3 removal might be an effective strategy to enhance the efficiency of reprogramming. Indeed, knockdown of the SUV39H1/H2 methyltransferases, thereby reducing H3K9me3, causes a dramatic increase in the number and rate of appearance of human iPS colonies (Soufi et al., 2012). Independently, in a screen of 22 chromatin modifiers, short hairpin RNA (shRNA) against SUV39H1 was found to cause the strongest increase in human iPS colony formation (Onder et al., 2012). Similar results have been obtained for the other H3K9 methyltransferases, in that

reprogramming efficiency is improved in murine neural progenitor cells after G9a inhibition (Shi et al., 2008) and in murine fibroblasts after depletion of G9a, GLP, or SETDB1 (with additive effects in combination) (Sridharan et al., 2013). It is thus unclear which methyltransferase is most responsible for stabilizing the differentiated state. The yield of fully reprogrammed murine iPS colonies is also enhanced by perturbation of other heterochromatin components, such as knockdown of individual HP1 isoforms (e.g., HP1γ/*Cbx3*) (Sridharan et al., 2013), inhibition of histone deacetylases (Huangfu et al., 2008; Liang et al., 2010; Mali et al., 2010), or inhibition of DNA methylation (Mikkelsen et al., 2008). Loss of DNA methylation enhances removal of H3K9me3 in the presence of a transcriptional stimulus (Hathaway et al., 2012), and thus the effects of DNA methyltransferase inhibition on reprogramming efficiency may act through similar mechanisms as SUV39H1 knockdown, although this has not been definitively investigated.

Other components of repressive chromatin that oppose iPS reprogramming appear to act at sites outside of DBRs. Demethylation of H3K27me3 by Utx is required for reprogramming (Mansour et al., 2012), while the repressive histone variant macroH2A inhibits it (Gaspar-Maia et al., 2013; Pasque et al., 2012), but both observations are linked to a common class of pluripotency genes that activate in early reprogramming (Gaspar-Maia et al., 2013; Mansour et al., 2012), in contrast to most DBR genes (Buganim et al., 2012; Soufi et al., 2012). In further contrast to H3K9me3, the H3K27me3 methyltransferase EZH2 is required for iPS reprogramming, consistent with its role in maintaining pluripotency (Buganim et al., 2012; Onder et al., 2012; Pereira et al., 2010). Thus, iPS reprogramming depends upon continued deposition of H3K27me3 at certain loci, simultaneous with H3K27me3 removal by Utx at other loci. Finally, reduction of

another mediator of gene silencing, MBD3 (a component of the NuRD histone remodeling and deacetylase complex), can allow a high fraction of cells to reprogram to the iPS state and to do so in a more synchronous manner (Rais et al., 2013; Luo et al., 2013). However, this co-repressor thwarts reprogramming factor activity at sites they already bind (Rais et al., 2013), and its role in regulating H3K9me3 domains or preventing factor binding to heterochromatic genes has not been explored.

Paucity of heterochromatin defines the pluripotent state

A reduction in inaccessible H3K9me3-marked heterochromatin not only speeds conversion to pluripotency by enhancing transcription factor binding, but it also appears to be a fundamental hallmark of the pluripotent state. Using electron spectroscopy imaging (ESI), Bazett-Jones and colleagues identified remarkable differences in chromatin compaction between embryonic epiblast cells and subsequent lineage-restricted stages of development (primitive endoderm and trophectoderm), with the former characterized by a highly dispersed network of 10-nm fibers and the latter showing blocks of highly compacted chromatin (Ahmed et al., 2010). Studies of mouse ES cells in culture revealed similar findings by ESI (Hiratani et al., 2010) and a reduction in the number and intensity of nuclear foci that stain positively for H3K9me3 (Meshorer and Misteli, 2006). Furthermore, the chromatin of pluripotent cells shows a higher rate of exchange of chromosomal proteins such as linker histone and HP1, indicative of a more dynamic and accessible chromatin state (Meshorer and Misteli, 2006). Consistent with such accessibility, ES cells have elevated levels of global transcriptional activity, including expression of repetitive sequences and mobile elements, which are repressed in differentiated cells (Efroni et al., 2008). Importantly, depletion of proteins involved in

maintaining chromatin accessibility (Efroni et al., 2008) or introduction of elements that promote heterochromatin formation (Savić et al., 2014) results in impaired ES cell self-renewal and altered differentiation capacity. Thus, the developmental plasticity of early embryonic cells, much like the ability of differentiated cells to complete reprogramming, is tightly linked to the accessibility of chromatin.

The necessity of heterochromatin erasure for the pluripotent state is further illustrated by studies of 'partially reprogrammed' cells that appear during iPS conversion. These cells lack induction of the pluripotency gene network and have limited developmental potential, but express the reprogramming factors and have downregulated their somatic program (Mikkelsen et al., 2008; Sridharan et al., 2009). Nuclear imaging with ESI revealed that partially reprogrammed cells, but not iPS cells, have highly compartmentalized heterochromatin structures containing dense chromatin fibers, similar to differentiated cells (Fussner et al., 2011). This is consistent with the persistence of DNA methylation and H3K9me3 over specific pluripotency loci, including *Oct4* and *Nanog*, in these cells (Mikkelsen et al., 2008; Fussner et al., 2011; Chen et al., 2013). Meanwhile, erasure of H3K9me3, via depletion of Setdb1 or Suv39h1 or overexpression of the Kdm4b demethylase, is sufficient to allow partially reprogrammed cells to progress to full iPS cells (Chen et al., 2013). These findings suggest that H3K9me3 is not only a barrier to pluripotency factor binding in the earliest stages of reprogramming (Soufi et al., 2012), but also opposes late maturation steps necessary for pluripotency.

Heterochromatin opposes reprogramming by somatic cell nuclear transfer

In contrast to the reliance of iPS reprogramming on defined factors, somatic cell nuclear transfer (SCNT, see Box 1) utilizes the diverse factors of the egg cytoplasm, acting *en*

masse, to restore pluripotency, and the resulting process proceeds more rapidly (Jullien et al., 2011). Yet, recent evidence suggests that H3K9me3 heterochromatin presents a barrier to even this form of reprogramming. Zhang and colleagues performed detailed transcriptomic analysis of two-cell mouse embryos derived by normal fertilization and SCNT, and they identified 'reprogramming resistant regions' (RRRs) containing transcripts that were silenced only in the SCNT condition (Matoba et al., 2014). The authors found that the RRRs had features of heterochromatin including selective marking by H3K9me3 and enrichment for LINE and LTR-type repeat elements in the genome. Injection of mRNA for the H3K9 demethylase *Kdm4d* into the embryo or knockdown of *Suv39h1/h2* in donor nuclei improved the expression of genes within the RRRs. Importantly, either approach for reducing H3K9me3 caused dramatic improvements in the developmental potential of the SCNT-derived embryos, with as much as 80% of the embryos reaching the blastocyst stage, compared with less than 20% in controls (Matoba et al., 2014). Similarly, dramatic increases in the number of cleavage stage SCNT-derived embryos were observed for donor nuclei lacking G9a (Epsztejn-Litman et al., 2008). Other methods to reduce heterochromatin integrity – including inhibition of histone deacetylases (Bui et al., 2011; Iager et al., 2008), reduction in DNA methylation (Blelloch et al., 2006), or depletion of macroH2A (Pasque et al., 2011b) – all improved embryo derivation by SCNT.

Taken together, the current evidence suggests that heterochromatin, and in particular H3K9me3-marked domains, presents a barrier to reprogramming to pluripotency. The H3K9me3 heterochromatic barrier applies regardless of the cell conversion methodology (SCNT versus defined factors) and impairs both the efficiency of reprogramming and the quality of the cells produced (Figure 2-2).

2.3 H3K9me3 as a regulator of cell fate in vivo

The crucial function of H3K9me3 in impeding cell reprogramming (Soufi et al., 2012; Chen et al., 2013; Matoba et al., 2014) and in silencing lineage-specific genes (Hawkins et al., 2010; Zhu et al., 2013) suggests that heterochromatin helps maintain cellular identity. Thus, patterns of H3K9me3 must be reorganized during cell fate transitions in development, both in the early embryo (see (Fadloun et al., 2013a; Burton and Torres-Padilla, 2014) for review) and in terminal lineage maturation.

Maintaining and exiting pluripotency: H3K9me3 and transcription factor crosstalk

In pluripotent stem cells, transcription factor networks ensure that H3K9me2/me3 is deposited over genes for cell differentiation and removed from essential pluripotency regulators. In murine ES cells, Setdb1 has been shown to occupy and repress genes encoding developmental regulators (Bilodeau et al., 2009) and to act as a co-repressor of Oct4, thereby suppressing trophoblast genes (Lohmann et al., 2010; Yeap et al., 2009; Yuan et al., 2009). Similarly, Loh et al. elegantly demonstrated that in murine ES cells Oct3/4 positively regulates the expression of the demethylases Kdm3a and Kdm4c to remove H3K9me2 and H3K9me3, respectively, from *Tcl1* and *Nanog*, guaranteeing the maintenance of cell renewal in ES cells (Loh et al., 2007).

Upon implantation of embryos *in vivo* or differentiation of ES cells *in vitro*, there is a progressive and irreversible silencing of *Oct3/4* and other pluripotency-associated genes, including *Nanog*, *Stella*, and *Rex-1*. Deposition of H3K9me2 at these sites, and in turn DNA methylation, is dependent on the methyltransferases GLP and G9a (Epsztejn-Litman et al., 2008; Feldman et al., 2006; Liu et al., 2015b). G9a prevents *Oct3/4*

reactivation when differentiated ES cells are returned to pluripotency culture conditions (Feldman et al., 2006) (Figure 2-2, dashed line). Meanwhile, mutations in GLP that disrupt its H3K9me1-recognition domain result in decreased H3K9me2, a delay in silencing of pluripotent genes during ES differentiation, and abnormal embryonic development *in vivo* (Liu et al., 2015b). The reverse H3K9me2 dynamics are seen at the master germ-line regulator genes *Ddx4* and *Dazl*, which show high levels of H3K9me2 in ES cells and lose the modification in *in vitro*-generated mature primordial germ cell-like cells (Kurimoto et al., 2015). A reduction in H3K9me2 occurs at lamina-associated domains (LADs), which normally associate with the nuclear periphery, and is coupled to a relative depletion in H3K27me3 (Kurimoto et al., 2015), a mark enriched at the borders of LADs (Harr et al., 2015). The overall picture highlights crosstalk between H3K9me2 and H3K27me3 and a direct role for H3K9me2/me3 in the developmental control of gene expression.

Requirement of H3K9me2/me3 deposition for normal embryonic development

The importance of H3K9me2/me3 establishment in completing developmental transitions is illustrated by genetic loss-of-function studies in mouse embryos. G9a- and GLP-null embryos show early lethality, characterized by dramatic morphological abnormalities associated with alteration in gene expression and chromatin organization (Tachibana et al., 2002, 2005). Homozygous inactivation of SETDB1 also leads to embryonic lethality around the time of implantation, an even earlier stage compared with G9a and GLP mutants, as well as defects in inner cell mass growth (Dodge et al., 2004). Although single knockouts of either SUV39H1 or SUV39H2 in mice show no developmental defects, double-null mice are born at sub-Mendelian ratios and show prenatal lethality linked to genome instability (Peters et al., 2001). Furthermore, knockout of HP1β results in

dramatic genomic instability and leads to perinatal lethality, likely caused by defects in the development of neuromuscular junctions and cerebral cortex (Aucott et al., 2008).

The distinct lethal phenotypes seen for the different classes of H3K9me-related methyltransferases and associated factors reflect their diverse contributions during development. G9a, GLP, and SETDB1 regulate early lineage commitment (Bilodeau et al., 2009; Lohmann et al., 2010; Yeap et al., 2009; Yuan et al., 2009; Feldman et al., 2006; Epsztejn-Litman et al., 2008; Tachibana et al., 2005), while SUV39H1/H2 are involved in genome stability (Peters et al., 2001) and maintenance of fully differentiated cell identity (Soufi et al., 2012; Onder et al., 2012; Matoba et al., 2014).

H3K9me3 contributes to lineage restriction in mature cell types

The role of H3K9me3-decorated heterochromatin in controlling terminal differentiation and ensuring the stability of cell identity emerges in two recent studies. Amigorena and co-workers, studying the molecular mechanisms underlying naive T cell differentiation into distinct T helper (Th) cells subtypes, revealed an interplay between SUV39H1 and HP1α to maintain a high ratio of H3K9me3 over H3K9ac at key Th1 genes, the latter of which must be silenced in Th2 cells (Allan et al., 2012). Applying both genetic and pharmacological loss-of-function approaches, the authors showed that in SUV39H1- and HP1α-deficient conditions, Th2 cell lineage stability is compromised and cellular plasticity towards the Th1 fate is increased. This phenotype is also seen in disease-related conditions: a Th2-mediated allergic lung inflammation is reduced upon depletion of H3K9me3 (Allan et al., 2012). In a genome-wide approach, Casaccia and collaborators analyzed differentiation processes in the brain and showed that silencing of H3K9-related, but not H3K27-related, methyltransferases impairs oligodendrocyte differentiation, altering

their response to electric stimulation (Liu et al., 2015a). Taken together, these studies indicate that H3K9me3 and H3K27me3 have different roles in developmental gene silencing and cell identity maintenance, depending on the cell lineage.

2.4 Molecular control of H3K9me3 deposition

Since H3K9 methyltransferases are broadly expressed (Aagaard et al., 1999; Tachibana et al., 2001) and are not known to make specific base contacts with DNA (Marmorstein, 2003), additional factors are required to explain the site-selectivity of H3K9me3 deposition and the developmental dynamics of H3K9me3 domains. In this section, we consider protein and RNA-based mechanisms by which H3K9me3-based heterochromatic domains are established.

Transcription factor-based recruitment of heterochromatin

A growing number of sequence-specific transcription factors have been found to recruit the heterochromatin machinery to particular gene promoters. The retinoblastoma (Rb) protein interacts with both SUV39H1 and HP1, and it is required for cell cycle-regulated H3K9me3 at the cyclin E promoter (Nielsen et al., 2001). Also important for heterochromatin establishment is a large, tetrapod-specific family of zinc finger (ZNF) transcription factors containing Krüppel-associated (KRAB) domains. Krüppel-associated box zinc finger proteins (KRAB-ZNFs), which mostly have lineage-specific expression, repress transcription of target genes by binding the co-repressor KAP1 (also known as TRIM28 and TIF1β), which in turn interacts with HP1, SETDB1, and histone deacetylases (Friedman et al., 1996; Ryan et al., 1999; Schultz et al., 2002). Experimental tethering of

a KRAB domain to a genomic site results in spreading of H3K9me3 and silencing of gene promoters as far as 15 kb away (Groner et al., 2010). Yet mutant forms of KAP1 that cannot bind KRAB-ZNFs nonetheless retain many of their genomic binding sites (Iyengar et al., 2011), suggesting that there is still much to learn about KAP1 recruitment.

Murine satellite repeats contain reiterated binding sequences for transcription factors, such as Gfi1b, Sall1, Zeb1, and select Pax family members (Vassen et al., 2006; Yamashita et al., 2007; Bulut-Karslioglu et al., 2012). Specifically, Pax3 and Pax9 contribute to H3K9me3 deposition and transcriptional repression at major satellites and are required for the integrity of pericentric heterochromatin (Bulut-Karslioglu et al., 2012), a startling finding given that these factors are expressed only in select cell types. The alternative factors that sustain constitutive heterochromatin in Pax3/9-negative cell types, and the role of these factors in recruiting H3K9me3 to domains containing genes, are not presently understood.

Contribution of RNA to heterochromatin formation

In addition to the role of transcription factors, noncoding RNA (ncRNA) can function as a binding platform to establish heterochromatin at specific genomic positions. In the fission yeast *S. pombe*, heterochromatin formation at pericentromeres (Volpe et al., 2002) and other sites (Hall et al., 2002) depends upon the components of the RNA interference (RNAi) pathway and, paradoxically, requires transcription of the locus to be silenced (Djupedal et al., 2005; Kato et al., 2005). Double-stranded RNAs (dsRNAs) transcribed from these regions are processed into small interfering RNAs (siRNAs) by Dicer, which in turn guide the silencing machinery, including an H3K9 methyltransferase, to the site of nascent heterochromatin transcription by RNA-RNA base pairing (Bühler et al., 2006;

Zhang et al., 2008; Bayne et al., 2010). (For review of the role of RNAi in heterochromatin, see (Grewal and Elgin, 2007; Bühler and Moazed, 2007)). In *S. pombe*, there are additional mechanisms by which ncRNA can establish heterochromatin, independent of RNAi, involving a growing number of RNA processing factors and components of the RNA exosome (Bühler et al., 2007; Reyes-Turcu et al., 2011). Meanwhile, in mammals, despite initial reports that Dicer was required for silencing of pericentric heterochromatin (Fukagawa et al., 2004; Kanellopoulou et al., 2005), heterochromatin-derived dsRNA has not been consistently detected across cell systems (Saksouk et al., 2015).

While the mechanisms by which ncRNA may establish heterochromatin in mammals remain poorly understood, emerging evidence suggests its contribution is significant. Transcription of mammalian heterochromatin has been observed despite the presence of H3K9me3 over the same regions (Martens et al., 2005; Terranova et al., 2005; Lu and Gilbert, 2007). The localization of HP1α at pericentromeric heterochromatin is dependent on its interaction with RNA, specifically its binding via its hinge domains to sense-oriented repeat transcripts (Maison et al., 2002; Muchardt et al., 2002; Maison et al., 2011). Strikingly, injection of pericentromere-derived dsRNA in the early mouse embryo is sufficient to rescue the phenotype of a mutant with defects in constitutive heterochromatin (Santenard et al., 2010). During normal development, an early burst of major satellite transcription precedes and is required for H3K9me3 deposition (Probst et al., 2010; Casanova et al., 2013). Also, dynamics in major satellite transcription, which in turn modulate pericentromeric binding of HP1, are important for the cell fate transition of epithelial-to-mesenchymal transition (Millanes-Romero et al., 2013). Along the chromosome arms, LINE-1 repeats undergo a wave of transcription early in development (Fadloun et al., 2013b), and transcripts derived from these elements may contribute to the

silencing of nearby genes on the inactive X chromosome (Chow et al., 2010). Finally, in human ES cells, addition of another ncRNA derived from rDNA arrays is sufficient to induce widespread H3K9me3 deposition outside the nucleolus and to promote exit from pluripotency (Savić et al., 2014). These findings suggest an intimate relationship between RNA and H3K9me3 establishment, although the nature of the interactions and the RNA-binding proteins involved are in need of further elucidation.

2.5 Concluding Remarks

Recent work suggests that large domains of H3K9me2/3 form in a cell type-specific manner (Hawkins et al., 2010; Soufi et al., 2012), but the protein machinery responsible for such precise developmental dynamics remain largely mysterious (see Box 2 - Outstanding Questions). First, the mechanisms and relative contributions of RNA-dependent versus transcription factor-dependent H3K9me3 recruitment must be defined for these regions, and it is not presently understood how either process can nucleate H3K9me3 deposition over a domain as large as multiple megabases (Hawkins et al., 2010; Zhu et al., 2013). In addition, simple inspection of the large H3K9me3 patches on a genomic level shows that they can terminate precisely over a local domain, suggesting a type of boundary. Elucidating these mechanisms for the initiation and delimitation of H3K9me2/3 domains will enable more targeted strategies to perturb H3K9me3-dependent heterochromatin at specific sites, possibly to enhance reprogramming in a manner tailored to the starting and desired cell types.

As RNAi-based knockdown of all five H3K9 methyltransferases has been found to promote reprogramming to pluripotency (Chen et al., 2013; Onder et al., 2012; Soufi et

al., 2012; Sridharan et al., 2013), it will be important to carefully dissect the unique and redundant roles of each enzyme in the establishment of specific H3K9me2/3 domains in diverse cellular contexts, and the relative contributions of the dimethyl and trimethyl forms. This would be facilitated by the creation of conditional knockouts of these genes, alone and in combination. Mapping of H3K9me2/3 domains in specific lineages and developmental stages, coupled with conditional deletion of methyltransferases, will reveal the enzymes responsible for tissue-specific domains and their contribution to developmental gene regulation.

Finally, studies of lineage-specific H3K9me2/3 domains should investigate if they similarly impede direct conversion or transdifferentiation between two differentiated fates (see Box 1). Whether perturbation of heterochromatin components can universally improve the fidelity of these direct conversions, or whether the role of H3K9 methylation in reprogramming is pluripotency- or tissue-specific, will be an exciting avenue for further investigation.

2.6 Supporting text boxes

Box 1. Methods of Cellular Reprogramming

Reprogramming refers to the erasure of the identity of a cell to convert it to a different type of cell, most commonly the conversion of a specialized fate to an earlier, undifferentiated state. Multiple techniques (see (Yamanaka and Blau, 2010)) now exist to transform differentiated cells into cells that are pluripotent, which means that they can give rise to any cell type in the embryo.

Somatic cell nuclear transfer (SCNT)

Seminal work in the 1950s established that transfer of a nucleus from a differentiated cell into an enucleated egg induces a restoration of developmental potential and the production of viable embryos (Gurdon et al., 1958). SCNT has been used to successfully clone mammals, such as sheep (Wilmut et al., 1997) and mice (Wakayama et al., 1998). Nonetheless, the frequency at which SCNT gives rise to viable organisms is low, with most resulting embryos exhibiting phenotypic and gene expression abnormalities (Matoba et al., 2014; Pasque et al., 2011a). Elegant studies have revealed specific molecular events required to complete reprogramming after SCNT (Jullien et al., 2011, 2014), but given the complexity of the egg cytoplasm that is mediating the process, the underlying mechanism is likely to involve myriad factors acting in concert.

Generating induced pluripotent stem (iPS) cells

Takahashi and Yamanaka made a critical breakthrough by defining a specific set of four transcription factors that, when ectopically overexpressed, are sufficient to convert a

differentiated cell into an iPS cell (Takahashi and Yamanaka, 2006). The factors originally identified – Oct4, Sox2, Klf4, and cMyc – are central regulators of the pluripotency gene network in ES cells (Boyer et al., 2005), and additional combinations of factors capable of generating iPS cells have since been reported (Buganim et al., 2012; Shu et al., 2013). In all cases, the resulting iPS cells meet stringent criteria for pluripotency, such as ability to rescue tetraploid blastocysts and contribute to the germline, and on the transcriptional level most iPS lines are highly similar to ES cells derived from the pluripotent inner cell mass (Buganim et al., 2012; Shu et al., 2013; Wernig et al., 2007). However, iPS reprogramming is a highly inefficient process, as it proceeds to completion only in a small fraction of cells (generally <0.1%) and at long latency (weeks to months) (Vierbuchen and Wernig, 2012; Papp and Plath, 2013).

Direct cell fate conversion

The strategy of ectopically expressing defined cocktails of lineage-specific transcription factors has been used to convert or transdifferentiate differentiated cells to other developmental lineages, without going through a pluripotent intermediate (Davis et al., 1987; Graf and Enver, 2009; Ladewig et al., 2013). Despite the promise of these techniques, the reprogrammed cells generally exhibit substantial gene expression differences from their native counterparts, limiting their in vivo functionality and therapeutic utility (Cahan et al., 2014; Morris et al., 2014).

Box 2. Outstanding questions

What are the proteins and ncRNAs that control the cell type-specific locations and boundaries of large H3K9me3 domains across the mammalian genome?

To what extent do Suv39h and Setdb1 have specific, non-redundant roles in repressing cell identity genes, and how does this relate to the distinct embryonic phenotypes upon deletion of these H3K9me3-related methyltransferases?

How do H3K9me2- and H3K9me3- arked chromatin domains differ in their exclusion of transcription factors and resistance to gene activation?

What are the relative contributions of H3K27me3 and H3K9me3 to developmental gene regulation in different lineages, and to what extent do they cooperate in cell fate establishment?

Does H3K9me3-dependent heterochromatin impede direct conversion between differentiated cell types, similar to its role in limiting reprogramming to pluripotency?

2.7 Tables and Figures

domain properties	H3K9me3 domains	H3K27me3 domains
Genomic distribution	Constitutive heterochromatin and tissue-specific sites[1-3]	Tissue-specific sites[1,4-8]
Chromatin accessibility	Prevent binding by diverse TFs[2]	Allow binding by general TFs and paused RNA polymerase[9-10]
Presence at 'poised' genes, competent for activation	Downstream of some H3K4me3-marked promoters[11]	Overlapping with H3K4me3 at many promoters[12-13]
Timing of gene reactivation during iPS reprogramming	Latest stages of reprogramming[2,14-15]	Early-to-mid stages of reprogramming[16]
Major methyltransferases and role in reprogramming	SETDB1, SUV39H1/H2: impede iPS conversion[2,15,17]	PRC2 complex (EZH2 or EZH1): required for generating iPS[14,17-18]

Table 2-1. Differences between H3K9me3 and H3K27me3 Heterochromatin Domains

TFs = transcription factors. References: [1](Hawkins et al., 2010), [2](Soufi et al., 2012), [3](Liu et al., 2015a), [4](Boyer et al., 2006), [5](Lee et al., 2006), [6](Ezhkova et al., 2009), [7](Zhu et al., 2013), [8](Xu et al., 2014a), [9](Breiling et al., 2001), [10](Dellino et al., 2004), [11](Matsumura et al., 2015), [12](Bernstein et al., 2006), [13](Voigt et al., 2012), [14](Buganim et al., 2012), [15](Chen et al., 2013), [16](Mansour et al., 2012), [17](Onder et al., 2012), [18](Pereira et al., 2010).

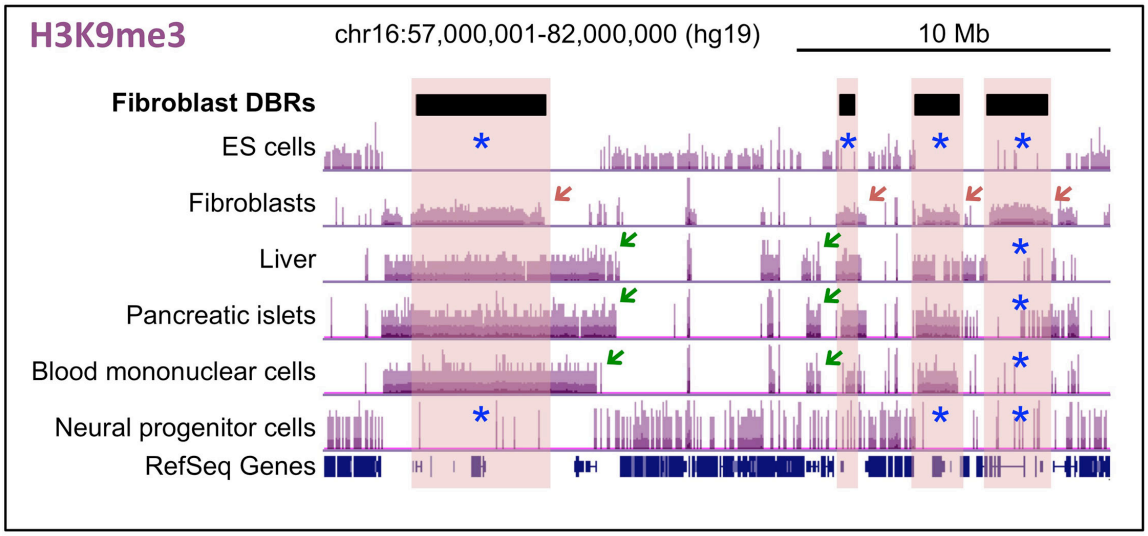


Figure 2-1. Megabase-scale domains of H3K9me3 vary by cell type and match regions resistant to reprogramming factor binding.

Shown is a 25-Mb segment of human chromosome 16, visualized in the UCSC Genome Browser. The purple tracks show H3K9me3 signals by chromatin immunoprecipitation and sequencing (ChIP-seq), normalized by input-subtraction, for the selected cell/tissue types. All ChIP-seq data come from the Roadmap Epigenomics Mapping Consortium (GSE16368). Note the close correspondence between the H3K9me3-enriched domains in foreskin fibroblasts (red arrows) and the fibroblast Differentially Bound Regions (DBRs, black bars), which are regions that fail to be targeted by pluripotency reprogramming factors in fibroblasts but are bound in ES cells (Soufi et al., 2012). Each of these regions lack H3K9me3 enrichment in ES cells, as well as in select other tissues (blue asterisks). Green arrows indicate representative H3K9me3 domains in other tissues that are absent in fibroblasts.

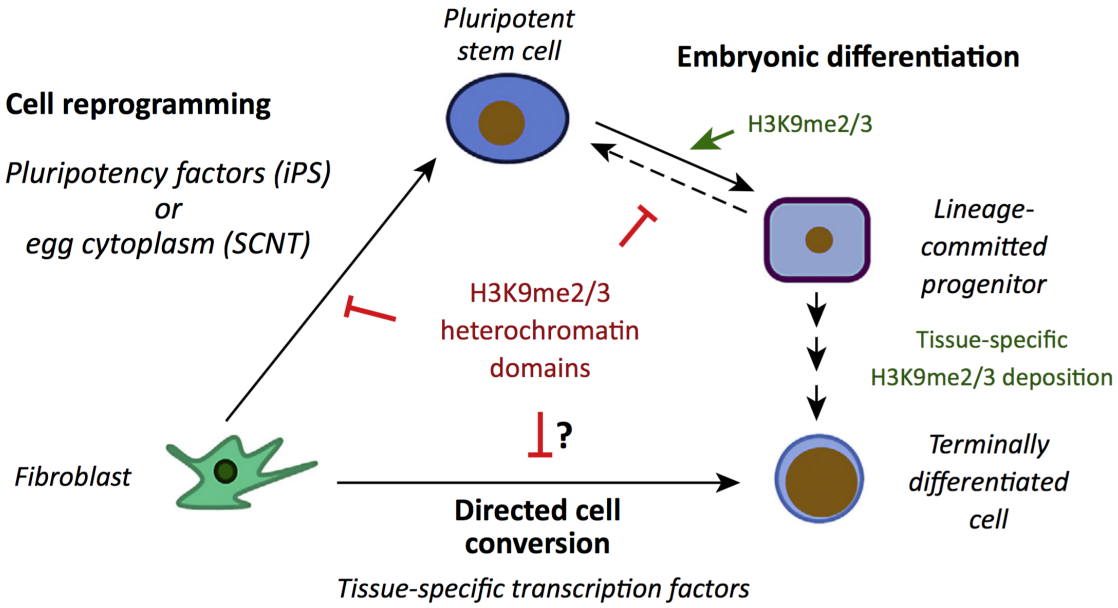


Figure 2-2. H3K9me2/3 heterochromatin domains impede diverse forms of cellular reprogramming.

The diagram shows major cell fate transitions (black arrows) that occur during differentiation and reprogramming and the role of H3K9me2/3 in these transitions. The leftmost black arrow indicates conversion of differentiated cells to pluripotency, which can be carried out by nuclear transfer to an enucleated egg or by overexpression of pluripotency transcription factors. In both cases, pluripotency genes inside H3K9me3 domains are more resistant to activation, and the success rate of reprogramming is improved when H3K9me3 levels are reduced (Chen et al., 2013; Matoba et al., 2014; Onder et al., 2012; Soufi et al., 2012). Thus, H3K9me3 domains impede reprogramming to pluripotency (red inhibitory arrows). When embryonic stem (ES) cell-derived differentiated cells are returned into ES culture conditions, thereby encouraging de-differentiation (dashed black arrow), the loss of an H3K9me2 methyltransferase increases

the appearance of undifferentiated colonies and the expression of pluripotency genes (Feldman et al., 2006). In contrast to reprogramming, the differentiation of pluripotent cells in culture (upper black arrow) is promoted by increases in H3K9me2/3 (Loh et al., 2007; Savić et al., 2014). Although H3K9me2/3 domains form in a tissue-specific manner over the course of development (rightmost black arrows), the role of these domains in the directed conversion of cells across developmental lineages (bottom black arrow) remains to be investigated. Abbreviations: iPS, induced pluripotent stem cell reprogramming; SCNT, somatic cell nuclear transfer.

CHAPTER 3. METHODS

3.1 Experimental methods

Cell Culture

Human BJ foreskin fibroblasts were obtained from Stemgent (08-0027) at passage 6 and cultured in Eagle's Minimum Essential Medium (EMEM) (Sigma-Aldrich M2279) supplemented with 10% fetal bovine serum (FBS, Hyclone SH30071) and 2mM L-glutamine (Gibco) at 37°C and 5% CO_2.

Preparation of crosslinked chromatin lysates

BJ fibroblast cells were grown to ~80% confluence in 150-mm dishes (5 - 20 plates per chromatin batch, ~4 x 10^6 cells per plate). Cells were crosslinked directly in the culture dishes, at room temperature, by addition of 2 ml formaldehyde solution (50 mM HEPES-KOH, pH 7.5, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 11% formaldehyde) to 20 ml media, for 1% final formaldehyde concentration. After 10 min, crosslinking was quenched by addition of glycine to a final concentration of 125 mM, followed by incubation for 5 min at room temperature. Cells were harvested from the plate with a plastic cell lifter, pelleted at 200 g for 4 min (4°C), and washed three times with ice-cold PBS. All subsequent steps were performed on ice or in centrifuges cooled to 4°C. To enrich for nuclei, cells were allowed to swell for 10 minutes in 4 ml Hypotonic Lysis Buffer (20 mM HEPES-KOH pH 7.5, 20 mM KCl, 1 mM EDTA, 10% glycerol, 1% IGEPAL CA-630, 0.25% Triton-X, 1 mM DTT, 0.2 mM PMSF, Protease Inhibitor Cocktail - Roche #11873580001) and were mechanically dounced for 50 strokes. Nuclei were pelleted at 1,350 g for 4 min, washed

with hypotonic buffer, and pelleted again at the same speed. Pellets were resuspended in 10 ml Nuclear Wash Buffer (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1 mM DTT, 0.2 mM PMSF, Protease Inhibitor Cocktail) and incubated for 10 min, rocking. Nuclei were pelleted at 1,350 g for 4 min and resuspended in 0.5 - 1 ml Sonication Lysis Buffer (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.5% N-lauroylsarcosine, 0.1% sodium deoxycholate, 1 mM DTT, 0.2 mM PMSF, Protease Inhibitor Cocktail) and transferred to 15-ml polystyrene tubes. Sonication was performed in polystyrene tubes using a Diagenode Bioruptor UCD-200 (power HI, cycles of 30s on / 30s off) for at least 25 cycles. After sonication, lysates were transferred to microcentrifuge tubes, supplemented with Triton-X to 1% final concentration (to promote chromatin solubility), and centrifuged at 20,000 g to pellet debris. Supernatants were transferred to new tubes, snap-frozen in liquid nitrogen, and stored at -80°C while chromatin shearing was assessed by agarose gel electrophoresis of purified DNA (as described for chromatin immunoprecipitation, below). Sonication was repeated as necessary until the majority of DNA was 200-400bp, with only a faint trail of larger material.

Chromatin immunoprecipitation

Immunoprecipitation was performed using Dynabead Protein G magnetic beads (Thermo Fisher, 10004D) saturated with the antibody of interest, such as anti-H3K9me3 (Abcam ab8898) or anti-mouse IgG control (Abcam ab46540). 5 µg of antibody and 25 ul of Dynabead slurry were used per 25 µg of chromatin (according to mass of purified DNA measured by nanodrop), with scaling as necessary. Dynabeads were first washed twice with 800 µl PBS, using a magnetic rack, and then resuspended in a volume of PBS that is 4X the original slurry volume. This suspension was supplemented with the desired

amount of antibody, mixed, and incubated on a rotating rack at 4°C for 2 - 6 hours, to allow antibody conjugation. Sonicated, crosslinked chromatin lysates (see above) were thawed on ice, and the desired mass of chromatin (25 µg or more) was aliquoted a low-retention 1.5-ml tube (Axygen MCT-150-L-C) tubes. Chromatin was diluted to 1 ml final volume with ice-cold Chromatin IP Buffer (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% N-lauroylsarcosine, 1% Triton 1 mM DTT, 0.2 mM PMSF, Protease Inhibitor Cocktail). A separate aliquot of chromatin (one-fifth the mass) was reserved as the input sample. Antibody-conjugated beads were washed twice with Chromatin IP buffer and then resuspended in the 1-ml sample of chromatin. Immunoprecipitations were incubated for 16 hours at 4°C on a rotating rack. Using a magnetic rack, the unbound lysate was aspirated, and the beads were washed 5 times with 900 µl ice-cold ChIP RIPA Buffer (50 mM HEPES, pH 7.5, 500 mM LiCl, 1 mM EDTA, 1% IGEPAL CA-630, 0.7% sodium deoxycholate, Protease Inhibitor Cocktail) and once with 900 µl ice-cold TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). Chromatin was eluted in 200 µl ChIP Elution Buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1% SDS) at 65°C for 30 min, shaking, and the eluate was transferred to a new tube. Reserved input sample was similarly diluted at least 3-fold in Chromatin IP Buffer, to 200 µl final volume. To purify DNA from ChIP eluate and Input, samples were decrosslinked by heating at 65°C for 18 hours. Chromatin was diluted with 200 µl TE and treated with 8 µl RNase A (10 mg/ml stock, Roche #10109169001), followed by incubation for 2 hours at 37°C. Protein was degraded by addition of 4 µl Proteinase K (20 mg/ml stock, Roche #03115828001) and incubation for 2 hours at 55°C. DNA was purified by two rounds of extraction with 400 µl phenol-chloroform-isoamyl alcohol. The extracted aqueous phase was supplemented with 16 µl of 5 M NaCl and 1.5 µl glycogen (20 mg/ml stock, Roche #10901393001). DNA was precipitated by addition of 2 volumes

(800 µl) 100% EtOH and overnight incubation at -20°C. DNA was pelleted at 20,000 g for 10 min (4°C), washed with 500 µl 80% EtOH, and pelleted again. The DNA pellet was air-dried and dissolved in 200 µl TE buffer. DNA yield was quantified by Quant-iT PicoGreen dsDNA Assay (Thermo Fisher P7589). Five-fold serial dilutions of the input DNA were prepared in DNA and were used as a standard curve when analyzing ChIP eluates by qPCR, in order to quantify sequence recovery as percent input.

Sucrose gradient sedimentation of chromatin

For preparative gradients used for proteomic and sequencing studies, crosslinked and sonicated chromatin was purified (see above) from near-confluent BJ fibroblasts in 20 150-mm plates (at least 8 x 10^7 cells, 0.5 mg DNA) per gradient. Chromatin lysates were prepared in 0.5 mL Sonication Lysis Buffer (see above) to allow the majority of the sample to be loaded on a single gradient. Prior to running gradients, chromatin shearing efficiency was verified by purifying DNA from a 5 µl chromatin aliquot (as described above for ChIP input DNA), while the remaining lysate was snap-frozen in liquid nitrogen and stored at -80°C. Empirically, we found that achieving sufficient levels of DNA shearing, comparable to standard ChIP, was important to achieve robust heterochromatin enrichment via sucrose gradient sedimentation.

6-40% linear sucrose gradients in Chromatin IP buffer were poured into 12-mL Ultra-Clear centrifuge tubes (Beckman #344059, 14 x 89 mm), using a Hoefer SG-15 gradient maker fitted with a two-way stopcock (Bio-Rad #7328102). Gradients were prepared by loading the chambers of the gradient maker with two solutions of approximately equal weight: 5.7 mL of 40% Sucrose Solution (40% sucrose, 10 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% Triton-X, 0.1% N-

lauroylsarcosine, 1 mM DTT, 0.2 mM PMSF, Protease Inhibitor Cocktail) and 6.4 mL of 6% Sucrose Solution (6 mL of 40% Sucrose Solution, plus 34 mL of: 10 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% Triton-X, 0.1% N-lauroylsarcosine, 1 mM DTT, 0.2 mM PMSF, Protease Inhibitor Cocktail), which are gradually mixed as the solutions exit the gradient maker. Gradients were filled bottom-up, beginning with the heavier 40% Sucrose Solution. To load the gradients, 455 µl of thawed chromatin lysate was mixed with 65 µl of the 40% Sucrose Solution, for a 5% final concentration of sucrose, and then the resulting sample was layered gently on the 6-40% gradients with a 1-mL pipet. Gradients were spun on a Beckman SW 41 Ti rotor at 41,000rpm for 3 hours (4°C), using conditions similar to those developed by Bickmore and colleages (Gilbert et al, 2004). Slow acceleration and deceleration settings were used. After sedimentation, gradients were fractionated top-down using a micropipet, generating 24 fractions of 500µL. 30µL of each fraction used for DNA purification (performed as for ChIP samples, see above) to allow fraction-specific qPCR studies; the remaining samples were snap-frozen and stored at -80°C.

The fractions found by qPCR to have the strongest enrichment for heterochromatic regions (fractions #10-17) were pooled as the “structural heterochromatin (strucHC) fraction”, while fraction #2 (second fraction from the top) was used as the “euchromatin fraction.” Both samples were dialyzed to Chromatin IP buffer (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% N-lauroylsarcosine, 1% Triton 1 mM DTT, 0.2 mM PMSF, Protease Inhibitor Cocktail) to remove sucrose. Dialysis was performed for two rounds of five hours each using Slide-A-Lyzer G2 Casettes 7K MWCO (Thermo Fisher). 5% of the dialyzed chromatin samples (approximately 250 ng DNA for strucHC

fraction) were used for DNA purification, using the same procedure as for ChIP eluates above, to enable subsequent qPCR and sequencing studies. For purification of proteins from gradient fractions, equal DNA equivalents were used for both the euchromatin and strucHC fractions, based on Quant-iT PicoGreen dsDNA Assay (Thermo Fisher P7589). Prior to protein precipitation, the dialyzed fractions were buffered by adding Tris-HCl pH 8.0 (to 100 mM final concentration) and then decrosslinked by heating at 65°C overnight followed by 99°C for 30 min. The proteins were precipitated in 6 volumes acetone (overnight at -20°C), pelleted at 3,600 g, washed once with ice-cold acetone, resuspended in 8M urea, and quantified by Bradford Assay (Bio-Rad #500-0006). Alternatively, the dialyzed strucHC and euchromatin fractions were used for chromatin IP against H3K9me3 or IgG control (using above ChIP protocol). For these IP eluates in ChIP Elution Buffer, one-twelfth of the sample was used for DNA purification, as above. For protein analysis, the remaining eluate was run in a cetrifugal evaporator to reduce the volume to ~25 µl, mixed with 4X NuPAGE LDS Sample Buffer (Thermo Fisher NP0007) and β-mercapto-ethanol (2.5% final concentration), and decrosslinked by heating at 99°C for 30 min.

Proteomics analysis of chromatin samples

Proteins were prepared for mass-spectrometry by in-gel digestion. Samples were mixed with 4X NuPAGE LDS Sample Buffer (Thermo Fisher NP0007) and β-mercaptoethanol (2.5% final concentration) and loaded into NuPAGE Novex 4-12% Bis-Tris Protein Gels (Thermo Fisher NP0335). Gels were run at 100V in NuPAGE MOPS SDS Running Buffer (Thermo Fisher NP0001) and washed three times with dH_2O for five minutes each. Gels were stained with SimplyBlue SafeStain (Thermo Fisher LC6060) for 1 hour at room

temperature, photographed, and destained in dH_2O overnight. Gels were then incubated in 40% ethanol, 10% acetic acid for 1-2 hours to fix and further destain. Lanes were excised and cut into five pieces that were digested in separate tubes (but pooled prior to nLC-MS/MS). Each piece was further diced into ~1 mm^3 cubes and transferred to clean microcentrifuge tubes.

Gel slices were reduced with 10 mM DTT for 1 hour at 56°C, alkylated using 55 mM iodoacetamide for 45 minutes (room temperature in dark) and digested overnight with trypsin at room temperature at a concentration of 12.5 ng/µl. Digestion was interrupted adding 1% formic acid. Samples were then desalted using in-house prepared C_{18} microcolumns, and the five peptide samples from the same gel lane were combined. Nano liquid chromatography was performed using a Thermo Scientific Easy nLC 1000 equipped with a 75 μm x 18 cm in-house packed column using Reprosil-Pur C18-AQ (3 μm, Dr. Maisch GmbH). Buffer A was 0.1% formic acid and Buffer B was 0.1% formic acid in acetonitrile. Peptides were resolved using a 165 min gradient from 2 to 28% B at a flow rate of 300 nL/min. The HPLC was coupled online to an Orbitrap Elite mass spectrometer (Thermo Scientific) for the first replicate and a Q-Exactive (Thermo Scientific) for the second and the third replicate operating in positive mode. Spectra were acquired using a data dependent acquisition (DDA) method, performing the full MS scan in the orbitrap at 60,000 (Elite) and 70,000 (Q-Exactive) resolution. The MS/MS was performed for both instruments at 17,500 resolution in the orbitrap mass analyzer. Loop count was set to 15 (Elite) and 12 (Q-Exactive) and collision energy was set to 20. Database searching was performed using MaxQuant v1.5.2.8 (Cox and Mann, 2008), using all standard settings unless otherwise stated. Database used was Human UniProt (v July 2015). For protein quantification the iBAQ option (Schwanhäusser et al., 2011) was enabled and adopted.

qPCR analysis of gradient fractions

Purified DNA samples from gradient fractions were quantified with Quant-iT PicoGreen dsDNA Assay (Thermo Fisher P7589) and diluted with TE buffer to 0.1 ng/µl, to allow loading of equal DNA mass per qPCR reaction. DNA concentrations were verified after dilution by repeat PicoGreen assay and were adjusted as needed. 10 µl qPCR reactions were prepared in 384-well optical plates with 2 µl DNA sample, 0.1 µl primer mix (10µM each primer), and 5µL Power SYBR Green PCR Master Mix (Thermo Fisher #4367659). Plates were in a 7900HT Real-Time PCR machine (Thermo Fisher #4329001), using the following thermal cycler protocol: 50°C for 4 min, 95°C for 10 min, followed by 40 cycles of 95°C for 15 s then 54°C for 15 s then 72°C for 45 s, with a dissociation curve generated to verify that a single PCR product was generated. qPCR results for each fraction were normalized to the input sample. Primer sequences for detecting sites inside and outside DBRs were from (Soufi et al., 2012) and are listed in the table below. All PCR amplicons correspond to Oct4/Sox2 binding sites: bound in fibroblasts and ES cells for non-DBR sites, and bound only in ES cells for DBR sites (Soufi et al., 2012).

Table 3-1. Primers used for qPCR analysis of Gradient Fractions and H3K9me3 ChIP.

primer name	forward primer (5' - 3')	reverse primer (5' - 3')	location
Ch3_DBR_3	TGGTCTTGAATTCCTGGCTG	GCTTAAGAATCGTCCGGAGG	DBR site
Ch3_DBR_4	ACCGCCATACCCAACTTG	GATGGCCCTAGGTCTTTAATGG	DBR site
Ch20_DBR_2	ATCAAGTGCCAGGAATGGAG	ATGGAGCCCGAATTTCTCAG	DBR site
Ch20_DBR_5	AATTTCAAGCGGAGCCCTAG	TCAGAAACCCTATTGAAGCCTC	DBR site
Ch22_DBR_2	GCCATTCGTGTGCAGAAAAG	CTGTCCATAGTCAGCGTTCC	DBR site
Ch22_DBR_5	CCTCAAGGGATTGGAAGATCTC	GGTGCCCAGATTAAATGTTCC	DBR site
DPPA4	TCCACCTCACCTCTTCTT	GTATTAGTAATTCAACCCAGACAA	DBR site

NANOG	TGTTGAACCATATTCCTGAT	TCTACCAGTCTCACCAAG	DBR site
Ch3_nonDBR_1	CATGGAGCAATTGTGAATAAATGTG	ATTAGGCTGGGGCTTTCTG	intergenic
Ch3_nonDBR_2	CCTCCAGTATCAACCGAAGAG	TCCGAAGACTCCTACTCACAC	promoter
Ch3_nonDBR_3	AAATGCTAAGAGGGTGTGGG	GAGAGTTGCCAGGAACAGAG	gene body
Ch20_nonDBR_1	CCCCGCAGACAATGACTATTAG	AGGTGTGAGCGTTCGATATG	intergenic
Ch20_nonDBR_3	GGACCACAGCACGGAAAC	CCTTCTCACTCCTCTTCTCCG	promoter
Ch22_nonDBR_2	GGGCTTGCATAGTGAAAACATG	ACGGTAGAGGACAGGGAAG	gene body
Ch22_nonDBR_3	CAGATTAATGTTTGCCAGGGC	AATATTTCCATTGCTCCAAAATTTCC	gene body
Ch22_nonDBR_4	CCCCTATCATTGTGAGAGTGTG	CAATTTACCCGCCACATCAC	promoter

Preparation of Gradient-seq and ChIP-seq libraries

Approximately 50 ng of purified DNA was used for each library, and two biological replicates (independent gradients) were sequenced per sample type. For samples containing large DNA fragments (including the structural heterochromatin fraction, IPs off of this fraction, and the input to the gradient), the pure DNA was first sheared in a Covaris S220 sonicator using snap-cap microTUBEs (Covaris #520045) for 5 min (settings: 175W peak power, 10% duty factor, 200 cycles/burst) to produce 150-350 bp fragments. Libraries were then prepared using the NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs E7370S), and amplified using 8 - 9 cycles of PCR using NEBNext Multiplex Oligos for Illumina (New England Biolabs E7335S and E7500S). For H3K9me3 ChIP-seq and ChIP Input libraries, the adapter-ligated DNA was size-selected using Agencourt Ampure XP beads (Beckman Coulter A63881), following the protocol in the NEBNext Ultra kit for a 200-bp average insert size. The libraries for Gradient-seq were not size-selected. Library yield and fragment size distribution was assessed on a Bioanalyzer 2100 instrument (Agilent Technologies), using the DNA 1000 kit (Agilent Technologies 5067-1504).

Western blotting

Whole-cell protein extracts were prepared by resuspending cells in RIPA extraction buffer (25 mM Tris, pH 7.5, 150 mM NaCl, 1% Na-deoxycholate, 1% IGEPAL CA-630, 0.1% SDS) supplemented with Protease Inhibitor Cocktail (Roche #11873580001). Suspensions were incubated on ice for 10 minutes and sonicated for 15 s on HI using a Diagenode Bioruptor UCD-200. Samples were centrifuged at 20,000 g for 10 min (4°C) to pellet debris, and the supernatant was transferred to new tubes. Protein content was quantified by BCA assay (Thermo Fisher Scientific #23227). Protein samples were mixed with 4X NuPAGE LDS Sample Buffer (Thermo Fisher Scientific NP0007) and 10X NuPAGE Sample Reducing Agent (Thermo Fisher Scientific NP0009), and were denatured at 70°C for 10 min. Samples were loaded in NuPAGE Novex 4-12% Bis-Tris Protein Gels (NP0335), and run using NuPAGE Running Buffers (NP0001; NP0002). Wet transfer to PVDF membranes (100V for 1.5 hr) was performed using NuPAGE Transfer Buffer (NP0006) containing 20% methanol, and membranes were blocked overnight in 5% nonfat dairy milk in TBS-T (20 mM Tris, pH 7.5, 150 mM NaCl, 0.1% Tween-20). Primary antibodies were diluted in 1% milk/TBS-T at the following concentrations: anti-GAPDH (1:1000, Santa Cruz Biotechnology sc-365062), anti-SUV39H1 (1:1000, Bethyl Laboratories A302-127A), anti-RBMX (1:1000, Cell Signaling Technology #14794), anti-H3K9me3 (1:1000, Abcam ab8898), and anti-Histone H3 (1:3000, Abcam ab1791). HRP-conjugated secondary antibodies (Santa Cruz Biotechnology sc-2004, sc-2005) were diluted 1:5,000 in 1% milk/TBS-T. Blots were developed using SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher Scientific #34080) and visualized with an Amersham Imager 600 (GE Healthcare Life Sciences).

Immunostaining

Cells were grown on plates coated with collagen I (Corning #354236), washed twice briefly with PBS, and fixed in 4% paraformaldehyde in PBS for 20 minutes at room temperature. Fixed cells were washed 3 times with PBS (all washes 5-10 minutes at room temperature, rocking), permeablized with ice-cold 0.1% Triton-X in PBS for 10 minutes, and washed twice TBS-T (20 mM Tris-HCl pH 7.4, 150mM NaCl, 0.05% Tween-20). Samples were blocked with 4% donkey serum (Sigma-Aldrich D9663) in PBS for 1-2 hours at room temperature or overnight at 4°C. Primary antibodies added in blocking solution and incubated overnight at 4°C, using the following concentrations: anti-human-albumin (1:100, Bethyl Laboratories A80-229A), anti-alpha-1-antitrypsin (1:200, Thermo Fisher Scientific RB-367), anti-FOXA3 (1:300, Santa Cruz Biotechnology sc-5361), anti-HNF1A (1:150, Santa Cruz Biotechnology sc-6547), and anti-HNF4A (1:150, Santa Cruz Biotechnology sc-8987). Cells were washed 3 times with TBS-T and then incubated with AlexaFluor 488- or 594-conjugated secondary antibodies raised in donkey (Thermo Fisher Scientific) at a 1:500 dilution in PBS, for 45 minutes at room temperature, protected from light. Samples were then washed 3 times with PBS, counterstained with 1μg/mL DAPI (Thermo Fisher, D1306) in PBS for 10 minutes, and washed with PBS once. Fluorescence images were taken using a Nikon eclipse TE2000-U microscope controlled by Nikon elements software and equipped with the appropriate filters.

Lentivirus production

Lentiviral plasmids pWPI.1-FOXA3, pWPI.1-HNF1A, and pWPI-HNF4A were kindly provided by the laboratory of Lijian Hui (Huang et al., 2014). Plasmid DNA was purified from bacteria using the EndoFree Plasmid Maxi Kit (Qiagen #12362). 293T cells were

grown in DMEM High Glucose (Thermo Fisher Scientific #11995) supplemented with 10% FBS (Hyclone SH30071) and seeded in 10-cm dishes at a density of 8 x 10^5 cells/plate. After 24 hours, transfection mixtures were prepared by mixing 2.5 µg lentiviral vector, 1.7 µg packaging plasmid psPAX2 (Addgene #12260), 0.8 µg envelope plasmid pMD2.G (Addgene #12259), and 30 µl Fugene 6 Transfection Reagent (Promega E2691) with 570 µl OptiMEM-I Reduced Serum Medium (Thermo Fisher Scientific #31985070), per plate. Transfection mixtures were vortexed, incubated for 15 min, and added drop-wise to plates containing 10 mL media. Media was changed 16 hours after transfection. 60 hours later, media containing viral particles was collected. Debris was pelleted at 800 g for 10 min (4°C), and the supernatant was passed through a 0.45 µm syringe filter (Millipore SLHV033RS). Viral particles were concentrated by ultracentrifugation at 25,000 rpm for 1.5 hr (4°C) with an SW-32 swinging bucket rotor (Beckman Coulter), and the viral pellet was resuspended in pure DMEM for 1/100th the original supernatant volume. Viral titer was determined immunostaining for FOXA3, HNF1A, and HNF4A in fibroblasts three days after infection with serial dilutions of concentrated virus. Dilutions of virus that produced 10-35% transgene-expressing cells were used to calculate the multiplicity of infection (MOI), and in turn the titer, using the relationship MOI = (-1) * ln(1 - [proportion infected]).

hiHep reprogramming

Transdifferentiation of fibroblasts to hiHep cells was carried out as previously described (Huang et al., 2014). BJ fibroblasts growing in complete EMEM (see above) were plated on collagen I-coated plates at a density of 3 x 10^4 cells per well in 12-well format. One day after plating, cells were infected with a cocktail of three lentiviruses (pWPI.1-FOXA3, pWPI.1-HNF1A, and pWPI.1-HNF4A), with an MOI of 1.25 per virus, in media containg

4.5 µg/ml polybrene. One day after infection, virus-containing media was removed, cells were washed twice with PBS, and fresh complete EMEM was added. On the second day after infection, medium was switched to Human Maintenance Medium (HMM): DMEM/F12 (Hyclone SH30023.01) supplemented with 0.544 mg/L $ZnCl_2$, 0.75 mg/L $ZnSO_4 \cdot 7H_2O$, 0.2 mg/L $CuSO_4 \cdot 5H_2O$, 0.025 mg/L $MnSO_4 \cdot 1H_2O$, 2 g/L Bovine serum albumin (Sigma-Aldrich), 2 g/L galactose (Sigma-Aldrich G0750), 0.1 g/L ornithine (Sigma-Aldrich O2375), 0.03 g/L proline (Sigma-Aldrich P5607), 0.61 g/L nicotinamide (N0636), 1X Insulin-transferrin-sodium selenite media supplement (Sigma-Aldrich I1884), 40 ng/ml TGF-α (Peprotech AF-100-16A), 40 ng/ml EGF (Peprotech AF-100-15), 10 µM dexamethasone (Sigma-Aldrich D4902), and 1% FBS (Hyclone SH30071). Media was changed with fresh HMM every 48 hours. Cells were analyzed for hepatic markers at 12-14 days after lentiviral infection.

siRNA transfection experiments

All knockdown experiments were performed with two cycles siRNA transfection, three days apart. The following Silencer Silect siRNAs from Thermo Fisher Scientific were used: Negative Control No. 1 siRNA (#4390843), SUV39H1 siRNA (s13660), RBMX siRNA "#1" (s56033), RBMX siRNA "#2" (s56035), and RBMX siRNA "#3" (s223747). Transfections were performed using 10 nM final concentration of siRNA and 2.5 µl/ml final concentration Lipofectamine RNAiMAX (Thermo Fisher Scientific #13778150). 10X transfection mixtures were prepared by adding 100 nM siRNA and 25 µl/ml Lipofectamine to OptiMEM-I Reduced Serum Medium (Thermo Fisher Scientific #31985070) and incubating for 10 min at room temperature. On day 0, cells were reverse transfected by seeding a 0.9X volume of cells/media into wells containing 0.1X volume of 10X transfection mixture. After

24 hours, siRNA-containing media was replaced with normal culture media. After 3 days, forward transfections were performed on adherent cells by adding 0.1X volume of 10X transfection mixture drop-wise to 0.9X volume of culture media. Media was again replaced 24 hours after transfection. RNA or protein was harvested on day 6, three days after the second siRNA transfection. For experiments performed with uninfected, proliferating fibroblasts, cells were passaged over the course of this six-day time-course, as needed to prevent over-growth.

Knockdown experiments in cells expressing hepatic transcription factors were performed as follows. First, on day -1, cells were infected as a batch (single well or plate) with pWPI.1-FOXA3, pWPI.1-HNF1A, and pWPI.1-HNF4A lentiviruses (MOI of 1.25 per virus), in media containg 4.5 μg/ml polybrene. 24 hours after infection, on day 0, virus-containing media was removed, cells were washed twice with PBS, and cells were detached from the plate and split for reverse siRNA transfection. The remainder of the six-day siRNA time-course was performed as described as above, except that, on day 4, media was changed to HMM (see above under "hiHep reprogramming") to promote hepatic induction. By infecting cells with factors prior to splitting for transfection with specific siRNAs, we ensure similar dosage of factors across the conditions being compared.

RNA isolation and Reverse Transcription-PCR

Total RNA was isolated using the ZR-96 Quick-RNA kit (Zymo Research, R1052), which includes an on-column DNase I treatment step, and eluted in 30 μL RNase-free dH_2O. cDNA was prepared using the High Capacity cDNA Reverse Transcription Kit (Thermo Fisher #4368814). For detection of transcripts from liver genes in fibroblast

heterochromatin, TaqMan-based quantitative PCR was performed using using the TaqMan assays listed below and TaqMan Gene Expression Master Mix (Thermo Fisher #4369016), and data was normalized to an average of GAPDH and 18SRNA endogenous controls. For detection of transcripts normally expressed in fibroblasts, qPCR was performed using Power SYBR Green PCR Master Mix (Thermo Fisher #4367659), and data was normalized using the GAPDH primer as an endogenous control (see primers listed below). qPCR reactions were run in 384-well format on an 7900HT Real-Time PCR machine (Thermo Fisher #4329001), using the following thermal cycler protocol: 50°C for 2 min, 95°C for 10 min, followed by 45 cycles of 95°C for 15 s then 60°C for 1 min. For SYBR-based qPCR reactions, a dissociation curve was generated to verify that a single PCR product was generated.

Table 3-2. TaqMan Gene Expression Assays used for RT-PCR experiments.

transcript target	**Thermo Fisher Scientific TaqMan ID**
human GAPDH	Hs02758991_g1
human RNA18S5	Hs03928990_g1
human FOXA2	Hs00232764_m1
human NR1H4	Hs01026590_m1
human DSC2	Hs00951428_m1
human DSG2	Hs00170071_m1
human ONECUT1	Hs00413554_m1
human CYP2C9	Hs04260376_m1
human CYP2C19	Hs00426380_m1
human SERPINA7	Hs02384980_m1
human HNF4G	Hs01071345_m1
human F9	Hs01592597_m1

Table 3-3. Primers for SYBR-based RT-PCR experiments.

transcript	forward primer (5' - 3')	reverse primer (5' - 3')
hGAPDH	CCAGGTGGTCTCCTCTGACTTC	TCATACCAGGAAATGAGCTTGACA
hSUV39H1	GTCATGGAGTACGTGGGAGAG	CCTGACGGTCGTAGATCTGG
hSUV39H2	TCGATACGGCAATGTGTCTC	ACAATGCTATTCGGGGAAGA
Hrbmx	CAGTTCGCAGTAGCAGTGGA	TCGAGGTGGACCTCCATAA
hRBMXL1	AGCAGCTCACGTGATGGATA	GATCACTTCGGCTGCTTGAG

Preparation of polyA-selected RNA libraries

Two biological replicates were sequenced per experimental condition. Purified total RNA was diluted to 50 µl in BTE buffer (10 mM Bis-tris, pH 6.7, 1 mM EDTA), denatured by heating at 65°C for 5 minutes, and place immediately on ice. Oligo$(dT)_{25}$ Dynabeads (Thermo Fisher Scientific #61002) were washed three times in 2x Oligo-dT Binding Buffer (2xOBB: 20 mM Tris, pH 7.5, 1 M LiCl, 2 mM EDTA), resuspended in 50 µl 2xOBB, and mixed with an equal volume of denatured RNA. RNA and beads were incubated at room temperature for 10 min, shaking. Beads were washed three times with Oligo-dT Washing Buffer (10 mM Tris, pH 7.5, 150 mM LiCl, 1mM EDTA), and eluted in 10 µl BTE buffer by heating at 80°C for 2 min. Strand-specific cDNA libraries were generated using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs, E7420S). This protocol includes a heat-based mRNA step (15 min at 94°C in first-strand cDNA synthesis buffer), and actinomycin D (Sigma-Aldrich A1410) is added during first-strand cDNA synthesis to inhibit DNA-dependent DNA polymerase activity and prevent template switching. Adapter-ligated cDNAs were amplified by 12 cycles of PCR with NEBNext Multiplex Oligos for Illumina (New England Biolabs E7335S and E7500S).

Library yield and fragment size distribution was assessed on a Bioanalyzer 2100 instrument (Agilent Technologies), using the DNA 1000 kit (Agilent Technologies 5067-1504).

Next-generation sequencing

Libraries were quantified by qPCR using the KAPA Library Quantifcation Kit for Illumina (KAPA Biosystems KK4824). Libraries were diluted to 8 nM concentration and pooled for multiplexing, and then their diluted concentrations were checked a second time using the KAPA kit, adjusting as necessary. Diluted libraries were denatured in 0.2 M NaOH, loaded into the cartridge of the NextSeq 500/550 High Output v2 kit (Illumina FC-404-2005, 75 cycles) at a concentration of 3.2 pM in the kit's Hybridization Buffer, and sequenced in an Illumina NextSeq 500 machine.

3.2 Computational and statistical analyses

Tests of statistical significance

Repeated measurements being compared between two samples were analyzed for significance by two-tailed Student's T-test. Distributions of unequal sizes, such as gene expression values for two different sets of genes, were compared by Wilcoxon rank sum test, implemented in R using the wilcox.test() function (paired=FALSE). Paired distributions (equal size) were analyzed by Wilcoxon signed rank test, using R's wilcox.test() function (paired=TRUE). Correction for multiple comparisons, where noted, were performed using the Benjamini-Hochberg procedure, implemented via DAVID

Bioinformatics tools or the p.adjust() function in R, and a 5% FDR cutoff was applied. For testing the significance of overlaps between two sets drawn from a common pool, the hypergeometric test was used, implemented with the dhyper() function in R.

Alignment and visualization of Gradient-seq and ChIP-seq data

For new sequencing data generated in these studies, sequencer output was demultiplexed (bcl2fastq using BaseSpace) to produce FASTQ files for individual samples. Sequenced reads were aligned to the hg19 assembly of the human genome using bowtie2 v2.1.0 (Langmead and Salzberg, 2012), run with the '--very-sensitive' parameter. Bowtie output files were converted to .bam files using samtools v1.1, and then to .bed files using bedtools v2.20.1 'bamtobed' (Quinlan and Hall, 2010). For each sample, reads mapping to the same genomic position (duplicate reads) were collapsed to a single entry (unique reads). 75bp sequencing reads were extended to 200 bp, to match the average size of DNA prior to library preparation. Meanwhile, all ChIP-seq data obtained from public consortia (see Table 3-4 below) were downloaded from GEO as aligned, unique reads with lengths of 200 bp.

To generate input-normalized genome coverage tracks, BED files were converted to BedGraph files using genomeCoverageBed (bedtools v2.20.1) and normalized to the number of millions of reads sequenced (rpm), to correct for lane or sample biases. For each sample's normalized BedGraph, the normalized BedGraph for the corresponding input sample was subtracted on a basepair-by-basepair basis. The resulting subtracted BedGraph was converted to a bigWig file using bedGraphToBigWig (v4).

Table 3-4. Publicly available ChIP-seq datasets used for genome browser tracks or domain calling.

Listed below are ChIP-seq datasets from the Roadmap Epigenomics Mapping Consortium ("Roadmap") (Bernstein et al., 2010) and the ENCODE project (ENCODE Project Consortium, 2012) that were used for generation of input-subtracted genome browser tracks. Asterisks indicate datasets that were also used for calling enriched genomic domains for a given histone mark.

Human cell type	ChIP type	Source	GEO accession #
foreskin fibroblasts	H3K9me3 *	Roadmap	GSM817236, GSM817239
foreskin fibroblasts	H3K27me3 *	Roadmap	GSM817237, GSM817240, GSM958154
foreskin fibroblasts	Input *	Roadmap	GSM817246, GSM817247, GSM958168
foreskin fibroblasts	H3K4me3	Roadmap	GSM817235, GSM941718, GSM958158
foreskin fibroblasts	H3K36me3	Roadmap	GSM817238, GSM817241, GSM958149
liver	H3K9me3 *	Roadmap	GSM537695, GSM537710, GSM669986
liver	Input *	Roadmap	GSM670008, GSM669910, GSM621629
IMR90 fibroblasts	H3K9me3	Roadmap	GSM469974
IMR90 fibroblasts	Input	Roadmap	GSM521926
skeletal muscle	H3K9me3	Roadmap	GSM621632
skeletal muscle	Input	Roadmap	GSM621641
cord blood $CD3^+$ cells	H3K9me3	Roadmap	GSM537637
cord blood $CD3^+$ cells	Input	Roadmap	GSM537619
naïve $CD8^+$ T cells	H3K9me3	Roadmap	GSM613812
naïve $CD8^+$ T cells	Input	Roadmap	GSM613816
PBMCs	H3K9me3	Roadmap	GSM613878
PBMCs	Input	Roadmap	GSM613893
K562 cells	H3K9me3	ENCODE	GSM733776 (2 replicates)
K562 cells	H3K27me3	ENCODE	GSM733658 (2 replicates)
K562 cells	H3K36me3	ENCODE	GSM733714 (2 replicates)
K562 cells	Input	ENCODE	GSM733780

Calling enriched genomic domains

Genomic domains enriched in ChIP-seq or Gradient-seq datasets were called using a 10-kb sliding window algorithm with a 500-bp sliding step, implemented using custom scripts. For every 10-kb window in the genome, the number of reads for the genomic feature of

interest (normalized to the number of reads sequenced) was divided by the number of reads in that window for the corresponding input file (also normalized to the number of reads sequenced). Divide-by-zero errors were avoided by spiking in a small value into both numerator and denominator (0.25 reads per million reads sequenced). Enriched domains were initially formed by taking all 10-kb windows whose signal-over-input value exceeds a threshold, after averaging all replicates. For human H3K9me3 ChIP-seq (produced by our lab or the Epigenomics Roadmap), the signal-over-input values form a bimodal distribution; consequently, we used kmeans clustering (via the kmeans() function in R) to find the partition between the "unenriched" and "enriched" modes of the distribution. This led to the selection of 1.23 as the threshold for the BJ fibroblast H3K9me3 ChIP-seq produced in this study and 1.41 as the threshold for the foreskin fibroblast H3K9me3 ChIP-seq produced by the Epigenomics Roadmap (see Table 3-4) – values that reflect the different dynamic ranges of the two datasets and yield highly concordant domain maps (86% overlap). Similarly, kmeans clustering was used to select a threshold of 1.30 for the human liver H3K9me3 ChIP-seq from the Epigenomics Roadmap (Table 3-4). For other genomic datasets, fixed enrichment thresholds were applied, instead of kmeans clustering, to ensure fair comparisons. For example, the strucHC sequencing has similar dynamic range as our H3K9me3 ChIP-seq data, and thus the same threshold of 1.23 was applied. For H3K27me3 ChIP-seq data from the Epigenomics Roadmap, the same threshold was used as the Roadmap H3K9me3 data (1.41). Rates of domain overlap among these datasets were similar across a wide range of chosen threshold values and were further corroborated by threshold-free correlation analyses.

Once enriched 10-kb windows were chosen, all adjacent enriched 10-kb windows were merged into contiguous domains. To increase the local resolution of the domain

calls, an edge-pruning step was applied, whereby 500-bp steps were removed from either end of each domain until a 500-bp step is encountered with a signal-over-input value of at least 1.2. (This step serves to eliminate “overhangs” where a local region of strong signal causes an entire 10kb window to be called as enriched. Empirically, this pruning step removes 5% or less of the total domain coverage.) Finally, any remaining overlaps among domains were merged, to produce the final domain calls.

For calling euchromatin domains, a small modification was made because of the close similarity of the euchromatin fraction (containing the majority of chromatin fragments) to the gradient input. To increase contrast for domain-calling, the euchromatin signal was normalized to the strucHC signal (instead of the input). A threshold of 1.2 was applied, based on the bimodal distribution of euchromatin-over-strucHC values. The remainder of the domain-calling procedure was as described above.

Intermediate signal domains were defined by taking all regions outside of strucHC and euchromatin domains that also had sequencing signal in both gradient replicates. Thus, uninformative regions without sequencing data were not assigned a structural type. The strucHC subtype of H3K9me3 or H3K27me3 domains are simply regions of overlap between the strucHC domains and histone mark domains. Among the remaining regions for each histone mark, the euchromatic subtype was called by selecting regions where the exact same 10-kb window met enrichment criteria for both euchromatin and the histone mark. (In other words, it was not sufficient for there merely to be overlap between one euchromatic 10-kb window and a nearby 10-kb window enriched for the histone mark; both properties had to co-occur in the same window, ensuring the stringency of the euchromatic calls.) Finally, remaining regions enriched for the histone mark that were neither strucHC nor euchromatic were called as intermediate.

Correlation analysis of Gradient-seq and ChIP-seq datasets

As for domain calling (see above), the genome was divided into 10-kb sliding windows (500-bp sliding step). For each window, the number of reads for the genomic sample of interest (normalized to the number of reads sequenced) was divided by the number of reads in that window for the corresponding input file (also normalized to the number of reads sequenced). As above, divide-by-zero errors were avoided by spiking in a small value into both numerator and denominator (0.25 reads per million reads sequenced). Using the cor() function in R, we computed the pairwise Spearman correlation among the samples in terms of their signal-over-input values, across all 10-kb windows genome-wide. Spearman correlation values were used to construct a heatmap using the "pheatmap" package in R. The dendrogram distances and clustering were set according to dissimilarity, where dissimilarity = [1 - (Spearman correlation)]. Similar correlations and identical clustering were obtained by Pearson correlation.

Defining genes sets overlapping chromatin domains

The hg19 Refseq gene table was downloaded from the UCSC Table Browser. Refseq genes were defined as "inside" a chromatin domain of interest if at least 50% of the gene body overlapped that domain class. If there were multiple overlaps of the Refseq gene with different domains in that category, these overlaps were summed together, and the Refseq gene was said to overlap the domains as long as the sum met the 50% cutoff. Many Refseq genes have the same gene symbol; for analyses at the gene symbol level (such as gene expression by mRNA-seq), genes symbols were said to overlap the domain if any of their associated Refseq genes met the 50% cutoff. Results regarding gene expression in chromatin domain types were highly invariant to the percentile cutoff chosen.

Computing histone mark enrichment over domains

A large number of ChIP-seq datasets for human fibroblast histone marks were downloaded from GEO (see Table 3-5 below for full list). This included all histone marks profiled in fibroblasts by the Roadmap Epigenomics Mapping Consortium (Bernstein et al., 2010), as well as data from Narita and colleagues (Chandra et al., 2012) to enable inclusion of H3K9me2. The Roadmap data was downloaded as aligned reads, while the Narita lab data was downloaded as raw reads and aligned using bowtie v1.

The following procedure was used to compute histone mark levels within domains of interest, such as strucHC domains (as in Figure 4-2B) or different types of H3K9me3 domains (as in Supplementary Figure 4-3A). First, each domain in the genome was divided into 100 equally sized bins. For each histone mark ChIP-seq experiment, the read pileup per bin was determined. These pileups per bin were then normalized for ChIP-seq sequencing depth (number of millions of reads) and the length of the bin in kb. For each bin of each domain, the normalized pileups were averaged across all replicate ChIP-seq datasets for that mark (from the same cell type and data source). These averages were then subtracted by the results for the corresponding input samples (from the matching replicates/batches of chromatin, also normalized for sequencing depth and bin size). This yielded a topography of histone mark enrichment or depletion across each individual domain, broken into a vector of 100 values for the 100 bins. An average of these vectors was then taken across all domains of that type in the genome, weighted by domain length.

For clarity of display, only the 13 most well-studied histone marks are plotted in Figure 4-2B and Supplementary Figure 4-3A, using the datasets indicated with a checkmark in Table 3-5 below. When available, datasets from primary foreskin fibroblasts were prioritized over IMR90 fibroblasts, given the greater similarity of the former to the BJ

foreskin fibroblasts used in our own experiments. However, we note that results between IMR90 and foreskin fibroblasts were in agreement in all cases, and all acetylated modifications not included in these figures were similarly depleted from strucHC and H3K9me3 domains (data not shown).

Table 3-5. ChIP-seq datasets used for computing histone mark signal over genomic domains.

Datasets used for results presented in Figure 4-2B and Supplementary Figure 4-3A are indicated with a check mark in the final column. Remaining datasets below yielded concordant results (data not shown), as explained above.

ChIP type	Cell type	Source	GEO accession #	Data shown?
H3K4me1	foreskin fibroblasts	Roadmap	GSM817234, GSM941717, GSM958164	✓
H3K4me3	foreskin fibroblasts	Roadmap	GSM817235, GSM941718, GSM958158	✓
H3K9me3	foreskin fibroblasts	Roadmap	GSM817236, GSM817239	✓
H3K27ac	foreskin fibroblasts	Roadmap	GSM1127076, GSM1127060, GSM958163	✓
H3K27me3	foreskin fibroblasts	Roadmap	GSM817237, GSM817240, GSM958154	✓
H3K36me3	foreskin fibroblasts	Roadmap	GSM817238, GSM817241, GSM958149	✓
Input	foreskin fibroblasts	Roadmap	GSM817246, GSM817247, GSM958168	✓
H3K4me2	IMR90 fibroblasts	Roadmap	GSM521899, GSM521900	✓
H3K9ac	IMR90 fibroblasts	Roadmap	GSM469973, GSM521912	✓
H3K9me1	IMR90 fibroblasts	Roadmap	GSM752986, GSM752987	✓
H3K79me1	IMR90 fibroblasts	Roadmap	GSM521904, GSM521906, GSM521907, GSM521908	✓
H3K79me2	IMR90 fibroblasts	Roadmap	GSM521909, GSM521911	✓
H4K20me1	IMR90 fibroblasts	Roadmap	GSM521915, GSM521917	✓
Input	IMR90 fibroblasts	Roadmap	GSM521926, GSM521927, GSM521928, GSM521929, GSM521930, GSM521931, GSM521932, GSM521933	✓
H3K9me2	IMR90 fibroblasts	Chandra et al., 2012	GSM942082, GSM942084	✓
Input	IMR90 fibroblasts	Chandra et al., 2012	GSM942119	✓
H3K9me3	IMR90 fibroblasts	Chandra et al., 2012	GSM942075	
H2AK5ac	IMR90 fibroblasts	Roadmap	GSM521866, GSM521868	
H2AK9ac	IMR90 fibroblasts	Roadmap	GSM818012, GSM818013	

H2BK5ac	IMR90 fibroblasts	Roadmap	GSM818017, GSM832837, GSM832838	
H2BK12ac	IMR90 fibroblasts	Roadmap	GSM521871, GSM521873, GSM521874	
H2BK15ac	IMR90 fibroblasts	Roadmap	GSM521875, GSM521877, GSM521878	
H2BK20ac	IMR90 fibroblasts	Roadmap	GSM521879, GSM521880	
H2BK120ac	IMR90 fibroblasts	Roadmap	GSM521869, GSM521870	
H3K4ac	IMR90 fibroblasts	Roadmap	GSM521893, GSM521894	
H3K4me1	IMR90 fibroblasts	Roadmap	GSM521895, GSM521897, GSM521898	
H3K4me3	IMR90 fibroblasts	Roadmap	GSM469970, GSM521901	
H3K9me3	IMR90 fibroblasts	Roadmap	GSM469974, GSM521913, GSM521914	
H3K14ac	IMR90 fibroblasts	Roadmap	GSM521881, GSM521883	
H3K18ac	IMR90 fibroblasts	Roadmap	GSM469965, GSM521884	
H3K23ac	IMR90 fibroblasts	Roadmap	GSM521885, GSM521886	
H3K27ac	IMR90 fibroblasts	Roadmap	GSM469966, GSM469967, GSM521887	
H3K27me3	IMR90 fibroblasts	Roadmap	GSM469968, GSM521889	
H3K36me3	IMR90 fibroblasts	Roadmap	GSM521890, GSM521892	
H3K56ac	IMR90 fibroblasts	Roadmap	GSM521902, GSM521903	
H4K5ac	IMR90 fibroblasts	Roadmap	GSM469975, GSM521918	
H4K8ac	IMR90 fibroblasts	Roadmap	GSM521919, GSM521921, GSM521922, GSM521923	
H4K91ac	IMR90 fibroblasts	Roadmap	GSM521924, GSM521925	

Analysis of repetitive elements enriched in chromatin domains

The Repeat Masker table for hg19 was downloaded from the UCSC table browser and was intersected with BED files listing non-overlapping domains, using Bedtools. For each type of repeat, the total number of base pairs falling within domains was computed as a percentage of the total genomic coverage of that repeat. This analysis was performed for each repeat class, repeat family, and individual repeat name listed in the Repeat Masker table. P-values were determined by permutation test, using 1000 simulations of domains randomly shuffled across the genome by Bedtools (preserving domain number and size). For each repeat, the P-value was the proportion of domain simulations that produced an equal or greater overlap with that repeat. To correct for multiple hypothesis testing, P values were used to calculate the FDR using the Benjamini-Hochberg procedure. Enriched repeats with an FDR < 0.05 were treated as statistically significant.

Analysis of DNA methylation in chromatin domains

Processed whole-genome bisulfite sequencing (WGBS) data for human foreskin fibroblasts was downloaded from GSM1127120 (Roadmap Epigenomics Consortium et al., 2015). This data file reports the percent methylation per CpG. CpGs were then divided into two categories based on whether they fell in annotated CpG islands (UCSC Table Browser). Both categories of CpG sites (inside and outside of CpG islands) were intersected with chromatin domains of interest, and the distribution of percent-methylation values was plotted.

Analysis of datasets for Lamin B1 ChIP-seq, DNase-seq, and MNase sensitivity

Lamin B1 ChIP-seq reads and input reads for IMR90 fibroblasts (Dou et al., 2015) were downloaded from GSE63440, aligned using bowtie2 v2.1.0 (parameters: --very-sensitive), and extended to 200 bp. PCR duplicates were collapsed to unique reads. The enrichment of Lamin B1 ChIP reads to input reads (after normalizing for sequencing depth) was computed over 10-kb genomic windows, with a 500-bp sliding step, using custom scripts. Lamin-B1 enrichment values over input were averaged across the two replicates and were plotted for all 10-kb windows falling entirely inside a chromatin domain type of interest.

ENCODE DNase-seq data for BJ fibroblasts (two replicates) were downloaded as aligned reads GSM736518 and GSM736596 (Thurman et al., 2012). The aligned reads were binned into 10-kb windows, with a 500-bp sliding step, and normalized based on the number of millions of reads sequenced. These normalized pileups were plotted for all 10-kb windows falling entirely inside a chromatin domain type of interest.

MACC scores, which represent sensitivity to MNase from an enzymatic titration, were downloaded for human K562 cells from GSE78984 (Mieczkowski et al., 2016).

MACC scores were provided at 500-bp resolution, and these scores were converted to 10-kb sliding window scores (500-bp slide) by averaging the component 500-bp values within each 10-kb window. During this averaging, missing values were ignored (rather than counting them as zero), and data was not reported for the 10-kb window if fewer than five of the constituent 500-bp bins contained data. The 10-kb MACC scores were then plotted for chromatin domains, similarly to the Lamin B1 and DNase I data.

Analysis of DNA replication timing data

ENCODE Repli-seq data for BJ fibroblasts (Pope et al., 2014) were downloaded as aligned reads from ENCSR894LZX. These files contain sequencing of newly replicated DNA in six cytometry-fractionated cell populations, with two replicates per timepoint. The downloaded alignments were intersected with each domain BED file using bedtools, and the number of intersecting reads was divided by the number of millions of reads in the sequencing file. These normalized intersection scores were then averaged between the two sequencing replicates per timepoint, and then they were expressed as a fraction of the sum of the intersection scores across the six timepoints. Thus, the final values for each domain type sum to 1.0 across the six cell cycle fractions.

Analysis of RNA microarray data for hiHep reprogramming

Microarray data for human fibroblasts, cultured human hepatocytes, hiHep cells, and immortalized hiHeps were downloaded from GSE42643 (Huang et al., 2014). The microarray data was quantile-normalized with median polish using the Partek Genomics Suite. For Refseq genes with multiple probes, only the probe with the highest variance across all samples was used. The full microarray was filtered down to genes expressed

significantly higher in normal hepatocytes compared to fibroblasts ($P < 0.05$, at least 2-fold). For analyses specifically related to silent genes in fibroblasts (eg, Extended Data Figure 1A), the list was further reduced to genes expressed in the bottom 40% among in fibroblasts, which corresponded to the lower mode of a biomodal distribution. Log2-normalized gene expression for hiHep cells was calculated on a relative scale, with 0% representing the log2-normalized fibroblast expression, and 100% representing the log2-normalized hepatocyte expression. Negative values (hepatic genes expressed lower in hiHeps than fibroblasts) were rounded up to 0%. Expression values on this scale were plotted as violin plots in R using the vioplot package, modified to display multiple colors.

Analysis of RNA microarray data for patient-derived fibroblasts

Exon microarray data were downloaded from GSE33855 (Highley et al., 2014) and quantile-normalized with median polish using the Partek Genomics Suite. This dataset includes fibroblasts from patients with Amyotrophic Lateral Sclerosis (ALS), some with germline mutations and some with sporadic disease, in addition to healthy controls. To extract whole gene-level expression from exon-directed probes (multiple per gene), we analyzed the microarray as described in the original publication (Highley et al., 2014): exon-specific probes were removed from consideration if their log2-normalized signal (averaged across all biological samples) was 3 standard deviations above or 1 standard deviation below the average signal for all the probes for that gene. This served to remove probes with nonspecific hybridization or affected by alternative splicing, respectively. An average was then taken of the remaining probes for each gene, for each patient sample. These log2-transformed values were then converted back to a linear scale and averaged

across the patients for a given genotype – TARDBP-mutated (n=3), sporadic ALS (n=6), or control (n=6) – in order to calculate, for each gene, the fold-change versus control.

Alignment and visualization of mRNA-seq data

Sequencer output was demultiplexed (bcl2fastq using BaseSpace) to produce FASTQ files for individual samples. Sequenced reads were aligned to the hg19 Refseq gene model, in a strand-specific manner, using TopHat v2.0.11 (parameters: --b2-very-sensitive --library-type fr-firststrand) (Kim et al., 2013). To generate genome coverage tracks, BED files were first pooled between biological replicates. Reads aligning to genomic regions longer than the sequencing length (75 bp), due to spanning of splice junctions, were discarded for genomic visualization purposes. Pooled BED files were then converted to BedGraph files using genomeCoverageBed (bedtools v2.20.1) and normalized to the number of millions of reads sequenced (rpm), to correct for lane or sample biases. The resulting subtracted BedGraph was converted to a bigWig file using bedGraphToBigWig (v4).

Gene expression analysis for mRNA-seq data

Sequenced reads were assigned in a strand-specific manner to genes from the hg19 Refseq table using HTSeq v0.6.1 (parameters: --stranded=reverse --mode=intersection-nonempty --type=exon -i gene_id) (Anders et al., 2015), such that all exons for all Refseq genes with the same official gene symbol were considered to belong to the same feature. The unnormalized HTSeq tables, after removal of lines for unassigned reads, were analyzed by the DESeq2 v1.11.45 (Love et al., 2014) package in R to produce normalized count tables. All samples for both the “hepatic-TF” and “no-TF” RNA-seq studies were

normalized in DESeq2 at the same time. For quantification of fibroblast gene expression across chromatin categories, the DESeq2-normalized counts for the control siRNA-transfected fibroblasts in the no-TF condition were used, after dividing each value by the length of that gene's exon model in kb and averaging the two biological replicates. Differentially expressed genes were determined in a pairwise manner using DESeq2, with a significance cutoff of p.adjust < 0.05. For analysis of genes upregulated by siRNAs compared to control siRNA in the "hepatic TF" condition, genes significantly downregulated by hepatic TFs alone (hepatic-TF control siRNA compared to no-TF control siRNA) were removed from consideration – 3,487 out of 26,839 total genes. This was to ensure that upregulated genes could be interpreted as being truly upregulated by the siRNA, rather than the siRNA treatment inhibiting the downregulation of the gene by the hepatic TFs. For mRNA-seq sample correlation analyses, a regularized log2-transformation was performed on the normalized counts using DESeq2, and Euclidean distances were calculated using the dist() function in R and visualized as a heatmap/dendrogram using the "pheatmap" package.

Defining protein sets enriched in heterochromatin and gradient top

Intensity-based absolute quantification (iBAQ) values were ranked in order of their iBAQ score for each of three biological replicates of each sample, to facilitate comparison among samples with very different numbers of identified proteins. Technical replicate MS runs were averaged. Proteins with only a single detected peptide were removed from consideration, and common contaminants (immunoglobulin chains, skin keratins) were filtered out. Protein ranks were compared between samples by two-tailed Student's T test. Proteins with a significantly higher rank ($P < 0.05$) in the strucHC+H3K9me3 sample

(H3K9me3-directed IP using strucHC fraction chromatin) compared to the Gradient Top fraction were used to define the “heterochromatin proteins.” Proteins detected in strucHC+H3K9me3, but without significant enrichment or depletion compared to the Gradient Top sample, were used to define the “shared” proteins. Meanwhile, for the strucHC whole fraction sample, proteins with significantly higher rank in the strucHC fraction compared to the Gradient Top fraction were defined as “strucHC enriched.” Finally, remaining proteins were classified as “Gradient Top” proteins if they were either significantly enriched in the Gradient Top fraction over strucHC+H3K9me3, significantly enriched in Gradient Top over strucHC, or unique to the Gradient Top fraction. Note that proteins not detected in the Gradient Top fraction were assigned a rank equal to the total number of proteins detected for that replicate; thus, proteins detected uniquely in strucHC+H3K9me3 or the strucHC fraction could still be quantified as significantly enriched. Note also that the significance by T-test requires at least two values per sample being compared, and thus the “heterochromatin” and “strucHC-enriched” proteins by definition had to have been detected in at least two out of three replicates of the strucHC+H3K9me3 and strucHC samples, respectively.

3.3 Genomic and proteomic dataset availability

All next-generation sequencing data generated by this study (FASTQ files of sequenced reads and bigwig files for browser visualization) were uploaded to GEO accession number GSE87041, and they will be made available upon publication. Proteomic raw data files are available on the Chorus database under Project ID 1172, Experiment ID 2585.

CHAPTER 4. ISOLATION OF HETEROCHROMATIN REVEALS STRUCTURAL SUBTYPES AND DIVERSE PROTEINS WITHIN REPRESSED CHROMATIN DOMAINS

Justin S. Becker[1-3], Simone Sidoli[2,4], Greg Donahue[1-3], Kelsey E. Kaeding[1-3], Zhiying He[1,3], Shu Lin[2,4], Benjamin A. Garcia[2,4], and Kenneth S. Zaret[1-3]

[1]Institute for Regenerative Medicine, [2]Epigenetics Program, [3]Department of Cell and Developmental Biology, [4]Department of Biochemistry and Molecular Biophysics

Perelman School of Medicine, University of Pennsylvania

Smilow Center for Translational Research, 3400 Civic Center Boulevard, Philadelphia, Pennsylvania 19104, USA.

Running title: Subtypes of structural heterochromatin

Corresponding author: Ken Zaret

e-mail: zaret@upenn.edu

Telephone: 215-573-5813

Fax: 215-573-5844

4.1 RESPECTIVE CONTRIBUTIONS

The majority of the experiments and analyses in this chapter were performed by me alone, with guidance from Kenneth S. Zaret. Simone Sidoli and Shu Lin, with guidance from Benjamin A. Garcia, performed the mass spectrometry experiments and the quantitative proteomics analysis on peptide samples that I provided. Kelsey Kaeding prepared and sequenced the mRNA-seq libraries for the transcriptome analysis after siRNA treatment, using RNA samples that I prepared, and she also validated the efficiency of the knockdowns by RT-PCR. Zhiying He assisted with some of the hiHep reprogramming experiments. Greg Donahue provided custom scripts and advice for computational analyses, worked directly with me for alignment of next-generation sequencing data, and performed the analysis of whole-genome bisfulfite sequencing data. Computational analyses performed by primarily by me included the mapping and analysis of chromatin domains enriched for various features, the differential gene expression analysis for mRNA-seq, and the analysis of available microarray data for studying hiHep reprogramming or the consequences of TARDBP mutation. The manuscript was written by me and Dr. Zaret.

4.2 ABSTRACT

Heterochromatic regions in mammalian cells suppress recombination, silence transcription, and are crucial for maintaining cell differentiation. Genomic and biochemical characterization of heterochromatin has relied on the associated histone modifications H3K9me3 and H3K27me3, yet these marks are also found in euchromatic regions that permit transcription. We employed a biophysical method to isolate structurally compact heterochromatin from human somatic cells, mapped its genomic organization compared to histone modifications, and used proteomics to reveal an extensive number of heterochromatin-bound proteins. We discriminate subtypes of H3K9me3- and H3K27me3-marked domains, in structural heterochromatin versus euchromatin, and we present a resource of hundreds of proteins that preferentially bind heterochromatin, revealing an enrichment for RNA-binding proteins and proteins that oppose iPS reprogramming. The structural heterochromatin landscape includes repressed genes for alternative lineages that are resistant to activation by introduced transcription factors. Depletion of identified heterochromatin-associated proteins reduces this barrier, rendering alternative-lineage genes more competent for transcriptional activation.

4.3 INTRODUCTION

Cell differentiation is achieved, in part, through the selective repression of transcription of a large fraction of the genome. Among such silent domains, regions of heterochromatin are distinguished by their high levels of physical compaction, causing intense staining with DNA dyes or by electron microscopy (Heitz, 1928; Underwood et al., 2016). The condensed structure of heterochromatin restricts the accessibility of DNA to binding by transcription factors (Soufi et al., 2012) and allows persistent gene silencing, despite the presence of transcriptional activators (Ayyanathan et al., 2003; Hathaway et al., 2012). The packaging of DNA into dense chromatin structures also suppresses recombination at repeat-rich sequences, promoting genome stability (Peters et al., 2001). But lacking has been a direct method to identify genomic domains with high physical compaction while allowing recovery and analysis of constituent proteins.

A hallmark of repeat-rich heterochromatic regions, conserved from fission yeast to humans, is the dimethylation and trimethylation of histone 3 lysine 9 (H3K9me2 and H3K9me3). In mammals, the methyltransferases SUV39H1/SUV39H2 and SETDB1 are responsible for H3K9me2 and H3K9me3 (Peters et al., 2001; Schultz et al., 2002), while G9a and GLP catalyze only H3K9me2 (Tachibana et al., 2005). The direct binding of H3K9me2/3 by heterochromatin protein 1 (HP1) recruits additional repressive complexes (Eissenberg and Elgin, 2014). Human fibroblasts contain broad, megabase-scale domains of H3K9me3 that include repressed genes for neural function and adhesion, as well as developmental transcription factors (Hawkins et al., 2010; Soufi et al., 2012). H3K9me3 and HP1 are also deposited over zinc finger (ZNF) genes that contain Krüppel-associated box (KRAB) domains, which is dependent on the recruitment of SETDB1 by the KRAB-

ZNFs themselves and their co-repressor KAP1 (Schultz et al., 2002; Vogel et al., 2006; O'Geen et al., 2007). The global pattern of H3K9me3 domains shows partial rearrangement across different cell lineages (Soufi et al., 2012; Becker et al., 2016), but the protein machinery that mediates the developmental dynamics of large-scale H3K9me3 domains is largely unknown..

Repressed genes also can be associated with H3 lysine 27 trimethylation (H3K27me3), catalyzed by the Polycomb Repressive Complex 2 (PRC2) (Margueron and Reinberg, 2011). This mark is prominent on the inactive X chromosome and at promoters for developmental transcription factors (Ezhkova et al., 2009; Hawkins et al., 2010; Xu et al., 2014a). However, H3K27me3-marked sites can be accessible to general transcription factors and RNA polymerase (Breiling et al., 2001; Dellino et al., 2004) and the literature is conflicted on whether the term "heterochromatin" includes H3K27me3-marked regions (Trojer and Reinberg, 2007; Beisel and Paro, 2011).

We previously described megabase-scale, H3K9me3-enriched domains that are resistant to binding by the pluripotency transcription factors Oct4, Sox2, Klf4, and cMyc (OSKM) in human fibroblasts, but where the factors bind in human embryonic stem (ES) cells (Soufi et al., 2012). These Differentially Bound Regions (DBRs) contain pluripotency genes whose activation during iPS reprogramming is restricted to rare cells late in the reprogramming process (Buganim et al., 2012). Reducing H3K9me3 levels, by depleting SUV39H1/H2 methyltransferases, improves the efficiency of iPS reprogramming (Onder et al., 2012; Soufi et al., 2012) as well as reprogramming by somatic cell nuclear transfer (Matoba et al., 2014). However, the chromatin components that mediate the resistance of H3K9me3 heterochromatin to gene activation during reprogramming are poorly

understood, and whether H3K9me3 heterochromatin impedes reprogramming between differentiated lineages is not clear.

To discover proteins embedded in heterochromatin, mass spectrometry (MS) has been performed after H3K9me3-directed chromatin immunoprecipitation (ChIP) (Soldi and Bonaldi, 2013; Engelen et al., 2015; Ji et al., 2015) or pulldown with a H3K9me3 bait (Vermeulen et al., 2010; Eberl et al., 2013), but the relationship between H3K9me3-associated proteins and heterochromatin per se is unclear. In two cases, proteomic analysis was performed in mouse pluripotent ES cells (Engelen et al., 2015; Ji et al., 2015), a cell type found to lack compacted blocks of heterochromatin by high-resolution imaging techniques (Fussner et al., 2011; Underwood et al., 2016). More importantly, H3K9me3 and HP1 are found outside of heterochromatin, including in the gene bodies of expressed genes, and can promote transcriptional elongation (Piacentini et al., 2003; Vakoc et al., 2005; Riddle et al., 2012). Gene clusters on *Drosophila* chromosome 4q and human ZNF clusters on chromosome 19 contain H3K9me3 and HP1, despite simultaneous transcription (Riddle et al., 2012; Vogel et al., 2006; Blahnik et al., 2011). A definitive accounting of transcriptionally permissive H3K9me3 domains is currently lacking. Although RNA-binding proteins appear abundant after H3K9me3 ChIP (Soldi and Bonaldi, 2013) or among readers of the H3K9me3 mark (Vermeulen et al., 2010), it is unclear of these proteins are specifically associated with the transcriptionally repressed, heterochromatic form of H3K9me3-marked chromatin.

In this study, we find that H3K9me3 domains, more than H3K27me3 domains, impede the activation of liver genes during reprogramming of fibroblasts to hepatocytes. In our investigation of H3K9me3 domains, we found that such regions are resistant to sonication, which causes their underrepresentation in ChIP-seq, and which allowed us to

physically isolate heterochromatic domains on sucrose gradients. We performed quantitative genomics and proteomics on the isolated heterochromatin, as well as functional analysis of selected heterochromatin proteins. The results provide resource maps of subsets of H3K9me3 and H3K27me3 domains that are in euchromatin versus heterochromatin, and they reveal unexpected complexity in the proteins that are enriched in what we define as "structural heterochromatin." We show that structural heterochromatin proteins identified by our approach inhibit the activation of developmentally silenced genes, establishing the proteins as functional regulators of heterochromatin.

4.4 RESULTS

Silent genes in H3K9me3 domains are more resistant to reprogramming than silent genes in H3K27me3 domains

Human fibroblasts can be reprogrammed to a hepatocyte-like fate by the ectopic expression of liver transcription factors (Huang et al., 2014; Du et al., 2014). The resulting human induced hepatocytes (hiHeps) express hepatic markers and recapitulate some features of liver metabolism, but expression profiling reveals a failure to fully express all hepatocyte genes (Huang et al., 2014; Du et al., 2014). We analyzed a published microarray comparing hiHep cells to fibroblasts and cultured human hepatocytes (Huang et al., 2014), focusing on genes that are silent in fibroblasts and upregulated in hepatocytes. We found that genes marked by H3K9me3 in the starting fibroblasts have a profound failure to activate during cell conversion to hiHeps (Supp. Fig. 4-1A), with the median barely above 0% induction. By contrast, genes marked by H3K27me3, another

mark associated with transcriptional silence (Margueron and Reinberg, 2011), show no greater defect in activation than genes lacking either mark (Supp. Fig. 4-1A). The H3K9me3-marked genes in fibroblasts that fail to activate include genes that crucial for terminal hepatocyte differentiation, including *NHR1H4* (encoding the bile acid receptor FXR), *FOXA2*, metabolic enzymes, cytochrome P450 subunits, secreted plasma proteins, and epithelial adhesion proteins. We conclude that H3K9me3-marked chromatin is the most refractory to gene activation by hepatic factors, similar to that seen at DBRs for reprogramming to pluripotency (Soufi et al., 2012; Matoba et al., 2014), leading us to further investigate the nature H3K9me3 domains in the human fibroblasts.

Heterochromatic Differentially Bound Regions are resistant to sonication

Using data from the NIH Epigenomics Roadmap (Bernstein et al., 2010), we find that genome mapping tracks for input-subtracted H3K9me3 show pronounced enrichment over DBRs (Figure 4-1A, purple track), but raw H3K9me3 signals are only modestly elevated in these regions (Figure 4-1A, yellow track). Surprisingly, the input signals alone show depletion over most H3K9me3 domains, including DBRs (Figure 4-1A, blue track, red arrows). Sequencing of both input and ChIP is typically performed after size-selecting short DNA fragments, thereby depleting regions that are more resistant to sonication. A study reported that megabase-scale H3K9me3 domains are largely dependent on culture conditions, but the ChIP data was analyzed without normalizing to input from the same sample (Zhu et al., 2013). Our own input-normalized analysis of the same ChIP-seq datasets showed that megabase-scale H3K9me3 domains are present at shared and different sites across diverse human cell types (Supp. Fig. 4-1B) (Becker et al., 2016).

Sites of high sonication susceptibility map to active promoters (Auerbach et al., 2009), but resistance to sonication has not been used to identify heterochromatic regions.

We prepared crosslinked, sonicated chromatin from human BJ foreskin fibroblasts, in which the DBRs were mapped (Soufi et al., 2012) and used sucrose gradient ultracentrifugation under conditions that separate chromatin fragments by size (Gilbert et al., 2004) (Figure 4-1B). By qPCR, we compared the migration sites previously shown to be functionally euchromatic versus heterochromatic in fibroblasts: euchromatic sites outside DBRs that are bound by Oct4/Sox2 in fibroblasts and ES cells, versus heterochromatic sites inside DBRs that are bound in ES cells but lack binding in fibroblasts (Soufi et al., 2012). DBR sites (Figure 4-1C, red) are significantly depleted from the fractions at the top of the gradient, which contain the most highly sheared DNA. Meanwhile, fractions from the middle of the gradient, containing longer DNA fragments normally excluded from ChIP-seq (Figure 4-1B), were selectively enriched for DBR sites (Figure 4-1C). The magnitude of DBR enrichment in middle gradient fractions is comparable to that achieved by conventional H3K9me3 ChIP-qPCR (Supp. Fig. 4-1C). Thus, gradient-mediated separation of sonication-resistant chromatin provides a biophysical readout of chromatin regions of heterochromatic domains.

We pooled the fractions of the gradient showing the greatest enrichment of DBR sites (fractions #10-17, red box in Figure 4-1B). This material was termed the “structural heterochromatin (strucHC) fraction” and was compared to a “euchromatin fraction” from the top of the gradient (fraction #2) in all subsequent analyses. Secondary H3K9me3 ChIP, performed from the strucHC fraction, resulted in additive enrichment of DBR sites (Supp. Fig. 4-1C). The chromatin recovery by H3K9me3 IP, as a fraction of input, was much greater for the strucHC fraction compared to bulk chromatin or the euchromatin

fraction (Supp. Fig. 4-1D), confirming that the gradient step enriches for H3K9me3-marked chromatin.

Genomic mapping of the structural heterochromatin landscape

We sequenced DNA from the strucHC and euchromatin fractions of the gradient ("Gradient-seq") (Figure 4-1D). For samples containing larger DNA fragments, including strucHC fractions, IPs off of this fraction, and input to the gradient, the purified DNA was sonicated further (Supp. Fig. 4-2A,B). In parallel, we performed conventional input-normalized H3K9me3 ChIP-seq and mRNA-seq of BJ fibroblast cells.

We found that of the 258 DBRs (Soufi et al., 2012), 256 are positively enriched in strucHC over input chromatin, in contrast to similarly sized regions that flank DBRs (Figure 4-1E). Browser views show that the strucHC sequencing signals form megabase-scale domains that correspond closely to H3K9me3 ChIP-seq domains, including DBRs, while the euchromatin fraction exhibits an inverse pattern (Figure 4-1F). Meanwhile, many H3K27me3-marked sites overlap with regions of euchromatin signal, while others show strong strucHC enrichment even in the absence of H3K9me3 (Figure 4-1F, green arrows). To quantitatively compare these samples without imposing numerical thresholds, we computed the pair-wise Spearman correlation coefficients among the samples after binning the input-normalized data into 10 kb windows. Genome-wide, the strucHC signal has a strong correlation with H3K9me3 ChIP-seq from either this study or the Epigenomics Roadmap (Figure 4-2A). Meanwhile, strucHC has weaker (but still positive) correlation with H3K27me3. Finally, the strucHC fraction and H3K9me3/H3K27me3 ChIP samples all anti-correlate with the euchromatin fraction (Figure 4-2A).

To identify genomic domains that were enriched in each sample (horizontal colored bars above each track in Figure 4-1F), we used a sliding window algorithm to select 10 kb intervals that exceeded a threshold for signal divided by input, followed by edge-pruning of enriched regions in 500-bp steps (see Chapter 2). The identified domains for strucHC cover 997 MB of the hg19 genome assembly and have weighted average length 135 kb. Euchromatin domains, which total 1240 MB, were identified using the same procedure, except the euchromatin sequencing was compared directly to strucHC, which provided greater contrast than input. Finally, mappable regions not in strucHC or euchromatin domains, meaning that they had comparable abundance in the two gradient fractions, were classified as intermediate domains.

To assess histone mark levels in strucHC domains, we downloaded ChIP-seq data for 27 marks profiled in human fibroblasts by the Roadmap Epigenomics consortium (Bernstein et al., 2010) as well as ChIP-seq for H3K9me2 (Chandra et al., 2012). We computed the enrichment or depletion of each mark, relative to input, over the width of each strucHC domain and plotted the average across all domains (Figure 4-2B). H3K9me3 is most strongly enriched in strucHC, while H3K27me3 and H3K9me2 have modest levels of enrichment. All other marks tested, including those associated with active promoters/enhancers or transcriptional elongation, are strongly depleted (Figure 4-2B). Consistent with these associations, 607 MB of the strucHC domains (60.8%) are also called as H3K9me3 domains, while 327 MB (32.8%) are H3K27me3 domains (Figure 4-2C). Interestingly, these two molecular forms of strucHC domains are largely distinct, with only 70 MB of overlap (7.0%), reminiscent of ChIP-Seq data (Hawkins et al., 2010; Chandra et al., 2012).

Genes within strucHC domains are transcriptionally repressed, in comparison to genes falling inside euchromatic domains, with the intermediate domains nearly as repressed as strucHC (Figure 4-2D). StrucHC is enriched for genes important for non-fibroblast lineages, including genes for cell surface (neuronal ion channels, G-protein-coupled receptors, and epithelial adhesion proteins), immune defense proteins, factors for endodermal and ectodermal development, keratins, muscle components, and enzymes for digestion and hepatic metabolism (Figure 4-2E). As expected, strucHC domains contain the majority of satellite repeat sequences and are significantly enriched for LINE elements and endogenous retroviruses. Using whole-genome bisulfite sequencing (WGBS) data from the Roadmap Epigenomics consortium (Bernstein et al., 2010), we find that DNA methylation rates at CpGs islands are highly divergent between the structural classes (Figure 4-2F, left). Domains of strucHC have the highest rate of methylation at CpG islands (median CpG is methylated at ~75% of copies), while at euchromatic CpG islands three-quarters of CpGs have no detectable methylation (Figure 4-2F, left). Meanwhile, outside of CpG islands, rates of DNA methylation are high in all three categories (Figure 4-2F, right), consistent with methylation being widespread outside of islands (Jones, 2012). Non-CpG-island methylation is actually highest in euchromatin (Figure 4-2F, right), perhaps because of the DNA methylation associated with transcribed gene bodies (Jones, 2012).

Gradient-seq reveals euchromatic subtypes of repressive histone mark domains

We observed that ZNF gene clusters on chromosome 19 have high levels of H3K9me3, despite expression of the genes (Figure 4-3A), as reported previously (Vogel et al., 2006; Blahnik et al., 2011). Notably, the expressed ZNF gene clusters are depleted from the

strucHC fraction and enriched in the euchromatin fraction (Figure 4-3A). These regions are further enriched by H3K9me3 IP off of the euchromatin fraction (Figure 4-3A), confirming that a sonication-sensitive chromatin structure can coexist with the H3K9me3 mark on the same chromatin fragment.

To identify similar regions on a genome-wide basis, we selected all 10-kb intervals that were enriched in the euchromatin fraction over strucHC fraction and for domains of H3K9me3 or H3K27me3. These euchromatic subtypes of these histone mark domains occupied 31 MB of sequence for H3K9me3 (3.2% of the total H3K9me3) and 193 MB (28.7%) for H3K27me3 (Figure 4-2C). Most of the genes in euchromatic H3K9me3 or H3K27me3 domains were expressed, although at lower levels than the euchromatin as a whole (Figure 4-3B). Consistent with this, euchromatic H3K9me3 domains, but not other H3K9me3 domains, were enriched for H3K36me3 (Supp. Fig. 4-3A), a mark associated with transcriptional elongation (Vakoc et al., 2006). At transcribed ZNF genes, H3K9me3 and H3K36me3 are both observed in the gene bodies, but the H3K9me3 signal is reduced or absent over the promoters (Supp. Fig. 4-3B).

The euchromatic H3K9me3 domains had robust selectivity for gene family clusters, with more than half of the genes inside these domains belonging to the KRAB-ZNF family (n=150 of 230 InterPro database matches), in addition to the protocadherin gamma cluster and certain homeobox (HOX) genes (Figure 4-3C). Genes in Euchromatic H3K9me3 are listed in Table 4-1. The 3% of fibroblast H3K9me3 domains that are euchromatic contain a majority of sites regulated by the HUSH complex (Tchasovnikarova et al., 2015), an enrichment of 39-fold compared to the strucHC subtype (Figure 4-3D). Indeed, the sites reported to have the greatest dependence on HUSH/SETDB1 for H3K9me3 levels fall in euchromatin (Supp. Fig. 4-3C), consistent with the complex localizing to genes that are

"dimmed" but nonetheless expressed (Tchasovnikarova et al., 2015). Euchromatic H3K27me3 domains include all four human HOX gene clusters (Figure 4-3E) and keratin type I and protocadherin gamma clusters, in addition to non-clustered genes for basic-helix-loop-helix and forkhead transcription factors (Figure 4-3C).

While as expected, the majority of satellite repeats fall in strucHC, the specific repeat type HSATII, which due to its constitutive silence might otherwise be classified as heterochromatic, is consistently found in euchromatic H3K9me3 domains (Figure 4-3F). That HSATII has an accessible structure is corroborated by its enrichment in available tracks of DNase-I sensitivity (Supp. Fig. 4-3D). A unique feature of HSATII is its massive overexpression in diverse human cancers (Ting et al., 2011) and cellular senescence (De Cecco et al., 2013). Heterochromatic regions have been found to have an elevated rate of mutation accumulation across human cancers (Lawrence et al., 2013), and we find that this effect is more pronounced in the strucHC and intermediate H3K9me3 subtypes, compared to euchromatic H3K9me3 or H3K27me3-marked regions (Supp. Fig. 4-3E). Thus, Gradient-seq provides structural insight into different repressed domains associated with human disease.

Relationship between strucHC and other markers of heterochromatin

Our finding that chromatin regions with the same repressive histone marks could be partitioned into structural subtypes with markedly different sonication sensitivity as measured by Gradient-seq (Figure 4-4A) led us to investigate to what extend to those differences are reflected in other properties associated with heterochromatin. The structural categories were indeed highly predictive of the levels of CpG methylation at CpG islands, as quantified by WGBS (Bernstein et al., 2010) (Figure 4-4B). Both inside and

outside H3K9me3 domains, methylation rates were consistently high at strucHC islands and low at euchromatic islands, with moderate levels in the intermediate category (Figure 4-4B), correlating with the levels of gene silencing for these subtypes (Figure 4-3B). Interestingly, methylation rates are highest in the 133 MB of strucHC that lack enrichment for H3K9me3 or H3K27me3 (Figure 4-4B, far right), confirming that the strucHC domains have heterochromatic features even where they diverge from traditional histone mark-based approaches. By contrast, outside of CpG islands, methylation rates are high in all categories (Supp. Fig. 4-4A), and highest in the euchromatin subtypes, similar to that observed for the total strucHC and euchromatin domains (Figure 4-2F).

We find that association of chromatin with the nuclear lamina, meanwhile, is related to both structural classification and the type of histone mark. Levels of Lamin B1 binding by ChIP-seq (Dou et al., 2015) are generally higher in H3K9me3 domains compared to H3K27me3 domains (Figure 4-4C), consistent with previous reports (Guelen et al., 2008). However, within H3K9me3 domains, the strucHC and intermediate subtypes are much more consistently enriched for Lamin B1 binding (Figure 4-4C) and overlap at a higher rate with previously mapped Lamina Associated Domains (Guelen et al., 2008) (Supp. Fig. 4-4D), compared to the euchromatic subtype of H3K9me3. Similarly, within H3K27me3 domains and regions without enrichment for H3K9me3 or H3K27me3, strucHC subtypes have higher levels of lamina association (Figure 4-4C, Supp. Fig. 4-4D).

Sensitivity to nucleases like DNase I or micrococcal nuclease is a classic measure of chromatin accessibility. We next compared our chromatin subtypes using ENCODE data for sequencing after DNase I digestion (DNase-seq) in BJ fibroblasts (Thurman et al., 2012). As expected, 10-kb windows within euchromatin domains are consistently more

DNase-sensitive than windows in strucHC or intermediate domains (Figure 4-4D, far left), further corroborating that Gradient-seq detects differences in the structural accessibility of chromatin. Meanwhile, DNase-seq does not clearly distinguish strucHC from intermediate domains, two forms of chromatin that are mostly transcriptional silent (Figure 4-2D) and yet are quantifiably different according to Gradient-seq (Figure 4-4A, far left), consistent with the specificity of DNase I for open chromatin regions (Thurman et al., 2012). Euchromatin domains are highly DNase-sensitive even where they overlapped with H3K27me3 (Figure 4-4D, middle right), but interestingly not when they overlap with H3K9me3 (Figure 4-4D, middle left), even though both types of regions are transcriptionally active (Figure 4-3B).

One recent study (Mieczkowski et al., 2016) performed MNase titrations to define the MNase accessibility (MACC) of chromatin regions, with a higher MACC score indicating sensitivity at lower concentrations of MNase. Similar to the results with DNase-seq, the MACC signal for human K562 cells (Mieczkowski et al., 2016) is elevated in euchromatin domains, both as a whole and in H3K27me3 domains, but not in euchromatin overlapping H3K9me3 (Supp. Fig. 4-4B). Although the MACC and Gradient-seq experiments were performed in different in different cell types, we confirmed these findings at regions with similar H3K9me3 or H3K27me3 patterns in both K562 cells and foreskin fibroblasts (Supp. Fig. 4-4C). Indeed, in such regions, the pattern of MNase resistance closely mirrors the pattern of strucHC versus euchromatin domains (Supp. Fig. 4-4C, left), indicating that there is widespread similarity of sonication-resistance and MNase-resistance. However, a clear exception occurs at regions bearing both H3K9me3 and H3K36me3 in both cell types (Supp. Fig. 4-4C, right), which are classified as euchromatic according to Gradient-seq but have negative MACC signals. Thus, in contrast to Gradient-

seq, nuclease-based methods are unable to distinguish transcriptionally active H3K9me3 domains from transcriptionally inactive ones, perhaps reflecting an intrinsic property of H3K9me3-marked chromatin to resist enzymatic cleavage reactions.

Finally, we analyzed replication timing within each chromatin subtype using Repli-seq data (Pope et al., 2014), in which newly replicated is sequenced in six cytometry-fractionated cell cycle phases. All forms of H3K9me3 domains, including both strucHC and euchromatin, replicate late, while H3K27me3 domains replicate in early-to-mid S phase (Figure 4-4E), consistent with previous observations (Chandra et al., 2012). Thus, in contrast to gene expression and methylation at CpG islands, late replication is a property more closely associated with the H3K9me3 histone mark than chromatin structure, and can co-occur with gene transcription.

In summary, we have used gradient-sedimentation of sonication-resistant chromatin to purify a set of transcriptional silent chromatin domains (strucHC), which includes the majority of H3K9me3 domains as well as a subset of DNase-resistant H3K27me3 or unmarked regions, but notably excludes the H3K9me3/H3K27me3 regions with transcribed genes and hypomethylated CpG islands.

The proteome of H3K9me3-marked structural heterochromatin

To identify proteins embedded in structural heterochromatin, as distinct from transcriptionally active H3K9me3 domains, we performed label-free quantitative proteomic studies on three biological replicates of the strucHC fraction, along with H3K9me3 IPs performed off of each strucHC fraction (Figure 4-5A). Results were compared to the euchromatin-containing fraction, which contains both euchromatic fragments as well as soluble proteins from the lysate, and thus for accuracy will be referred to as the “gradient

top" fraction. Proteins were decrosslinked, digested with trypsin, and analyzed by nano liquid chromatography coupled to tandem mass spectrometry (MS/MS). Protein abundance in each sample was determined using intensity-based absolute quantification (iBAQ) (Cox and Mann, 2008; Schwanhäusser et al., 2011), and protein ranks were compared between samples. In total, across the different samples, we identified 3,097 proteins with at least two peptides. Of the 1,864 proteins detected in the strucHC fraction, 217 proteins had a significantly higher rank in strucHC compared to the gradient top fraction (T-test, $P < 0.05$), and these were termed "strucHC-enriched proteins" (red dots in Figure 4-5B, left). The strucHC-enriched proteins are listed in Table 4-2. However, we cannot exclude the possibility that some non-chromatin-bound proteins may sediment to the strucHC fraction and be included among the proteins enriched in this fraction.

To specifically identify proteins that crosslink to H3K9me3-marked heterochromatin, we focused our proteomic analysis on the H3K9me3-directed IPs performed using strucHC chromatin (Figure 4-5A, henceforth "strucHC+H3K9me3"). Analysis of the DNA isolated from the same samples reveal a high enrichment for heterochromatic regions (Figure 4-1F, Supp. Fig. 4-1C). The strucHC+H3K9me3 chromatin was eluted after stringent high-salt washes, and thus has comparable specificity for chromatin-bound proteins as traditional ChIP-mass spec approaches, but inclusion of the gradient sedimentation step allows depletion of transcriptionally active H3K9me3 domains. Out of 716 proteins detected in the strucHC+H3K9me3 sample, we identified 172 "heterochromatin proteins" (Table 4-3) that were significantly enriched in strucHC+H3K9me3 compared to the gradient top fraction (Figure 4-5B, right, orange dots). As expected, there was substantial overlap between the heterochromatin proteins and strucHC-enriched proteins (Figure 4-5C). An additional 429 proteins detected in both

strucHC+H3K9me3 and the gradient top fraction, but without significantly different ranks between the two samples, and hence were classified as “shared” (Figure 4-5B, right, grey dots). Finally, among the remaining proteins, 1,474 proteins were classified as belonging the gradient top, meaning they were uniquely detected in that sample or were significantly enriched by rank in comparison to the strucHC or strucHC+H3K9me3 samples.

The 172 identified heterochromatin proteins include several proteins known to compose or associate with compacted chromatin (Figure 4-5D), including the linker histone (H1.1, H1x, and H1.0), histone variant macroH2A, and lamin B1, as well as the repressive histone modifier histone deacetylase 2. Also included among the heterochromatin proteins are NONO and SFPQ, proteins that form a co-repressor complex (Mathur et al., 2001; Dong et al., 2009) and are known to interact with the H3K9me3 mark (Vermeulen et al., 2010), and we find strongly enriched in strucHC+H3K9me3 (Figure 4-5D, bottom). Strong enrichment was also observed HNRNPK, which is required for silencing and H3K9me3 deposition at genes and repetitive elements (Bao et al., 2015; Thompson et al., 2015). Furthermore, the identified heterochromatin proteins include proteins that interact directly with HP1 and have been linked to gene repression, such as HP1BP3 , THRAP3, and BCLAF1 (Vermeulen et al., 2010). Meanwhile, HPγ itself was in the category shared between heterochromatin and euchromatin, consistent with our genomic data showing that HP1-associated ZNF domains (Vogel et al., 2006) are enriched in euchromatin.

To our knowledge, only one previous study used proteomics to comprehensively identify proteins present in H3K9me3 chromatin in differentiated mammalian cells (Soldi and Bonaldi, 2013). The results of this study, using H3K9me3-directed ChIP-MS in HeLa cells, agree strongly with our own: 60 (37%) of the 172 heterochromatin proteins identified

by our approach were previously found to be enriched by H3K9me3 ChIP (Soldi and Bonaldi, 2013) (Figure 4-5E, left orange bar), an enrichment of 27-fold compared to the gradient top proteins (Figure 4-5E, blue). Moreover, 124 out of the 208 (60%) proteins discovered by H3K9me3 ChIP-MS (Soldi and Bonaldi, 2013) fall into the heterochromatin or shared categories from our study. Meanwhile, other studies have used proteomics to characterize specific subsets of H3K9me3-related proteins, such as “readers” that interact with a H3K9me3 bait (Vermeulen et al., 2010) or proteins associated with pericentromeric satellites in mouse ES cells (Saksouk et al., 2014). The heterochromatin protein group also has a significantly higher rate of overlap with both of these studies compared to gradient top proteins (Figure 4-5E, middle panels), though as expected fewer proteins fall into these specific categories. The strucHC-enriched proteins similarly have an elevated rate of overlap with all three studies (Figure 4-5E, red bars). Similar to previous studies of constitutive heterochromatin (Saksouk et al., 2014), we find nucleolar proteins among the heterochromatin proteins, including nucleolin (NCL), BRIX1, NOP2/NOP56/NOP58, Pescadillo homolog (PES1), and GTPBP4. This is consistent with role of H3K9me2/3 in repressing rDNA arrays (Santoro et al., 2002) and the physical association of heterochromatin at the nucleolar periphery (Politz et al., 2013).

Remarkably, despite the low levels of gene transcription within strucHC domains (Figure 4-2D), 119 of the 172 heterochromatin proteins have RNA-binding activity according to a published database (Gerstberger et al., 2014), including known co-repressors HNRNPK, NONO, and SFPQ (Mathur et al., 2001; Dong et al., 2009; Thompson et al., 2015). RNA binding by HP1 has long been known to be important for its localization to heterochromatin (Muchardt et al., 2002), but RNA binding has not been appreciated as a common feature among mammalian heterochromatin proteins. Yet the

RNA-binding proteins we find in heterochromatin, compared to the total heterochromatin proteins, have an even higher rate of agreement of previously published sets of H3K9me3 ChIP-MS proteins or H3K9me3 readers (Figure 4-5F). Thus, the association of such proteins with H3K9me3-marked chromatin has been demonstrated through multiple independent approaches, and our use of gradient sedimentation ensures that such proteins are indeed associated with the heterochromatic, transcriptionally repressed portion of H3K9me3 domains.

A major functional property of the H3K9me3 form of heterochromatin is impeding reprogramming to pluripotency (Onder et al., 2012; Soufi et al., 2012; Matoba et al., 2014). Using data from a genome-wide screen (Toh et al., 2016), we find that heterochromatin proteins are enriched for factors that are repressors of iPS reprogramming (siRNAs that increase iPS conversion) and depleted for factors that enhance reprogramming (Supp. Fig. 4-5A). To our knowledge, across several studies, 22 of the 172 heterochromatin proteins have been shown to impede iPS reprogramming (Supp. Fig. 4-5B), several of which are RNA binding proteins, including HNRNPK (Thompson et al., 2015) and muscleblind-like protein 1 (MNBL1) (Han et al., 2013).

Surprisingly, we noted that the heterochromatin proteins included six proteins implicated in the neurodegenerative disease amyotrophic lateral sclerosis (ALS), with strong quantitative enrichment in strucHC+H3K9me3 (Figure 4-5G). Based on the total number of proteins known to be recurrently mutated in ALS (Cirulli et al., 2015), this rate of overlap with heterochromatin and strucHC-enriched proteins is significantly higher than expected (Figure 4-5E, far right). Among such proteins, TARDBP (also called TDP-43) has been well-studied as a biochemical component of the pathological inclusions found in ALS neurons and a genetic cause of certain forms of familial ALS (Lee et al., 2012).

TARDBP has been shown to oppose iPS reprogramming in two independent genome-wide screens (Qin et al., 2014; Toh et al., 2016) (Supp. Fig. 4-5B), suggesting that it might be important for heterochromatic silencing. Using transcriptome data from fibroblasts isolated from ALS patients (Highley et al., 2014), we find that germline mutation of TDP-43 is associated with a modest but widespread upregulation of genes in strucHC, compared to healthy controls (Supp. Fig. 4-5C,D,E). Such upregulation is absent in sporadic ALS cases, where the fibroblast copies of TDP-43 are normal. These findings help explain overexpression of heterochromatic repeat elements and the altered organization of compacted chromatin in ALS disease models (Li et al., 2012; Amlie-Wolf et al., 2015). Thus, identification of proteins crosslinked to chromatin in the strucHC fraction has revealed new proteins with functional roles in human heterochromatin.

Activation of liver-specific genes is impeded by H3K9me3-marked strucHC and associated proteins RBMX/RBMXL1

We next investigated whether strucHC and its bound proteins, in addition to impeding reprogramming to pluripotency, oppose conversion to the hepatic lineage. Indeed, in human fibroblasts, liver genes in H3K9me3 domains in strucHC or intermediate forms, but not liver genes in H3K9me3 domains in euchromatin, have a profound failure to activate during cell conversion to hiHep cells (Figure 4-6A, median indicated by white circles). These include genes for important for hepatic cell fate, metabolism, epithelialization, and secretion into the blood plasma (Figure 4-6B; genes listed in Table 4-4). Compared to fibroblasts, these genes show a diminution of H3K9me3 in human liver (Figure 4-6C). While genes in H3K27me3 in fibroblasts allowed greater levels of hepatic gene induction, the strucHC form of H3K27me3 was slightly more refractory than the intermediate or

euchromatic forms of this mark (Figure 4-6A). We conclude that H3K9me3-marked chromatin, of either the strucHC or intermediate form, is most refractory to gene activation by hepatic factors.

We reasoned that disruption of heterochromatin might render this class of H3K9me3-marked genes more competent for activation by hepatic transcription factors (TFs). We used a cocktail of three factors previously shown to produce hiHep cells (FOXA3, HNF1A, and HNF4A) (Huang et al., 2014) and confirmed that expression of these factors was sufficient to convert foreskin fibroblasts to hepatocyte-like cells expressing liver markers (Supp. Fig. 4-6A). We found that two cycles of siRNA against SUV39H1, which reduced H3K9me3 levels (Supp. Fig. 4-6B), greatly enhanced the activation of a panel of hepatic genes marked by H3K9me3 in fibroblasts (Figure 4-6D). SUV39H1 knockdown alone, in the absence of hepatic TFs, did not result in detectable expression by RT-qPCR (Figure 4-6D).

The protein RNA Binding Motif Protein, X-linked (RBMX) had the second highest enrichment in strucHC+H3K9me3 among heterochromatin proteins (Figure 4-5D, bottom, red dot), and the related protein RBMXL1 was enriched in the strucHC fraction and detected in one strucHC+H3K9me3 replicate. RBMX is known to bind chromatin independent of its RNA recognition motif (Matsunaga et al., 2012), and its depletion causes defects in pericentromeric cohesion (Matsunaga et al., 2012), a phenotype reminiscent of SUV39H1 mutants (Peters et al., 2001). However, RBMX is not known to function in gene silencing or heterochromatin. We find that siRNAs co-targeting RBMX and RBMXL1 (RBMX/L1) (Supp. Fig. 4-6C,D) enable the activation of liver-specific genes in strucHC by hepatic TFs, comparable to SUV39H1 knockdown (Figure 4-6D). The

enhanced expression persisted in hiHep cells 14 days after induction (Supp. Fig. 4-6E), demonstrating gene expression stability substantially later than the RBMX knockdown.

Regulation of diverse genes in strucHC by RBMX/L1

To understand the role of RBMX/L1 in impeding hepatic TF activity genome-wide, we performed mRNA-seq in fibroblasts treated with RBMX/L1 siRNA, compared to control siRNA and siRNA against SUV39H1, after 7 days of hepatic TF expression (Figure 4-7A). Unsupervised clustering of the mRNA-seq data by Euclidean distance confirmed the similarity of biological replicates and showed that the RBMX/L1 siRNA treatment clustered with the SUV39H1 knockdown (Supp. Fig. 4-7A, green box). Out of 1333 genes upregulated by RBMX/L1 knockdown (compared to non-targeting siRNA) in the presence of hepatic TFs, 65% were also upregulated by SUV39H1 ($P < 10^{-200}$) (Figure 4-7B).

Within strucHC, the SUV39H1 knockdown condition revealed 281 genes that are responsive to heterochromatin disruption and can be activated by either endogenous factors or the hepatic TF cocktail. Among these 281 genes, the effect of SUV39H1 siRNA was well correlated with the effect of RBMX/L1 siRNA (Figure 4-7C, left), with upregulation by RBMX/L1 siRNA reaching statistical significance for 102 of these genes. Genes upregulated by both SUV39H1 and RBMX/L1 siRNA include the liver-specific genes encoding C-Reactive Protein (CRP) and the nuclear receptor NR1H4/FXR (Figure 4-7D). We note that certain strucHC genes are uniquely regulated by either SUV39H1 or RBMX/L1 (Figure 4-7C). A list of strucHC genes upregulated by each siRNA is provided in Table 4-5.

We also performed mRNA-seq in cells treated with SUV39H1, RBMX/L1, and control siRNA in the absence of ectopic transcription factors (Figure 4-7A, Supp. Fig. 4-

7). Fewer genes in strucHC were upregulated in this condition (109 after SUV39H1 knockdown and 67 for RBMX/L1 knockdown, Supp. Fig. 4-7C), consistent with our qPCR studies (Figure 4-6D) and prior studies on SUV39H1 (Bulut-Karslioglu et al., 2014). Importantly, in the absence of ectopic TFs, knockdown of RBMX/L1 was sufficient for the activation of genes within strucHC and H3K9me3 domains, such as the TMEM178 (Figure 4-7E) and the imprinted gene Paternally Expressed 3 (PEG3). Thus, the role of RBMX/L1 in impeding gene activation is not dependent upon ectopic TF expression. Taken together, our studies with RBMX/L1 validate the proteomic study of the strucHC gradient fraction as a resource for discovering novel regulators of genes within heterochromatin.

4.5 DISCUSSION

It is becoming clear that silent chromatin adopts diverse structural and functional forms, and that understanding the mechanisms of gene silencing provides insight that is crucial for being able to reprogram cells by modifying silent chromatin. Genome mapping of histone modifications such as H3K9me3 and H3K27me3 has been integral to the study of lineage-specific gene repression (Ezhkova et al., 2009; Hawkins et al., 2010; Xu et al., 2014a), but these marks are also found in transcriptionally active or accessible chromatin (Vakoc et al., 2005; Blahnik et al., 2011; Riddle et al., 2012; Breiling et al., 2001). Here we show that the recovery of sonication-resistance chromatin fragments using sucrose gradients is an effective method of heterochromatin purification, with efficiency comparable of conventional ChIP. This approach allows us, for the first time, to distinguish heterochromatic forms of H3K9me3 and H3K27me3 domains from those that are competent for transcription, and to comprehensively identify proteins that are associated

with definitive regions of heterochromatin. Using this dataset, we show that particular heterochromatin proteins are required to prevent activation of cell type-inappropriate genes, either under homeostatic conditions or in the presence of ectopic transcription factors. We also provide diverse resources for future studies on heterochromatin subtypes and the genes and repeat elements they contain.

Our use of sucrose gradients was inspired by previous studies applying gradient sedimentation to MNase I-digested or sonicated chromatin (Gilbert et al., 2004; Ishihara et al., 2010), but these approaches focused on fragments that changed buoyancy without changing size or resistance to fragmentation. Our method, in addition to its enrichment of heterochromatic regions, has the advantage of avoiding an electrophoresis-based size-selection step, thereby enabling subsequent IP or analysis of protein content.

The literature has been conflicted on whether heterochromatin includes H3K27me3-marked regions or applies solely to H3K9me3-marked and HP1-bound regions (Trojer and Reinberg, 2007; Beisel and Paro, 2011). We find that many H3K27me3 domains are enriched in the sonication-resistant strucHC fraction comparable to heterochromatic DBRs (Figure 4-1F, green arrows versus purple arrows), with 49% of the H3K27me3 domains overlapping with strucHC. Meanwhile, 29% of H3K27me3 domains are enriched in the euchromatic gradient fraction, correlating with enhanced permissiveness to transcription. Thus, H3K27me3-marked chromatin spans a wide range of physical states, suggesting that this histone mark alone is insufficient to predict or exclude heterochromatin status. This also highlights a potential danger of generalizing from model loci to all sites bearing the same histone mark, as we find that all four HOX clusters, which have been pivotal to the study of Polycomb biology (Margueron and Reinberg, 2011), reside in euchromatin domains.

Compared to H3K27me3, H3K9me3 domains have a much stronger tendency to be heterochromatic, with only 3.2% being significantly enriched in the euchromatic fraction. Although the euchromatic subtype represents a small fraction of the total H3K9me3 domains, it contains genes that have been used to normalize H3K9me3 ChIP-seq data (Zhu et al., 2013) or to study the HUSH complex, KAP1, and SETDB1 (Tchasovnikarova et al., 2015; Groner et al., 2010), without recognition that these sites are euchromatic and atypical for H3K9me3. Although most genes in euchromatic H3K9me3 domains are expressed (Figure 4-3B) and are more readily activated by ectopic transcription factors (Figure 4-6B), H3K9me3 has a repressive role at these sites in dimming gene expression (Tchasovnikarova et al., 2015; Groner et al., 2010). As active genes marked by H3K9me3 are still bound by multiple isoforms of HP1 (Vogel et al., 2006), our data allow new studies on which chromatin components distinguish the heterochromatic and euchromatic forms of H3K9me3 domains.

We find that genes marked by H3K9me3, if in strucHC or intermediate domains, are markedly deficient in activation by hepatic transcription factors during hiHep reprogramming, with the majority of hepatocyte genes persisting at near-fibroblast levels (Figure 4-6A). A comparable defect was not observed for the heterochromatic form of H3K27me3 domains, demonstrating a significant distinction between H3K9me3- and H3K27me3-marked chromatin. This finding shows, for the first time, that H3K9me3-marked chromatin impedes direct conversion between two mature lineages. This agrees with prior work showing that such chromatin impedes binding by iPS reprogramming factors and reprogramming itself (Soufi et al., 2012; Onder et al., 2012), and it suggests that diverse other forms of cell reprogramming might similarly be limited by the extent to which strucHC domains can be accessed.

The ability to purify heterochromatin physically provided the opportunity to uncover the protein composition of repressed chromatin domains, without contaminating euchromatin or bias for particular histone modifications. The proteins we find to be enriched in H3K9me3-marked heterochromatin include many proteins known to be involved in gene repression or associated with H3K9me3 (Figure 4-5D). Beyond specific examples corroborated by the literature, we show that this set of heterochromatin proteins as a whole is enriched for functional features of heterochromatin, such as impeding reprogramming to pluripotency (Supp. Fig. 4-5A,B). Using this dataset, we were able to discover that two related proteins, RBMX and RBMX/L1, although not previously implicated in transcriptional repression, impede gene activation, with knockdown phenotypes comparable in magnitude to a classic heterochromatin regulator (Figure 4-6D). We find that RBMX/L1 and SUV39H1 inhibit TF-mediated gene induction at many sites where they are otherwise dispensable for silencing, suggesting that gene induction assays using alternative-lineage TFs may be important to test the functionality of heterochromatin proteins.

Heterochromatin regulation is instrumental to cell fate control in development and iPS reprogramming. Our findings support a role for heterochromatin in impeding conversion between mature lineages, they define subtypes of heterochromatin that contribute differentially to this barrier, and they reveal a large resource of proteins that can be investigated in order to overcome it. Understanding the general and specific aspects of heterochromatin, starting with the datasets provided here, could lead to the selective opening of silent chromatin domains and more efficient cellular reprogramming to new, fully differentiated types of cells.

4.6 MAIN FIGURES

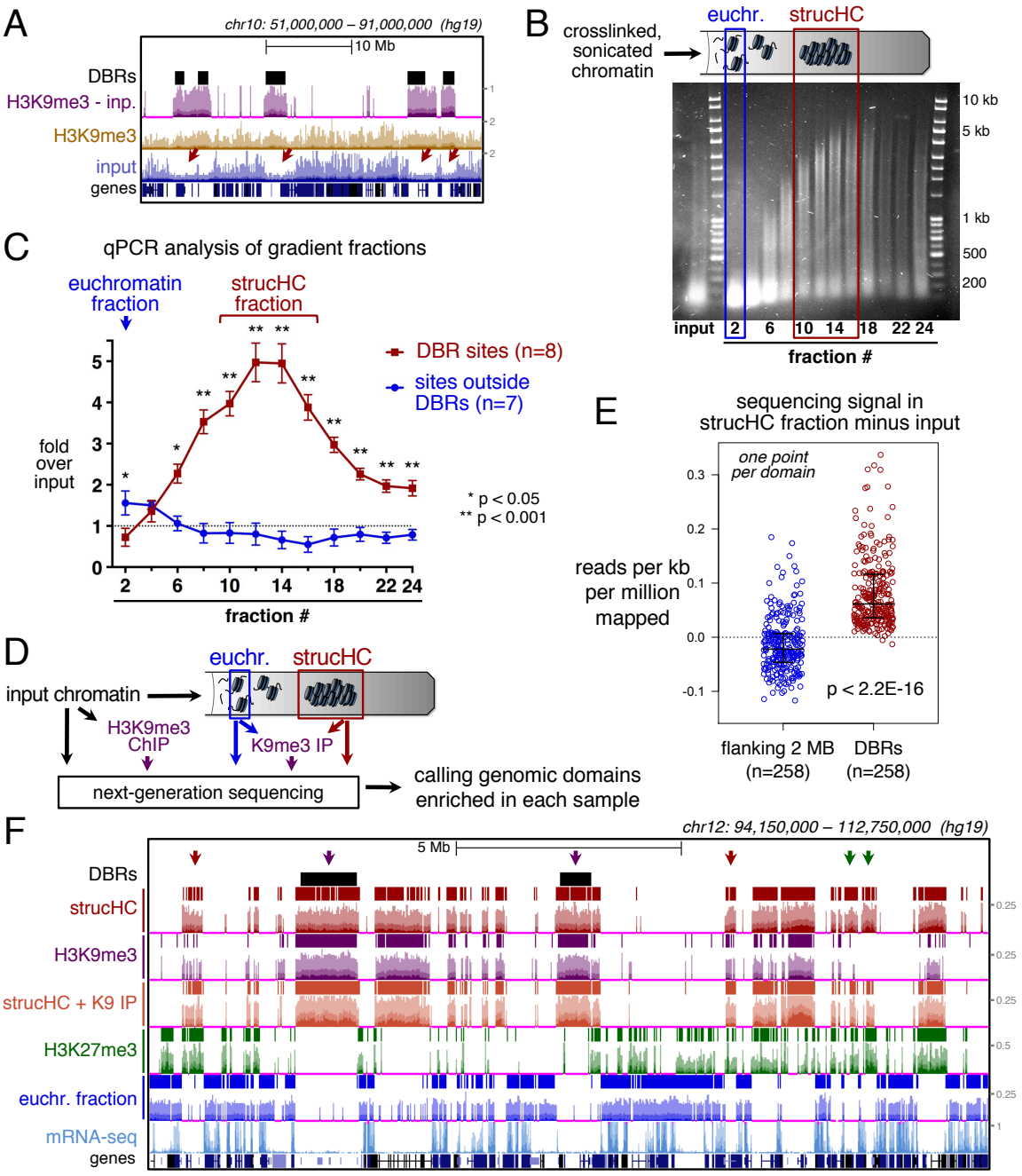


Figure 4-1. Heterochromatic regions can be enriched and mapped genomically by isolating sonication-resistant chromatin.

(A) The positions of DBRs (black bars; Soufi et al., 2012) compared to H3K9m3 ChIP-seq for IMR90 fibroblasts (Roadmap Epigenomics consortium, GSE16256) with and without normalization to input. Red arrows indicate regions of depletion in the input track alone.

(B) Crosslinked, sonicated chromatin from human BJ fibroblasts was separated on a sucrose gradient. Agarose gel shows the size ranges of DNA purified from each fraction. Boxes indicate the fractions used as the euchromatin and structural heterochromatin (strucHC) samples in subsequent analyses.

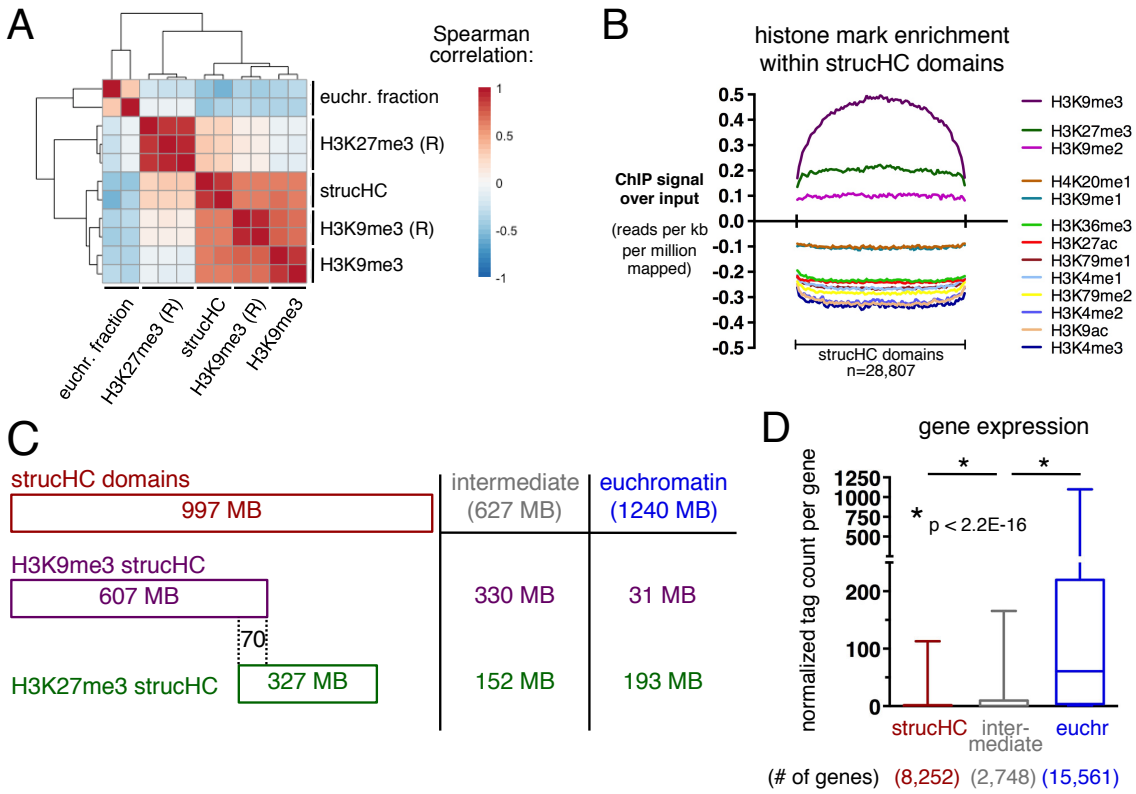
(C) qPCR on DNA isolated from sucrose gradient fractions from BJ fibroblasts. Based on prior work (Soufi et al., 2012), all qPCR sites (red and blue) contain binding sites for OCT4/SOX2 in ES cells, but red sites fall inside DBR domains and lack OCT4/SOX2 binding in fibroblasts. An equal DNA mass was loaded per fraction, and results were averaged from two biological replicates. Error bars show SEM among tested qPCR sites.

(D) Plan for ChIP and gradient-purified samples used for DNA sequencing.

(E) Sequencing signal per domain in strucHC gradient fraction minus input. For each point, signal is averaged over an entire domain, which is either a DBR (red) or the flanking 2 MB on both sides of a DBR (blue). Black lines show the median and interquartile ranges. P value by Wilcoxon rank sum test.

(F) Browser view of 18 MB segment of chr12 showing physical mapping of chromatin structure by Gradient-seq in foreskin fibroblasts compared to ChIP-seq signals for repressive histone marks. Horizontal, colored bars above sequencing tracks indicate enriched domains called by a sliding window algorithm. The pattern in the strucHC fraction has broad similarity to H3K9me3 domains, which include DBRs (purple arrows), and anti-correlates with the signal in the Euchromatin fraction and gene expression by mRNA-seq.

However, some strucHC domains contain chromatin co-marked by H3K9me3 and H3K27me3 (red arrows), or H3K27me3 alone (green arrows). “strucHC + K9 IP” indicates the H3K9me3 IP performed using the strucHC fraction chromatin. H3K27me3 ChIP data was obtained from the Epigenomics Roadmap consortium.



	# of genes	enrichment	FDR
intrinsic to membrane	1,984	1.3	$<10^{-54}$
neurological system process	572	1.8	$<10^{-55}$
G-protein coupled receptor signaling	534	1.8	$<10^{-52}$
calcium ion binding	378	1.6	$<10^{-19}$
ion transport	322	1.6	$<10^{-18}$
cell adhesion	271	1.5	$<10^{-9}$
defense response	265	1.6	$<10^{-16}$
immune response	253	1.4	$<10^{-6}$
olfactory receptor activity	247	2.2	$<10^{-38}$
secretion	126	1.6	$<10^{-6}$
ectoderm development	95	1.8	$<10^{-7}$
sugar binding	92	1.8	$<10^{-7}$
neurotransmitter binding	59	2.2	$<10^{-7}$
digestion	48	2.0	$<10^{-4}$
oxidoreductase activity	23	2.9	$<10^{-5}$

Figure 4-2. The genomic landscape of structural heterochromatin contains distinct histone modification signatures and cell type-inappropriate genes.

(A) Spearman correlation analysis and unsupervised clustering of ChIP- and Gradient-seq datasets. Samples were compared using their input-normalized tag density binned into 10

kb sliding windows. Data from the Epigenomics Roadmap is labeled with "(R)"; all other samples were sequenced in this study.

(B) Histone mark levels, normalized for input, over the width of strucHC domains, using Roadmap Epigenomics data. The plots show an average of all 28,807 strucHC domains, weighted by domain length. An additional 14 acetyl marks depleted in strucHC are not shown.

(C) Schematic showing the distribution of H3K9me3 and H3K27me3 domains across the three structural categories defined by Gradient-seq. In the strucHC category, domains are drawn to scale, to indicate the number of megabases (MB) of overlap.

(D) Expression levels of genes within the three structural categories of chromatin defined by Gradient-seq. Boxplots show mRNA-seq tag counts per gene, normalized by DESeq2 and divided by gene length (whiskers: 5^{th} and 95^{th} percentiles). P values by Wilcoxon rank sum test.

(E) Gene Ontology (GO) analysis of Refseq genes inside strucHC domains. GO terms were manually filtered to remove redundant categories. FDR was computed by Benjamini-Hochberg. Shown are the top 15 most significant terms, sorted by gene count.

(F) Rates of methylation per CpG were determined by whole-genome bisulfite sequencing (Roadmap Epigenomics consortium data for foreskin fibroblasts, GSM1127120). The distribution of methylation rates was plotted for each category (whiskers: 10^{th} and 90^{th} percentiles). P values by Wilcoxon rank sum test.

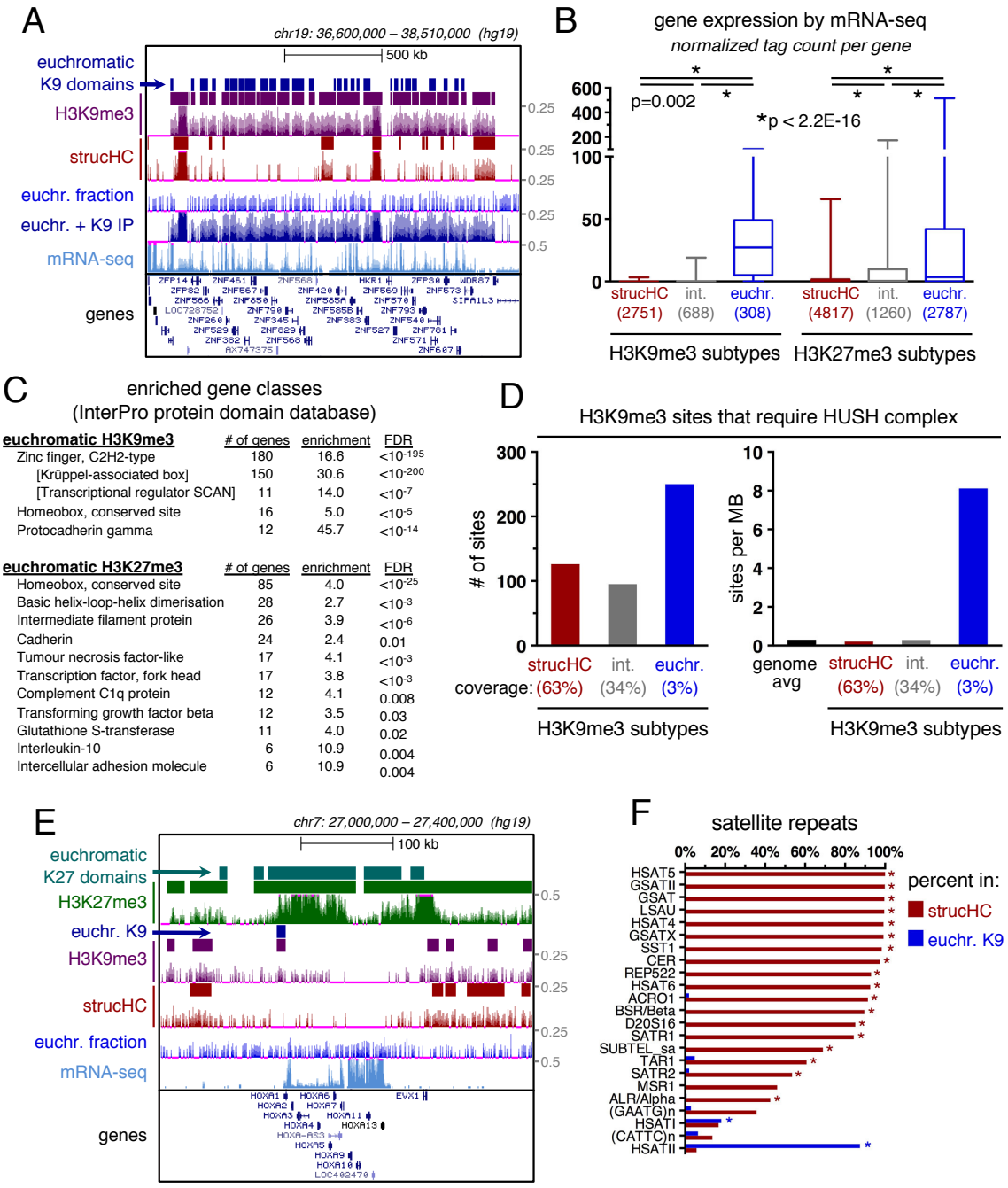


Figure 4-3. A subset of H3K9me3 and H3K27me3 domains are structurally euchromatic and permissive to transcription.

(A) Browser view of euchromatic H3K9me3 domains over ZNF gene family cluster, which is depleted from strucHC. Note the mRNA-seq signal indicating transcription of these

genes. The “euchr. + K9 IP” track shows the H3K9me3 ChIP performed off of the euchromatic (“Euchr.”) fraction of the gradient.

(B) Genes in euchromatic H3K9me3 and H3K27me3 domains are largely expressed, in contrast to the strucHC or intermediate (“int.”) subtypes, which contain mostly silent genes. Boxplots show mRNA-seq tag counts per gene, normalized by DESeq2 and divided by gene length (whiskers: 5^{th} and 95^{th} percentiles). The number of genes per category is shown in parentheses. P values by Wilcoxon rank sum test.

(C) Gene annotation for euchromatic subtypes of H3K9me3 and H3K27me3 domains, using the InterPro database of protein families. FDR was computed by Benjamini-Hochberg. Redundant terms were omitted.

(D) Analysis of genomic sites reported to lose H3K9me3 after depletion of the HUSH complex/SETDB1 (Tchasovnikarova et al., 2015). Shown are the number of such sites (left) and the frequency of sites per MB (right) per chromatin category.

(E) Browser view of euchromatic H3K27me3 (“K27”) domain over HOX-A gene cluster.

(F) Fraction of satellite repeat types overlapping with strucHC versus euchromatic H3K9me3 domains. Asterisks indicate significantly enrichment (FDR < 0.05) in strucHC (red) or euchromatic H3K9me3 (blue) based on 1000 simulations of randomly shuffled domains.

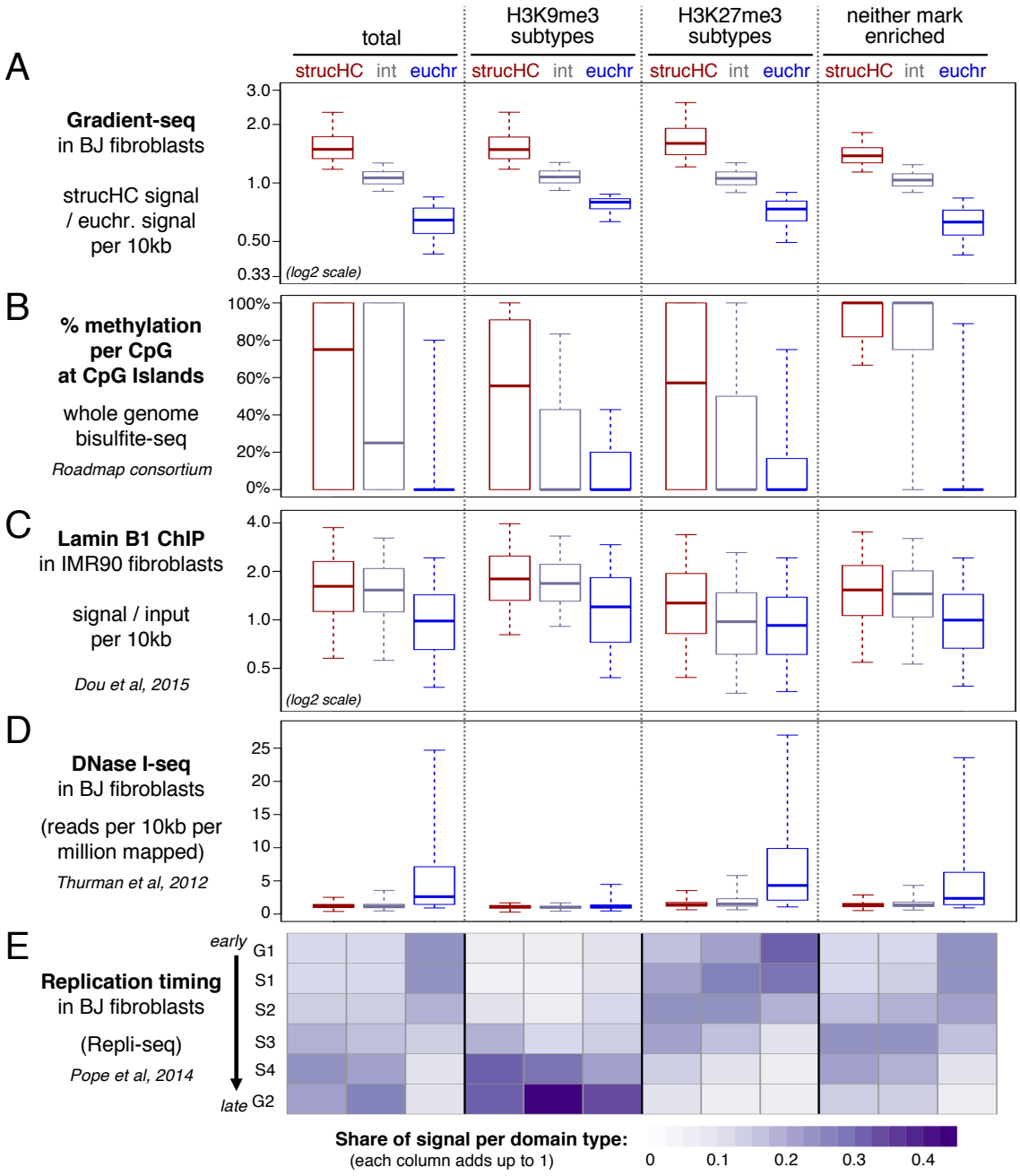


Figure 4-4. Relationship between structural subtypes of histone mark domains and hallmark properties of heterochromatin.

(A) Genomic windows (10-kb size, 500-bp sliding step) were classified by their association with strucHC, intermediate ("int"), or euchromatin ("euchr") domains, and also by their

presence inside or outside H3K9me3 and H3K27me3 domains. Windows were then scored for their sequencing signal in the strucHC fraction divided by the euchromatin fraction, and the distribution of scores per category is plotted (whiskers: 5^{th} and 95^{th} percentiles).

(B) The frequency of methylation per CpG was determined by whole-genome bisulfite sequencing (Roadmap consortium data for foreskin fibroblasts, GSM1127120). For CpGs falling within CpG islands, the distribution of methylation frequencies was plotted for each chromatin category (whiskers: 10^{th} and 90^{th} percentiles).

(C) As in Panel A, except that 10-kb windows are scored based on the IMR90 lamin B1 ChIP-seq signal (reads per million sequenced) divided by the corresponding input signal (Dou et al., 2015).

(D) As in Panel A, except that 10-kb windows are scored based the number of DNase-seq reads per million reads mapped, averaged across two replicates, using ENCODE data for BJ fibroblasts (Thurman et al., 2012).

(E) BJ fibroblast Repli-seq reads (Pope et al., 2014) for each cell cycle phase were intersected with each domain type, and the number of intersecting reads were normalized for sequencing depth and expressed as a fraction of the total signal for that domain type, so that each column sums to 1.

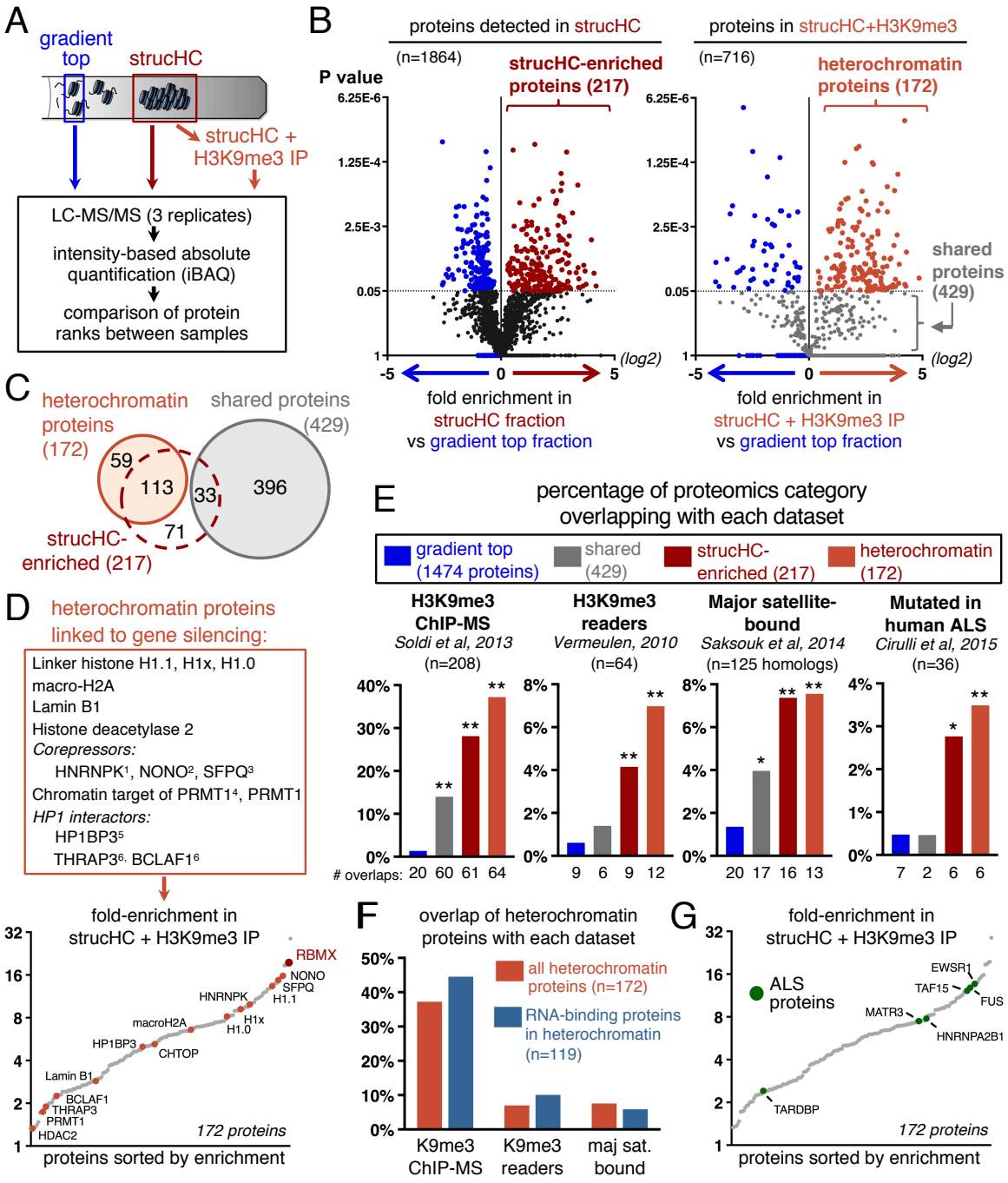


Figure 4-5. Proteomic analysis of purified H3K9me3-marked heterochromatin.

(A) Summary of quantitative proteomics study of 3 purified fractions.

(B) Proteins were scored by their relative rank between samples (x axis) and the significance of the difference between samples (T-test, y-axis). Proteins significantly

enriched in the strucHC + H3K9me3 IP compared to gradient top (orange dots) were defined as heterochromatin proteins. Proteins detected in strucHC + H3K9me3 IP, but without significant difference from gradient top (grey dots) were defined as shared proteins.

(C) Overlap of strucHC-enriched proteins with two groups of proteins (heterochromatin and shared) that were detected in the strucHC + H3K9me3 IP.

(D) Top: List of a selection of heterochromatin proteins that were previously shown to contribute to heterochromatin and/or gene repression. Bottom: the 172 heterochromatin proteins were sorted and plotted by their fold-enrichment in the strucHC + H3K9me3 (relative to gradient top), with the selected proteins above indicated by orange dots. RBMX, a focus of subsequent studies, is indicated in red. References: [1]Thompson et al., 2015; [2]Dong et al., 2009; [3]Mathur et al., 2001; [4]van Dijk et al., 2010; [5]Hayashihara et al., 2010; [6]Vermeulen et al., 2010.

(E) Percentage of each proteomic category (eg, heterochromatin proteins) that is found in each published dataset. The raw number of proteins found in common is listed below the bars. Significance were computed relative to the background overlap for the total set of 3,097 MS-detected proteins: *$p<0.05$, **$p<0.001$.

(F) Percentage of proteins that overlap with each dataset, as in Panel E, plotted for the total set of 172 heterochromatins and for the heterochromatin proteins that have RNA-binding activity according to (Gerstberger et al., 2014).

(G) As in Panel D (bottom), but proteins recurrently mutated in ALS (Cirulli et al., 2015) are indicated in green.

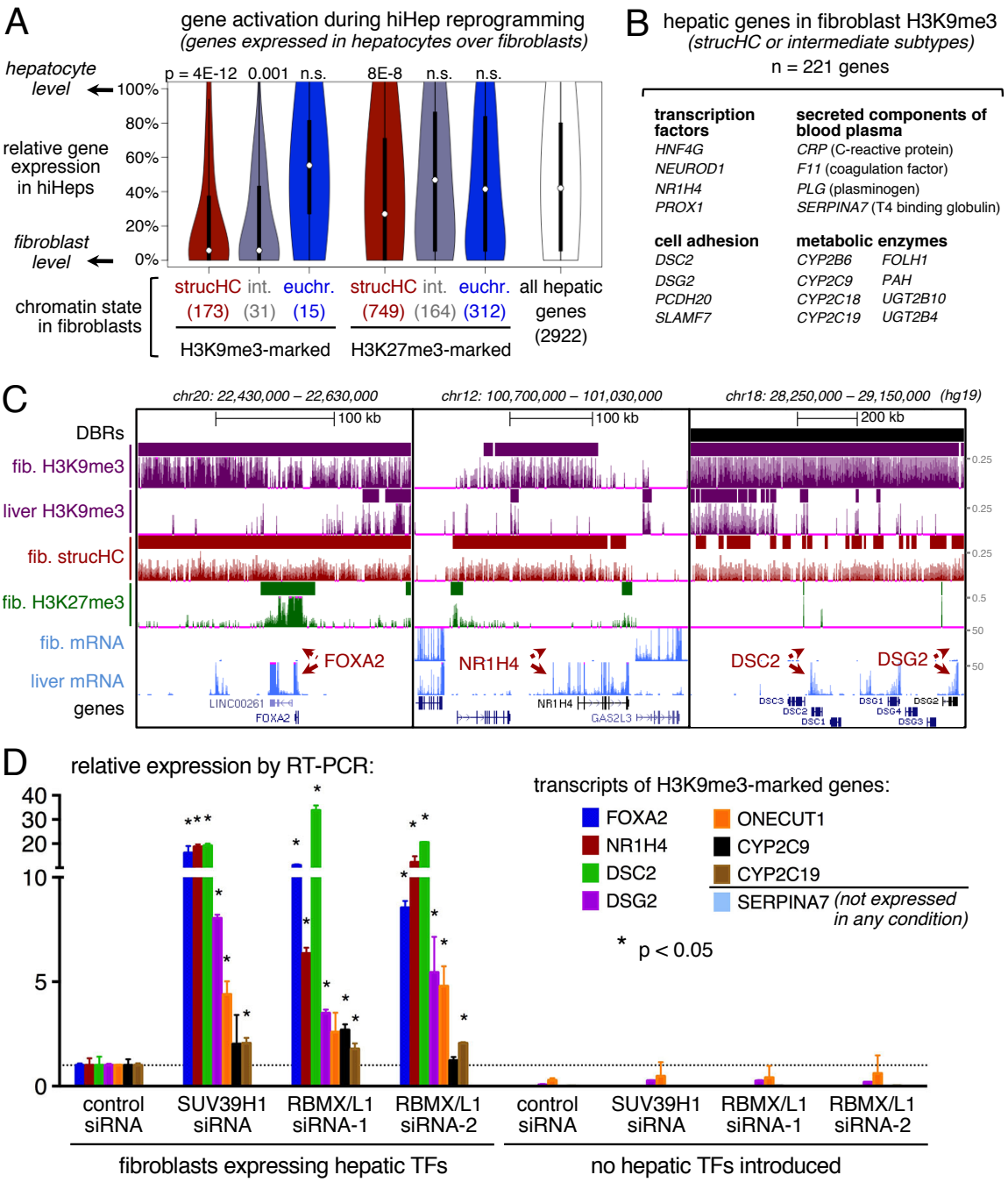


Figure 4-6. H3K9me3 domains and heterochromatin proteins impede the activation of hepatic genes in fibroblasts.

(A) Violin plots showing the levels of gene activation during hiHep reprogramming. Microarray data (Huang et al., 2014) was curated for genes expressed in cultured

hepatocytes over fibroblasts. Expression of these genes in hiHep cells was plotted on a relative scale ranging from fibroblast levels (0%) to hepatocyte levels (100%), using log2-transformed values. Median values indicated by white circles. P values calculated versus all hepatic genes by Wilcoxon rank sum test.

(B) Examples of hepatic genes in fibroblast H3K9me3 domains (strucHC or intermediate subtypes) that were analyzed in part (A), grouped according to protein function.

(C) Browser views comparing fibroblast ("fib.") and liver chromatin state and expression levels at hepatic genes. For both tissues, ChIP-seq data for H3K9me3 and H3K27me3 and mRNA-seq data were obtained from the Epigenomics Roadmap (Bernstein et al., 2010). Red arrows indicate liver-specific mRNA-seq signal for the labeled genes.

(D) RT-PCR for hepatic genes marked by H3K9me3 in fibroblasts, after two cycles of siRNA transfection. Knockdowns were performed in fibroblasts expressing the hepatic TFs FOXA3, HNF1A, and HNF4A (left) and in fibroblasts without ectopic TF expression (right). Error bars show the SEM of two biological replicates. P values by Student's T test.

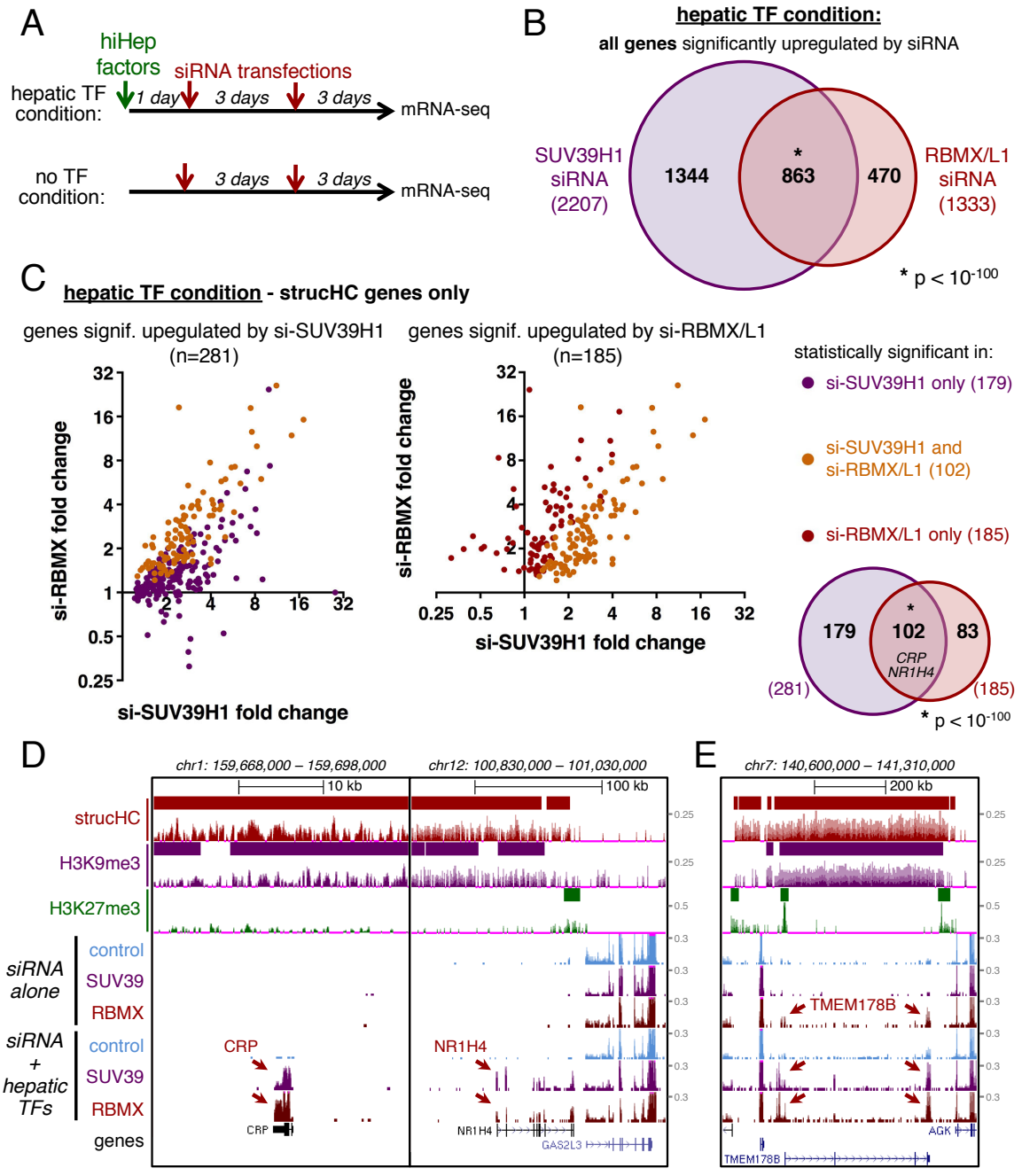


Figure 4-7. RBMX is required to suppress transcriptional activation at diverse genes in strucHC domains.

(A) Diagram of mRNA-seq experiment, performed under two conditions: with and without expression of the hepatic factors FOXA3, HNF1A, and HNF4A for hiHep reprogramming.

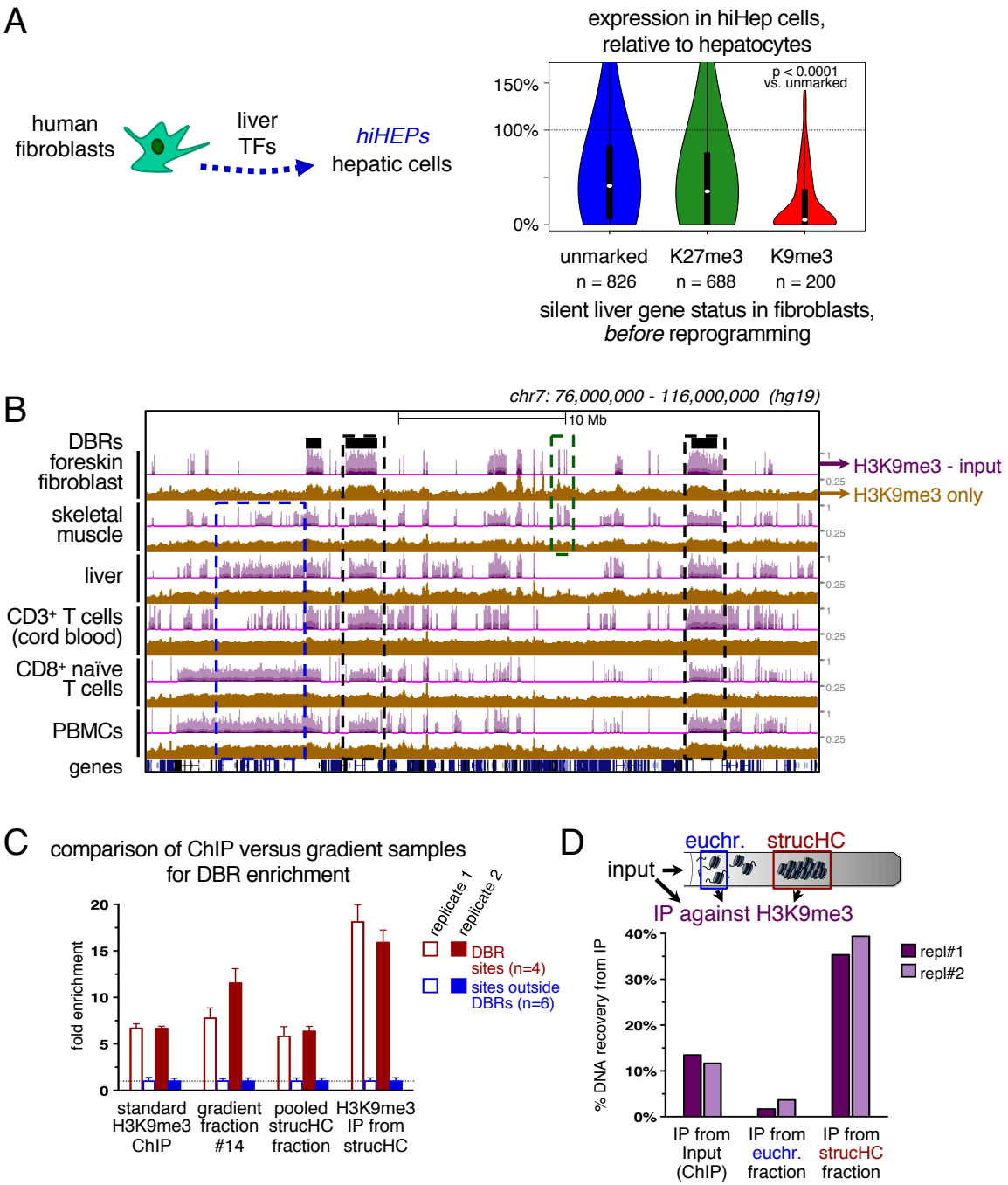
(B) Genes significantly upregulated by SUV39H1 and RBMX/L1 siRNA, compared to control siRNA, in the presence of hiHep factors. Genes downregulated by the hiHep factors in the control siRNA condition were removed from consideration. P-value indicates the significance of the overlap between gene sets.

(C) Comparison of fold-changes induced by SUV39H1 siRNA versus RBMX/L1 siRNA, in the presence of hihep factors, at genes inside strucHC domains. Genes are color-coded based on whether they were upregulated with statistical significance ($P_{adj} < 0.05$) in one sample, the other, or both. The extent of overlap of statistically upregulated strucHC genes is shown in the Venn diagram.

(D) Browser views of genes in strucHC that are upregulated by si-SUV39H1 and si-RBMX/L1 in the presence of hepatic factors. Note the mRNA-seq signal when either siRNA is added in the presence of factors (red arrows), but not in control siRNA samples or in the absence of factors. H3K27me3 data was obtained from the Epigenomics Roadmap; all other samples were sequenced in this study.

(E) Browser view of gene in strucHC that is upregulated by si-RBMX/L1 both in the presence and absence of hepatic factors.

4.7 SUPPLEMENTARY FIGURES



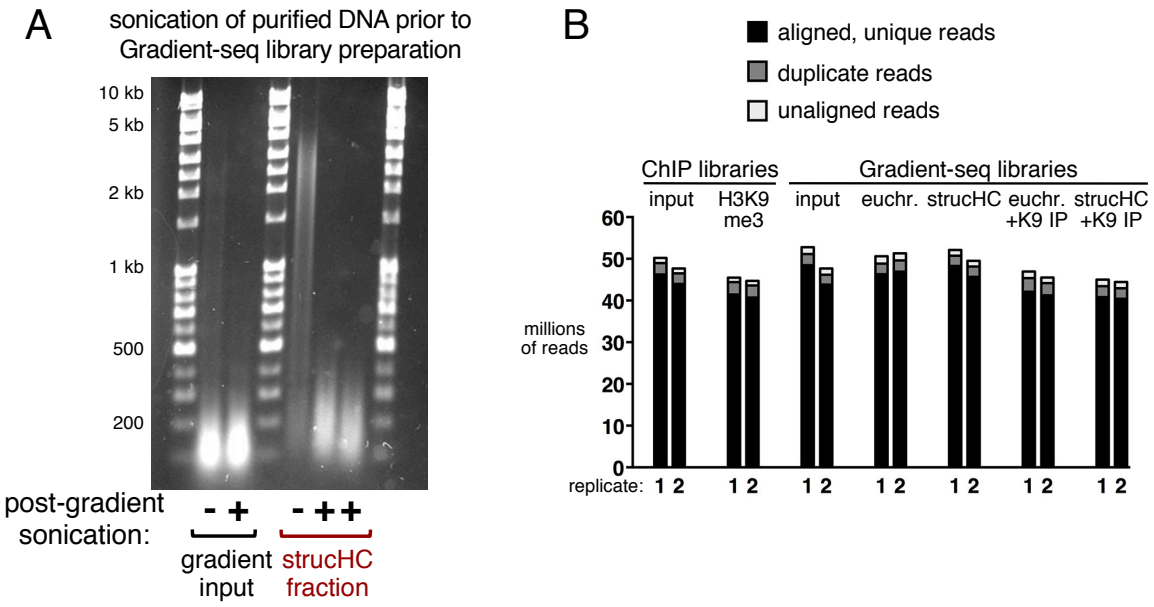
Supplementary Figure 4-1. H3K9me3-marked heterochromatin impedes hiHep reprogramming, resists sonication, and is enriched by gradient sedimentation.

(A) Violin plots showing the levels of gene activation during hiHep reprogramming. Microarray data (Huang et al., 2014) was curated for genes silent in fibroblasts and expressed in cultured hepatocytes. Genes were then classified by histone mark in fibroblasts (starting cell type) using Roadmap Epigenomics ChIP-seq data. Expression of these genes in hiHep cells was plotted on a relative scale ranging from fibroblast levels (0%) to hepatocyte levels (100%), using log2-transformed values. Median values indicated by white circles. P values calculated by Wilcoxon rank sum test.

(B) H3K9me3 ChIP-seq profiles for various human cell and tissue types, comparing input-normalized signal (purple) to unnormalized signal (yellow). Data obtained from the Epigenomics Roadmap consortium (Bernstein et al, 2010). Some domains of H3K9me3 enrichment over input are present constitutively across differentiated tissues (black box), whereas others are present in non-fibroblast tissues (blue box) or are specific to fibroblasts and similar tissue types (green box).

(C) qPCR for DBR and non-DBR sites, plotted relative to the average of the non-DBR sites. Fraction #14 was the gradient fraction showing greatest DBR enrichment. Error bars show SEM of qPCR sites tested.

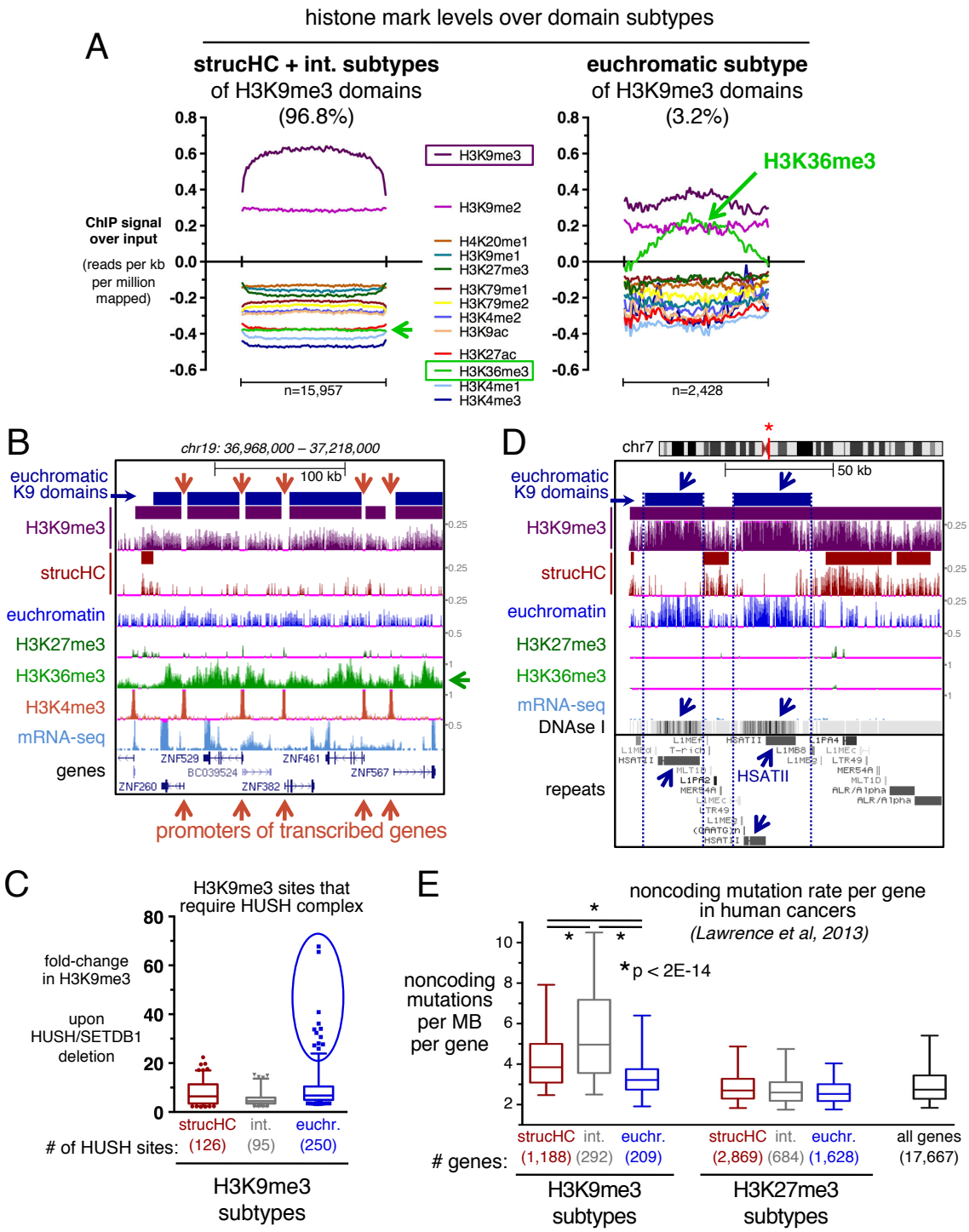
(D) Rate of chromatin recovery by H3K9me3-directed IP, comparing IP from gradient fractions versus IP from input chromatin (conventional ChIP). DNA yield measured by PicoGreen assay.



Supplementary Figure 4-2. Preparation of sequencing libraries for Gradient-seq.

(A) Agarose gel showing sonication of purified DNA for Gradient-seq experiment, to shear large fragments down to a ~200 bp size for library preparation and sequencing.

(B) Number of reads for each DNA sequencing sample, showing the proportion of reads that aligned to the genome and passed filtering to remove duplicate reads. “+K9 IP” indicates the H3K9me3 IPs performed from the strucHC and euchromatin (“euchr.”) fractions.



Supplementary Figure 4-3. Characterization of euchromatic subtype of H3K9me3 domains.

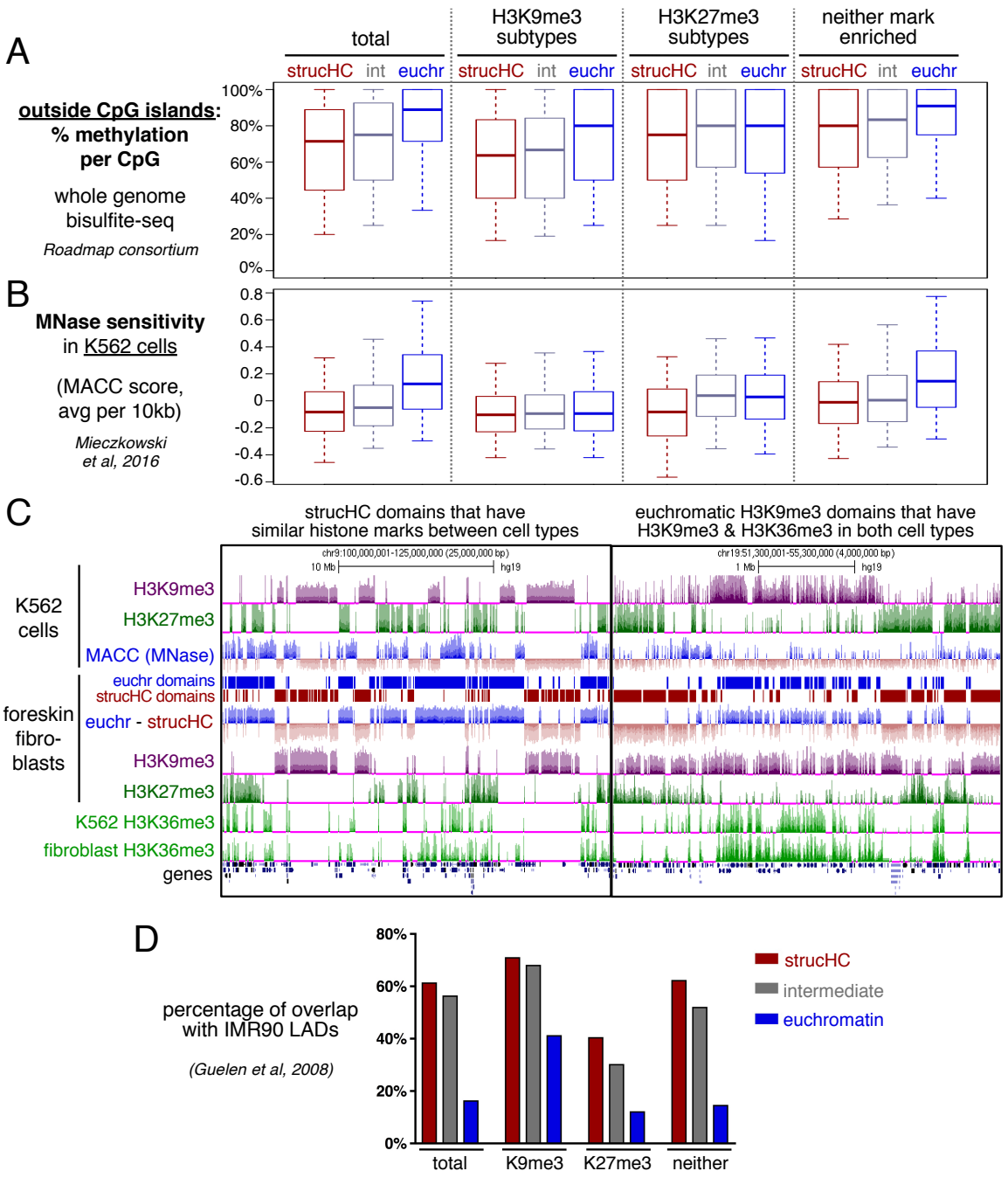
(A) Histone mark levels in human fibroblasts, normalized for input, plotted over subtypes of H3K9me3 domains: the strucHC/intermediate subtypes (left), and the euchromatic subtype (right). All ChIP-seq data was obtained from the Roadmap Epigenomics consortium (Bernstein et al, 2010), except for H3K9me2 ChIP-seq data (Chandra et al, 2012).

(B) Browser view showing enrichment of H3K36me3 (green arrow) in euchromatic H3K9me3 domains containing with transcribed genes. Note that the gaps in H3K9me3 domains (diminution of H3K9me3 signal) coincide with H3K4me3-marked active gene promoters.

(C) Among genomic sites that depend upon the HUSH complex and SETDB1 for H3K9me3 levels, we plotted the fold-change in H3K9me3 upon deletion of HUSH components and SETDB1, using data from (Tchasovnikarova et al, 2015). The 12 most HUSH-dependent sites (points in blue circle) all fell inside the euchromatic subtype of H3K9me3.

(D) Brower view showing overlap of euchromatic H3K9me3 domains over HSATII repeats, which are also Dnase I-hypersensitive (Roadmap Epigenomics track for IMR90 fibroblasts). View is proximal to the chr7 centromere (coordinates 61,721,000-61,871,000).

(E) Structural subtypes of H3K9me3 and H3K27me3 domains were compared for rates of noncoding mutations per gene, averaged across all human cancers, using data from (Lawrence et al, 2013). P values calculated by Wilcoxon rank sum test. Whiskers: 5^{th} - 95^{th} percentiles.



Supplementary Figure 4-4. Relationship of chromatin subtypes from Gradient-seq to DNA methylation, MNase sensitivity, and lamina association.

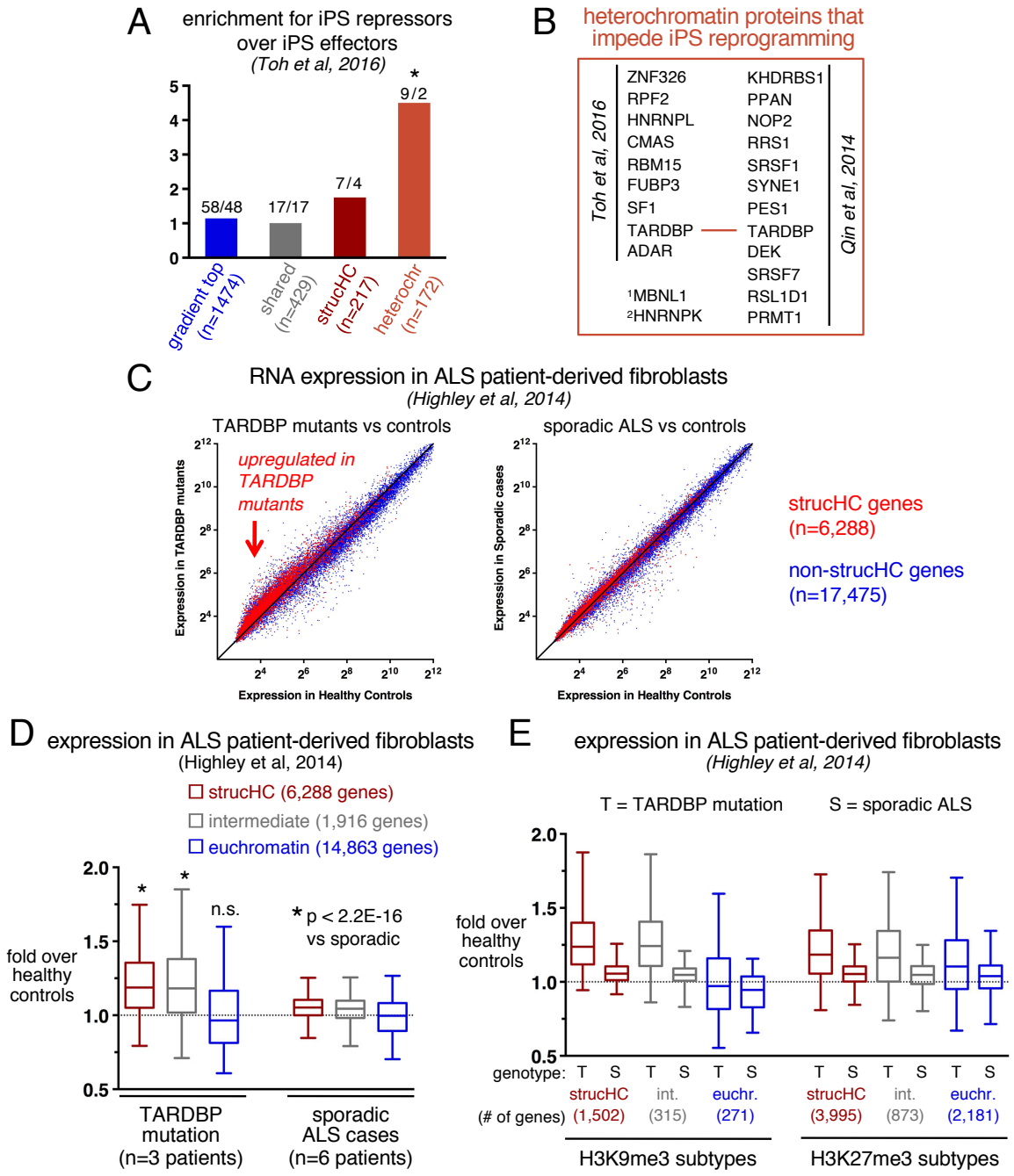
(A) This plot considers only CpG dinucleotides that are not part of CpG islands. Using Gradient-seq data, chromatin was classified into strucHC, intermediate ("int"), and

euchromatin (“euchr”) categories, for the total mappable genome (“total”), within H3K9me3 and H3K27me3 domains, or outside of those domains (“neither mark enriched”). Whole-genome bisulfite sequencing (Roadmap consortium data for foreskin fibroblasts, GSM1127120) was used to determine the frequency of methylation per CpG dinucleotide. The distribution of methylation frequencies is plotted for non-CpG-island CpGs falling within each chromatin subtype (whiskers: 10^{th} and 90^{th} percentiles).

(B) MACC scores, where higher MACC indicate that chromatin is more sensitive to MNase, were determined on a 500-bin basis for K562 cells by Mieczkowski et al., 2016. MACC scores were averaged across 10-kb sliding windows, and the distribution of MACC scores is plotted for all 10-kb windows falling inside a given chromatin subtype (whiskers: 5^{th} and 95^{th} percentiles).

(C) Browser views comparing MACC signal in K562 cells (Mieczkowski et al., 2016) and Gradient-seq signal in foreskin fibroblasts, using regions with similar patterns of H3K9me3, H3K27me3, and H3K36me3 histone marks between the two cell types. Gradient-seq signal is plotted as the euchromatin signal minus the strucHC signal (“euchr - strucHC"). **Left:** Note that the pattern of euchr/strucHC domains is highly similar to the pattern of positive/negative MACC signal, in regions where both cell types have similar H3K9me3 (without H3K36me3) domains or H3K27me3 domains. **Right:** Sites bearing H3K9me3 and H3K36me3 in both cell types are classified as euchromatin according to Gradient-seq, but have negative MACC scores, indicating MNase resistance. All histone mark tracks are replicate-pooled and input-subtracted. Foreskin fibroblast H3K27me3 and H3K36me3 ChIP-seq data was obtained from the Epigenomics Roadmap (GSE16368); H3K9me3 ChIP-seq was from this study. K562 ChIP-seq data was downloaded from ENCODE (GSE29611).

(D) Rate of overlap between each chromatin subtype and Lamina Associated Domains (LADs) mapped in IMR90 fibroblasts (Guelen et al., 2008). The number of megabases of overlap divided by the number of megabases total in the chromatin subtype was used to calculate the percentage.



Supplementary Figure 4-5. Heterochromatin proteins such as TARDBP impede iPS reprogramming and contribute to heterochromatic silencing in human cells.

(A) Proteomic categories were compared to lists of iPS repressors (proteins whose knockdown increases reprogramming) and iPS effectors (knockdown inhibits

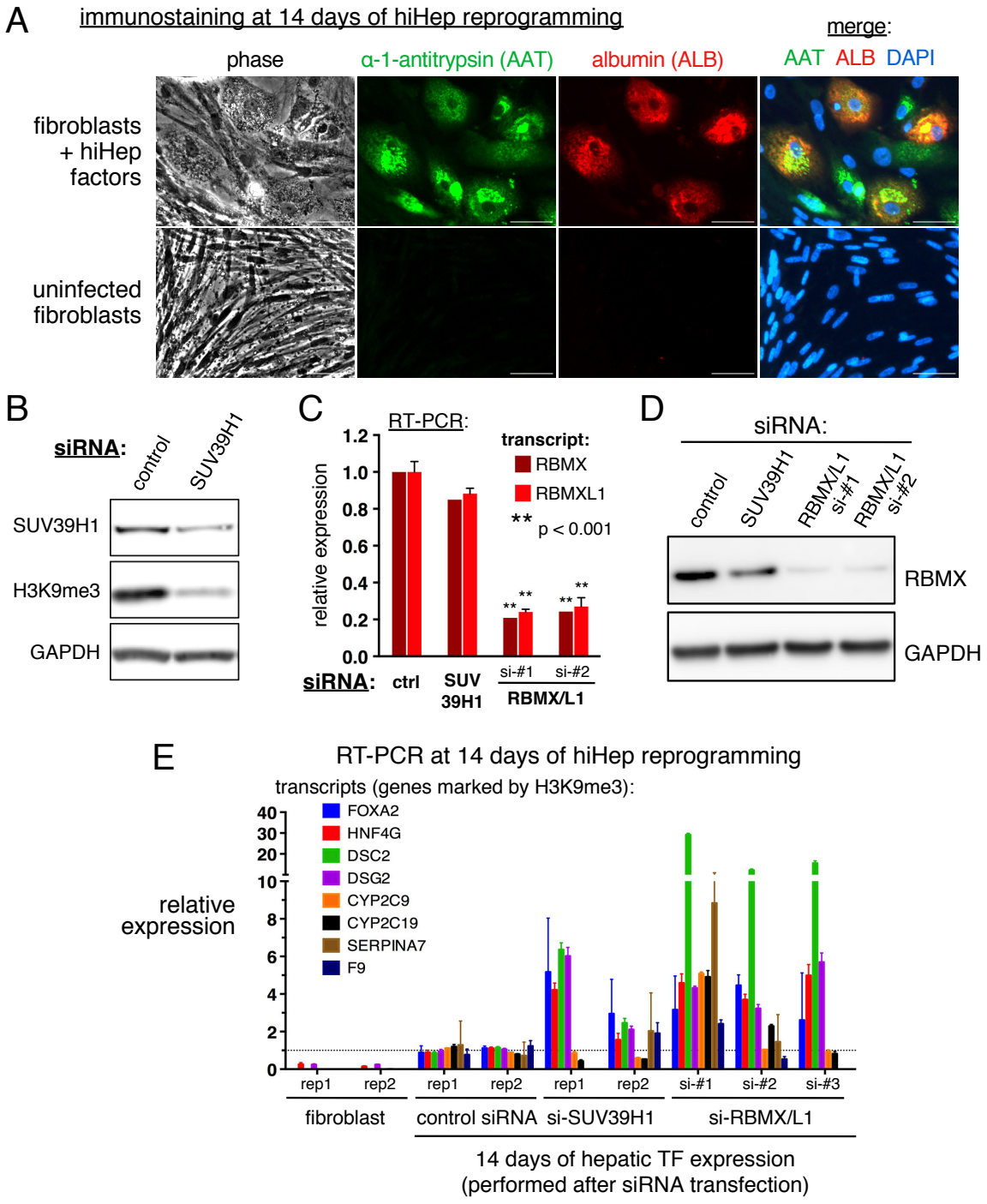
repreogramming) from a genome-wide RNAi screen (Toh et al., 2016). The ratio of repressors to effectors is plotted for each category, and the numbers of repressors and effectors are listed above the bar. Asterisk indicates $P < 0.05$ by simulation test.

(B) Listed are heterochromatin proteins whose knockdown has previously been shown to enhance iPS reprogramming efficiency, along with the corresponding reference. Note that TARDBP knockdown was found to enhance reprogramming in two genome-wide screens. [1]Han et al., 2013; [2]Bao et al., 2015.

(C) Both scatterplots compare gene expression in fibroblasts from ALS patients to fibroblasts from healthy controls, using published microarray data (Highley et al, 2014). Left plot: in fibroblasts from TARDBP-mutant ALS patients, there is upregulation of genes in strucHC (red dots) compared to controls. Right plot: no upregulation of strucHC genes is observed in fibroblasts from sporadic ALS patients, which lack germline TARDBP mutation.

(D) Analysis of microarray data for ALS patient-derived fibroblasts (Highley et al, 2014), with expression plotted for genes in the 3 structural chromatin categories defined by Gradient-seq. Gene expression is normalized relative to the average of healthy patients (n=6). Sporadic ALS cases (no germline TARDBP mutation) serve as a control for TARDBP-mutant fibroblasts. P values calculated using Wilcoxon signed rank test (paired), comparing TARDBP-mutant samples to sporadic ALS samples for the same chromatin category. Whiskers: 5^{th} - 95^{th} percentiles.

(E) Same as Panel D, but showing gene expression for the subtypes of H3K9me3 and H3K27me3 domains. Boxplots are shown for TARDBP mutant fibroblasts (T) and fibroblasts from patients with sporadic ALS (S).



Supplementary Figure 4-6. Knockdown of SUV39H1 and RBMX/L1 enhances hepatic gene activation during hiHep reprogramming.

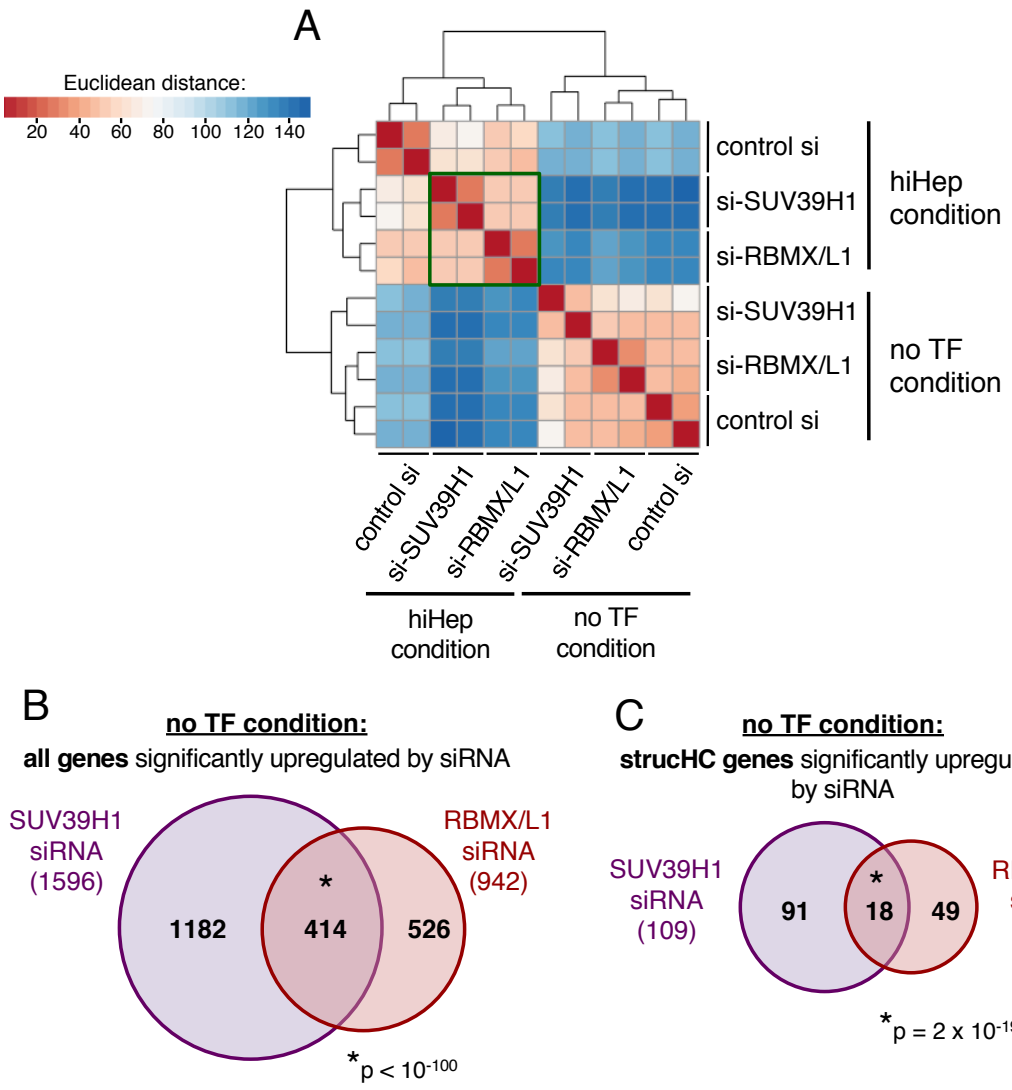
(A) Immunofluorescence imaging confirming induction of hepatic markers after 14 days of hiHep reprogramming with FOXA3, HNF1A, and HNF4A, compared to uninfected fibroblasts cultured under the same conditions. Scale bars, 50μM.

(B) Western blots in fibroblasts after two cycles of SUV39H1 siRNA, compared to non-targeting control siRNA.

(C) RT-PCR showing that RBMX/L1 siRNAs deplete transcripts for both RBMX and RBMXL1, after two cycles of siRNA transfection. Error bars show the SEM of three biological replicates. P values by Student's T test.

(D) Western blots for RBMX protein in fibroblasts after two cycles of siRNA transfection.

(E) RT-PCR analysis of fibroblasts treated with two cycles of siRNA, followed by hiHep reprogramming for 14 days. Three independent siRNAs were tested for the RBMX/L1 knockdown condition; all other samples have two biological replicates shown.



Supplementary Figure 4-7. RBMX represses select strucHC genes in the absence of ectopic transcription factors.

(A) Comparison of mRNA-seq samples by Euclidean distance, using the normalized gene expression scores for each sample. Each sample type has two biological replicates that cluster together. The two experimental conditions, “hiHep condition” and “no TF condition”, refer to siRNA knockdowns performed in the presence and absence of hiHep transcription factors, respectively. The green box highlights the clustering of SUV39H1 siRNA and RBMX/L1 siRNA in the hiHep condition.

(B) Venn diagram of genes significantly upregulated by SUV39H1 and RBMX/L1 siRNA, compared to control siRNA, in the absence of ectopic transcription factors. P-value indicates the significance of the overlap between gene sets.

(C) Among genes inside strucHC domains, comparison of genes upregulated by SUV39H1 and RBMX/L1 siRNA in the no TF condition.

4.8 TABLES

Table 4-1. List of genes in the euchromatic subtype of fibroblast H3K9me3 domains.

Refseq genes with at least 50% overlap with euchromatic H3K9me3 domains were identified and collapsed down to 308 unique gene symbols. The amount of overlap is reported in the right-most column, as a fraction of the gene body length. (If there were multiple Refseq genes for a given symbol, the one with the highest rate of overlap was chosen.)

RefSeq ID	gene symbol	name	fraction of gene in domain
NR_002451	**ABCA11P**	ATP binding cassette subfamily A member 11, pseudogene	0.94
NM_000677	**ADORA3**	adenosine A3 receptor	0.84
NM_139058	**ARX**	aristaless related homeobox	0.98
NM_001098169	**BSX**	brain specific homeobox	1
NM_144645	**C4orf36**	chromosome 4 open reading frame 36	0.52
NM_001017978	**CT83**	cancer/testis antigen 83	1
NR_033339	**CTBP1-AS2**	CTBP1 antisense RNA 2 (head to head)	0.5
NM_001565	**CXCL10**	C-X-C motif chemokine ligand 10	1
NM_001029865	**DBX1**	developing brain homeobox 1	1
NM_001037499	**DEFB114**	defensin beta 114	1
NM_001426	**EN1**	engrailed homeobox 1	1
NM_001278182	**EOMES**	eomesodermin	1
NM_001007253	**ERV3-1**	endogenous retrovirus group 3 member 1	0.78
NM_001080458	**EVX2**	even-skipped homeobox 2	1
NM_022774	**EXO5**	exonuclease 5	0.86
NM_012159	**FBXL21**	F-box and leucine rich repeat protein 21 (gene/pseudogene)	0.88
NM_005249	**FOXG1**	forkhead box G1	1
NR_026718	**FOXO3B**	forkhead box O3B pseudogene	1
NR_102763	**GATA6-AS1**	GATA6 antisense RNA 1 (head to head)	1
NM_001301687	**GBX2**	gastrulation brain homeobox 2	1
NM_004821	**HAND1**	heart and neural crest derivatives expressed 1	1
NM_001105574	**HMX3**	H6 family homeobox 3	1
NR_033205	**HOXB-AS3**	HOXB cluster antisense RNA 3	0.74
NM_018952	**HOXB6**	homeobox B6	0.79
NM_004502	**HOXB7**	homeobox B7	1
NM_024016	**HOXB8**	homeobox B8	0.88
NM_017410	**HOXC13**	homeobox C13	1
NR_047507	**HOXC13-AS**	HOXC13 antisense RNA	1
NR_038435	**HOXD-AS2**	HOXD cluster antisense RNA 2	1

NR_023915	**IPW**	imprinted in Prader-Willi syndrome (non-protein coding)	0.78
NM_001320334	**KIAA1143**	KIAA1143	0.58
NM_152349	**KRT222**	keratin 222	0.9
NM_005568	**LHX1**	LIM homeobox 1	0.87
NR_038400	**LINC00308**	long intergenic non-protein coding RNA 308	0.53
NR_135279	**LINC00334**	long intergenic non-protein coding RNA 334	0.61
NR_103734	**LINC00342**	long intergenic non-protein coding RNA 342	0.84
NR_104059	**LINC00397**	long intergenic non-protein coding RNA 397	1
NR_024383	**LINC00461**	long intergenic non-protein coding RNA 461	1
NR_047483	**LINC00554**	long intergenic non-protein coding RNA 554	0.95
NR_038914	**LINC01349**	long intergenic non-protein coding RNA 1349	0.55
NR_120618	**LINC01475**	long intergenic non-protein coding RNA 1475	1
NR_026732	**LINC01551**	long intergenic non-protein coding RNA 1551	1
NR_036508	**LOC100128398**		1
NR_104179	**LOC100289333**		0.64
NR_135129	**LOC100505588**		1
NR_134920	**LOC101928238**		1
NR_110091	**LOC101928597**		1
NR_125869	**LOC102723883**		0.85
NR_110544	**LOC102724188**		0.79
NR_135110	**LOC105376633**		0.59
NM_032517	**LYZL1**	lysozyme like 1	0.61
NM_001308244	**MAPK9**	mitogen-activated protein kinase 9	0.61
NM_001190787	**MCIDAS**	multiciliate differentiation and DNA synthesis associated cell cycle protein	1
NR_036053	**MIR1270**	microRNA 1270	1
NR_036235	**MIR4273**	microRNA 4273	1
NR_039722	**MIR4500**	microRNA 4500	1
NR_030297	**MIR571**	microRNA 571	1
NR_030373	**MIR643**	microRNA 643	1
NR_106898	**MIR6839**	microRNA 6839	1
NR_030741	**MIR9-2**	microRNA 9-2	1
NM_145285	**NKX2-3**	NK2 homeobox 3	1
NM_030905	**OR2J2**	olfactory receptor family 2 subfamily J member 2	1
NM_032109	**OTP**	orthopedia homeobox	1
NM_001310159	**PAX6**	paired box 6	0.66
NR_046636	**PCDH9-AS3**	PCDH9 antisense RNA 3	0.7
NM_032009	**PCDHGA2**	protocadherin gamma subfamily A, 2	1
NM_032011	**PCDHGA3**	protocadherin gamma subfamily A, 3	1
NM_032053	**PCDHGA4**	protocadherin gamma subfamily A, 4	1
NM_032054	**PCDHGA5**	protocadherin gamma subfamily A, 5	1
NM_032086	**PCDHGA6**	protocadherin gamma subfamily A, 6	1
NM_032087	**PCDHGA7**	protocadherin gamma subfamily A, 7	0.93
NM_014004	**PCDHGA8**	protocadherin gamma subfamily A, 8	1
NM_032089	**PCDHGA9**	protocadherin gamma subfamily A, 9	1
NM_032096	**PCDHGB2**	protocadherin gamma subfamily B, 2	1
NM_032097	**PCDHGB3**	protocadherin gamma subfamily B, 3	1
NM_032098	**PCDHGB4**	protocadherin gamma subfamily B, 4	1
NM_032099	**PCDHGB5**	protocadherin gamma subfamily B, 5	1
NM_003924	**PHOX2B**	paired like homeobox 2b	1
NM_021620	**PRDM13**	PR/SET domain 13	1
NR_023917	**PTENP1**	phosphatase and tensin homolog pseudogene 1	1
NR_103745	**PTENP1-AS**	PTENP1 antisense RNA	0.81
NM_001080521	**RASSF10**	Ras association domain family member 10	0.62

NM_021163	**RBAK**	RB associated KRAB zinc finger	0.64
NM_001204513	**RBAK-RBAKDN**	RBAK-RBAKDN readthrough	0.55
NR_125730	**RNU6-2**	RNA, U6 small nuclear 2	1
NM_001193307	**SAMD9**	sterile alpha motif domain containing 9	0.77
NR_026830	**SATB2-AS1**	SATB2 antisense RNA 1	1
NM_152679	**SLC10A4**	solute carrier family 10 member 4	0.82
NM_005642	**TAF7**	TATA-box binding protein associated factor 7	1
NM_006593	**TBR1**	T-box, brain 1	0.55
NM_033208	**TIGD7**	tigger transposable element derived 7	0.56
NM_016170	**TLX2**	T-cell leukemia homeobox 2	1
NR_037893	**TMCC1-AS1**	TMCC1 antisense RNA 1 (head to head)	0.55
NM_020683	**TMIGD3**	transmembrane and immunoglobulin domain containing 3	0.53
NR_003148	**TPM3P9**	tropomyosin 3 pseudogene 9	0.78
NR_001524	**TTTY3**	testis-specific transcript, Y-linked 3 (non-protein coding)	1
NR_002176	**TTTY3B**	testis-specific transcript, Y-linked 3B (non-protein coding)	1
NM_020633	**VN1R1**	vomeronasal 1 receptor 1	1
NM_175619	**ZAR1**	zygote arrest 1	0.78
NM_203303	**ZCCHC13**	zinc finger CCHC-type containing 13	1
NM_001318475	**ZFP1**	ZFP1 zinc finger protein	0.5
NM_001308440	**ZFP28**	ZFP28 zinc finger protein	1
NM_153018	**ZFP3**	ZFP3 zinc finger protein	0.71
NM_001320666	**ZFP30**	ZFP30 zinc finger protein	0.81
NM_001282515	**ZFP37**	ZFP37 zinc finger protein	1
NM_001320178	**ZFP69**	ZFP69 zinc finger protein	0.68
NM_023070	**ZFP69B**	ZFP69 zinc finger protein B	0.91
NM_133466	**ZFP82**	ZFP82 zinc finger protein	0.96
NM_001305203	**ZFP90**	ZFP90 zinc finger protein	0.84
NM_033132	**ZIC5**	Zic family member 5	0.69
NM_001010879	**ZIK1**	zinc finger protein interacting with K protein 1	0.77
NM_173531	**ZNF100**	zinc finger protein 100	0.61
NM_033204	**ZNF101**	zinc finger protein 101	0.75
NM_001013746	**ZNF107**	zinc finger protein 107	0.6
NM_015852	**ZNF117**	zinc finger protein 117	1
NM_006956	**ZNF12**	zinc finger protein 12	0.81
NM_001008727	**ZNF121**	zinc finger protein 121	0.78
NM_001297568	**ZNF124**	zinc finger protein 124	0.81
NM_003433	**ZNF132**	zinc finger protein 132	1
NM_003435	**ZNF134**	zinc finger protein 134	1
NR_023311	**ZNF137P**	zinc finger protein 137, pseudogene	0.59
NM_021030	**ZNF14**	zinc finger protein 14	1
NM_001300776	**ZNF140**	zinc finger protein 140	1
NM_003441	**ZNF141**	zinc finger protein 141	0.93
NM_001029976	**ZNF16**	zinc finger protein 16	0.63
NM_003450	**ZNF174**	zinc finger protein 174	0.84
NM_007147	**ZNF175**	zinc finger protein 175	0.51
NM_001172651	**ZNF177**	zinc finger protein 177	1
NM_001278508	**ZNF180**	zinc finger protein 180	0.67
NM_001029997	**ZNF181**	zinc finger protein 181	0.64
NM_001007088	**ZNF182**	zinc finger protein 182	0.53
NM_001318892	**ZNF184**	zinc finger protein 184	0.79
NM_001130519	**ZNF195**	zinc finger protein 195	0.84

NM_006991	**ZNF197**	zinc finger protein 197	0.51
NM_001203250	**ZNF20**	zinc finger protein 20	0.75
NM_001145447	**ZNF200**	zinc finger protein 200	0.64
NM_001265597	**ZNF211**	zinc finger protein 211	0.85
NM_013359	**ZNF221**	zinc finger protein 221	0.54
NM_001129996	**ZNF222**	zinc finger protein 222	0.68
NM_001032372	**ZNF226**	zinc finger protein 226	0.75
NM_014518	**ZNF229**	zinc finger protein 229	0.93
NM_145911	**ZNF23**	zinc finger protein 23	0.79
NM_001320952	**ZNF232**	zinc finger protein 232	0.84
NM_005674	**ZNF239**	zinc finger protein 239	0.51
NR_023392	**ZNF252P**	zinc finger protein 252, pseudogene	0.63
NM_021047	**ZNF253**	zinc finger protein 253	0.94
NM_005773	**ZNF256**	zinc finger protein 256	0.55
NM_001256279	**ZNF26**	zinc finger protein 26	0.94
NM_001012756	**ZNF260**	zinc finger protein 260	0.87
NM_005741	**ZNF263**	zinc finger protein 263	0.56
NM_001265588	**ZNF267**	zinc finger protein 267	0.65
NM_001165881	**ZNF268**	zinc finger protein 268	1
NR_024565	**ZNF271P**	zinc finger protein 271, pseudogene	0.69
NM_021148	**ZNF273**	zinc finger protein 273	0.65
NM_001278734	**ZNF274**	zinc finger protein 274	0.77
NM_006969	**ZNF28**	zinc finger protein 28	0.51
NM_001297752	**ZNF283**	zinc finger protein 283	0.88
NM_001130842	**ZNF286A**	zinc finger protein 286A	0.78
NM_001145045	**ZNF286B**	zinc finger protein 286B	0.89
NM_020653	**ZNF287**	zinc finger protein 287	0.81
NM_001172831	**ZNF300**	zinc finger protein 300	1
NR_026867	**ZNF300P1**	zinc finger protein 300 pseudogene 1	1
NM_001012320	**ZNF302**	zinc finger protein 302	0.74
NM_001190791	**ZNF317**	zinc finger protein 317	0.63
NM_001005368	**ZNF32**	zinc finger protein 32	0.66
NR_047557	**ZNF32-AS1**	ZNF32 antisense RNA 1	1
NR_047558	**ZNF32-AS2**	ZNF32 antisense RNA 2	0.53
NM_207333	**ZNF320**	zinc finger protein 320	0.76
NM_001242797	**ZNF322**	zinc finger protein 322	0.65
NM_024620	**ZNF329**	zinc finger protein 329	0.72
NM_001242475	**ZNF345**	zinc finger protein 345	0.97
NM_001172674	**ZNF347**	zinc finger protein 347	0.74
NM_021632	**ZNF350**	zinc finger protein 350	1
NR_103847	**ZNF350-AS1**	ZNF350 antisense RNA 1	0.77
NM_014594	**ZNF354C**	zinc finger protein 354C	0.95
NM_001007094	**ZNF37A**	zinc finger protein 37A	0.89
NR_026777	**ZNF37BP**	zinc finger protein 37B, pseudogene	0.78
NM_001256838	**ZNF382**	zinc finger protein 382	0.84
NM_152604	**ZNF383**	zinc finger protein 383	1
NM_032347	**ZNF397**	zinc finger protein 397	0.88
NM_001033719	**ZNF404**	zinc finger protein 404	1
NM_017879	**ZNF416**	zinc finger protein 416	1
NM_001297734	**ZNF417**	zinc finger protein 417	0.86
NM_144689	**ZNF420**	zinc finger protein 420	0.74
NM_001300883	**ZNF426**	zinc finger protein 426	1
NM_001001415	**ZNF429**	zinc finger protein 429	0.97
NM_001256654	**ZNF43**	zinc finger protein 43	0.95

NM_014650	**ZNF432**	zinc finger protein 432	0.83
NM_001080411	**ZNF433**	zinc finger protein 433	0.93
NM_030634	**ZNF436**	zinc finger protein 436	0.85
NM_152262	**ZNF439**	zinc finger protein 439	1
NM_001164276	**ZNF44**	zinc finger protein 44	0.71
NM_152357	**ZNF440**	zinc finger protein 440	0.53
NM_152355	**ZNF441**	zinc finger protein 441	0.81
NM_030824	**ZNF442**	zinc finger protein 442	0.51
NM_005815	**ZNF443**	zinc finger protein 443	0.66
NM_001297623	**ZNF461**	zinc finger protein 461	0.94
NM_001008801	**ZNF468**	zinc finger protein 468	0.69
NM_001001668	**ZNF470**	zinc finger protein 470	1
NM_020813	**ZNF471**	zinc finger protein 471	0.87
NM_001297624	**ZNF480**	zinc finger protein 480	0.91
NM_133464	**ZNF483**	zinc finger protein 483	0.61
NM_152356	**ZNF491**	zinc finger protein 491	0.79
NM_001076678	**ZNF493**	zinc finger protein 493	0.95
NM_001258280	**ZNF501**	zinc finger protein 501	1
NM_001134440	**ZNF502**	zinc finger protein 502	1
NM_001099269	**ZNF506**	zinc finger protein 506	1
NM_001314059	**ZNF510**	zinc finger protein 510	0.79
NM_001318005	**ZNF514**	zinc finger protein 514	0.94
NM_145287	**ZNF519**	zinc finger protein 519	0.51
NR_003699	**ZNF525**	zinc finger protein 525	0.54
NM_032453	**ZNF527**	zinc finger protein 527	0.66
NR_125345	**ZNF528-AS1**	ZNF528 antisense RNA 1	1
NR_027239	**ZNF529**	zinc finger protein 529	0.93
NR_110703	**ZNF529-AS1**	ZNF529 antisense RNA 1	0.88
NM_152606	**ZNF540**	zinc finger protein 540	0.61
NR_003127	**ZNF542P**	zinc finger protein 542, pseudogene	0.91
NM_001297763	**ZNF546**	zinc finger protein 546	0.56
NM_001199295	**ZNF549**	zinc finger protein 549	0.9
NM_001277090	**ZNF550**	zinc finger protein 550	0.63
NM_001270938	**ZNF551**	zinc finger protein 551	1
NM_001172775	**ZNF555**	zinc finger protein 555	0.74
NM_001044387	**ZNF557**	zinc finger protein 557	0.81
NM_001304350	**ZNF558**	zinc finger protein 558	0.62
NM_001202406	**ZNF559**	zinc finger protein 559	1
NM_001172650	**ZNF559-ZNF177**	ZNF559-ZNF177 readthrough	0.85
NM_152289	**ZNF561**	zinc finger protein 561	0.75
NM_001130031	**ZNF562**	zinc finger protein 562	0.91
NM_145276	**ZNF563**	zinc finger protein 563	0.69
NM_144976	**ZNF564**	zinc finger protein 564	0.72
NM_152603	**ZNF567**	zinc finger protein 567	0.96
NM_152484	**ZNF569**	zinc finger protein 569	0.81
NM_144694	**ZNF570**	zinc finger protein 570	0.94
NM_001290314	**ZNF571**	zinc finger protein 571	0.93
NR_038249	**ZNF571-AS1**	ZNF571 antisense RNA 1	0.91
NM_152412	**ZNF572**	zinc finger protein 572	1
NR_037159	**ZNF582-AS1**	ZNF582 antisense RNA 1 (head to head)	0.55
NM_001159860	**ZNF583**	zinc finger protein 583	1
NR_110152	**ZNF585A**	zinc finger protein 585A	1
NM_152279	**ZNF585B**	zinc finger protein 585B	0.57

NM_001077426	**ZNF586**	zinc finger protein 586	0.86
NM_001204817	**ZNF587**	zinc finger protein 587	0.65
NM_032530	**ZNF594**	zinc finger protein 594	0.95
NM_001286052	**ZNF595**	zinc finger protein 595	0.66
NM_001042415	**ZNF596**	zinc finger protein 596	0.73
NM_001007248	**ZNF599**	zinc finger protein 599	0.89
NM_001164715	**ZNF605**	zinc finger protein 605	0.83
NM_025027	**ZNF606**	zinc finger protein 606	0.9
NM_001031721	**ZNF613**	zinc finger protein 613	0.57
NM_025040	**ZNF614**	zinc finger protein 614	0.92
NM_001199324	**ZNF615**	zinc finger protein 615	0.91
NM_178523	**ZNF616**	zinc finger protein 616	0.91
NM_014789	**ZNF623**	zinc finger protein 623	0.77
NM_145233	**ZNF625**	zinc finger protein 625	0.74
NR_037802	**ZNF625-ZNF20**	ZNF625-ZNF20 readthrough (NMD candidate)	0.68
NM_145297	**ZNF626**	zinc finger protein 626	1
NM_173658	**ZNF660**	zinc finger protein 660	0.55
NR_036521	**ZNF667-AS1**	ZNF667 antisense RNA 1 (head to head)	0.9
NM_001142572	**ZNF669**	zinc finger protein 669	0.51
NM_001317998	**ZNF677**	zinc finger protein 677	1
NM_138286	**ZNF681**	zinc finger protein 681	0.79
NM_021915	**ZNF69**	zinc finger protein 69	0.74
NM_020394	**ZNF695**	zinc finger protein 695	0.54
NM_198535	**ZNF699**	zinc finger protein 699	0.82
NM_001282796	**ZNF7**	zinc finger protein 7	0.77
NM_001271848	**ZNF700**	zinc finger protein 700	0.9
NM_001172655	**ZNF701**	zinc finger protein 701	0.7
NM_001290210	**ZNF717**	zinc finger protein 717	0.53
NM_001289930	**ZNF718**	zinc finger protein 718	0.95
NM_001130913	**ZNF720**	zinc finger protein 720	0.57
NM_133474	**ZNF721**	zinc finger protein 721	0.88
NR_045525	**ZNF724P**		0.94
NM_001170905	**ZNF736**	zinc finger protein 736	0.8
NM_001159293	**ZNF737**	zinc finger protein 737	0.96
NR_027130	**ZNF738**	zinc finger protein 738	0.74
NR_126069	**ZNF75A**	zinc finger protein 75a	1
NM_001185063	**ZNF75D**	zinc finger protein 75D	1
NM_001289952	**ZNF761**	zinc finger protein 761	1
NM_001012753	**ZNF763**	zinc finger protein 763	0.53
NM_001040185	**ZNF765**	zinc finger protein 765	0.79
NM_001010851	**ZNF766**	zinc finger protein 766	0.75
NM_001024596	**ZNF772**	zinc finger protein 772	1
NR_130705	**ZNF773**	zinc finger protein 773	0.71
NM_173632	**ZNF776**	zinc finger protein 776	0.93
NM_001005851	**ZNF780B**	zinc finger protein 780B	0.85
NM_001242802	**ZNF790**	zinc finger protein 790	0.88
NR_040027	**ZNF790-AS1**	ZNF790 antisense RNA 1	1
NM_153358	**ZNF791**	zinc finger protein 791	0.59
NM_001080821	**ZNF799**	zinc finger protein 799	0.59
NM_001031665	**ZNF816**	zinc finger protein 816	0.91
NR_073396	**ZNF818P**	zinc finger protein 818, pseudogene	0.73
NM_001080493	**ZNF823**	zinc finger protein 823	0.59
NM_001171979	**ZNF829**	zinc finger protein 829	0.9
NM_001102657	**ZNF836**	zinc finger protein 836	0.65

NM_001127372	**ZNF84**	zinc finger protein 84	1
NM_001136499	**ZNF841**	zinc finger protein 841	1
NM_001136501	**ZNF844**	zinc finger protein 844	0.92
NM_138374	**ZNF845**	zinc finger protein 845	0.62
NM_001077624	**ZNF846**	zinc finger protein 846	0.79
NM_001193552	**ZNF850**	zinc finger protein 850	0.94
NM_001287349	**ZNF852**	zinc finger protein 852	0.77
NM_001136116	**ZNF879**	zinc finger protein 879	0.92
NM_001145434	**ZNF880**	zinc finger protein 880	1
NM_001277291	**ZNF891**	zinc finger protein 891	0.9
NM_001300951	**ZNF91**	zinc finger protein 91	0.94
NM_031218	**ZNF93**	zinc finger protein 93	0.72
NM_152455	**ZSCAN29**	zinc finger and SCAN domain containing 29	0.94
NM_001112734	**ZSCAN30**	zinc finger and SCAN domain containing 30	0.56

Table 4-2. List of strucHC-enriched proteins.

Shown are the 217 proteins enriched in the strucHC fraction compared to the euchromatin fraction. The fold-enrichement column is computed by the relative change in protein rank between the samples. The right-most column indicates “Y” (yes) if the protein was also one of the 172 heterochromatin proteins (enriched in the strucHC + H3K9me3 IP sample).

gene symbol	protein	fold-enrichment in strucHC vs euchromatin	t-test	het protein?
EBNA1BP2	Probable rRNA-processing protein EBP2	18.2	0.0398	Y
SRSF10	Serine/arginine-rich splicing factor 10	17.1	0.0221	Y
RBMX	RNA-binding motif protein, X chromosome	15.8	0.0047	Y
SRSF9	Serine/arginine-rich splicing factor 9	13.4	0.0369	Y
HNRNPA2B1	Heterogeneous nuclear ribonucleoproteins A2/B1	13.0	0.0199	Y
BRIX1	Ribosome biogenesis protein BRX1 homolog	12.5	0.0152	Y
RRS1	Ribosome biogenesis regulatory protein homolog	12.2	0.0401	Y
LLPH	Protein LLP homolog	11.2	0.0430	
HNRNPC	Heterogeneous nuclear ribonucleoproteins C1/C2	10.8	0.0285	Y
KHDRBS3	KH domain-containing, RNA-binding, signal transduction-associated protein 3	10.7	0.0426	Y
POLDIP3	Polymerase delta-interacting protein 3	10.5	0.0170	Y
HNRNPM	Heterogeneous nuclear ribonucleoprotein M	10.4	0.0004	Y
RPF2	Ribosome production factor 2 homolog	9.9	0.0215	Y
TRA2B	Transformer-2 protein homolog beta	8.9	0.0013	Y
ELN	Elastin	8.4	0.0450	
RBM14	RNA-binding protein 14	8.1	0.0213	Y
FTSJ3	pre-rRNA processing protein FTSJ3	8.1	0.0442	Y
SRSF1	Serine/arginine-rich splicing factor 1	8.0	0.0039	Y
NCOA5	Nuclear receptor coactivator 5	8.0	0.0416	Y
HIST1H1A	Histone H1.1	8.0	0.0096	Y
GNL3	Guanine nucleotide-binding protein-like 3	7.8	0.0137	Y
DDX27	Probable ATP-dependent RNA helicase DDX27	7.7	0.0325	Y
DDX17	Probable ATP-dependent RNA helicase DDX17	7.7	0.0319	Y
HNRNPK	Heterogeneous nuclear ribonucleoprotein K	7.5	0.0057	Y
GTPBP4	Nucleolar GTP-binding protein 1	7.4	0.0001	Y
MATR3	Matrin-3	7.4	0.0143	Y
FYTTD1	UAP56-interacting factor	7.1	0.0438	Y
TAF15	TATA-binding protein-associated factor 2N	7.1	0.0086	Y
ZNF326	DBIRD complex subunit ZNF326	7.1	0.0405	Y
DDX5	Probable ATP-dependent RNA helicase DDX5	7.0	0.0383	Y
RPL4	60S ribosomal protein L4	6.9	0.0021	
DCD	Dermcidin	6.7	0.0047	Y
HNRNPA0	Heterogeneous nuclear ribonucleoprotein A0	6.7	0.0017	Y
HNRNPU	Heterogeneous nuclear ribonucleoprotein U	6.5	0.0117	Y
NOP2	Probable 28S rRNA (cytosine(4447)-C(5))-methyltransferase	6.4	0.0188	Y
HNRNPR	Heterogeneous nuclear ribonucleoprotein R	6.4	0.0385	Y

HNRNPH3	Heterogeneous nuclear ribonucleoprotein H3	6.4	0.0003	Y
DDX54	ATP-dependent RNA helicase DDX54	6.3	0.0374	Y
HNRNPF	Heterogeneous nuclear ribonucleoprotein F	6.3	0.0044	Y
NOP16	Nucleolar protein 16	6.3	0.0489	
TRA2A	Transformer-2 protein homolog alpha	6.2	0.0002	Y
HNRNPH1	Heterogeneous nuclear ribonucleoprotein H	6.2	0.0005	Y
WDR46	WD repeat-containing protein 46	6.1	0.0500	Y
NSA2	Ribosome biogenesis protein NSA2 homolog	6.0	0.0378	
H1F0	Histone H1.0	6.0	0.0048	Y
CHTOP	Chromatin target of PRMT1 protein	6.0	0.0015	Y
SRSF7	Serine/arginine-rich splicing factor 7	6.0	0.0125	Y
PPAN	Suppressor of SWI4 1 homolog	6.0	0.0472	Y
NOC3L	Nucleolar complex protein 3 homolog	5.9	0.0351	Y
ALYREF	THO complex subunit 4	5.9	0.0039	Y
KHDRBS1	KH domain-containing, RNA-binding, signal transduction-associated protein 1	5.7	0.0076	Y
HNRNPL	Heterogeneous nuclear ribonucleoprotein L	5.7	0.0024	Y
HNRNPDL	Heterogeneous nuclear ribonucleoprotein D-like	5.7	0.0129	Y
CCDC137	Coiled-coil domain-containing protein 137	5.6	0.0422	
BOP1	Ribosome biogenesis protein BOP1	5.6	0.0438	
EWSR1	RNA-binding protein EWS	5.5	0.0008	Y
MAK16	Protein MAK16 homolog	5.3	0.0443	
SLTM	SAFB-like transcription modulator	5.3	0.0294	
RALY	RNA-binding protein Raly	5.3	0.0009	Y
FUS	RNA-binding protein FUS	5.2	0.0482	Y
DDX56	Probable ATP-dependent RNA helicase DDX56	5.1	0.0484	
NONO	Non-POU domain-containing octamer-binding protein	5.0	0.0079	Y
FN1	Fibronectin	5.0	0.0298	
PES1	Pescadillo homolog	4.9	0.0005	Y
UTP14A	U3 small nucleolar RNA-associated protein 14 homolog A	4.9	0.0500	
RBMXL1	RNA binding motif protein, X-linked-like-1	4.8	0.0480	
CCDC86	Coiled-coil domain-containing protein 86	4.7	0.0483	
RRP15	RRP15-like protein	4.6	0.0254	
RNPS1	RNA-binding protein with serine-rich domain 1	4.6	0.0362	
TRIM4	Tripartite motif-containing protein 4	4.6	0.0351	
NIFK	MKI67 FHA domain-interacting nucleolar phosphoprotein	4.5	0.0168	
RBFOX2	RNA binding protein fox-1 homolog 2	4.5	0.0083	
RBM28	RNA-binding protein 28	4.4	0.0412	Y
S100A9	Protein S100-A9	4.4	0.0414	Y
SAFB	Scaffold attachment factor B1	4.4	0.0203	Y
S100A8	Protein S100-A8	4.4	0.0439	Y
SFPQ	Splicing factor, proline- and glutamine-rich	4.2	0.0037	Y
YTHDC1	YTH domain-containing protein 1	4.2	0.0401	
AKAP8	A-kinase anchor protein 8	4.2	0.0472	Y
SRSF4	Serine/arginine-rich splicing factor 4	4.2	0.0354	
PWP2	Periodic tryptophan protein 2 homolog	4.1	0.0385	Y
SAFB2	Scaffold attachment factor B2	4.1	0.0394	
NOL7	Nucleolar protein 7	4.1	0.0103	
GLTSCR2	Glioma tumor suppressor candidate region gene 2 protein	4.1	0.0401	Y
MYEF2	Myelin expression factor 2	4.1	0.0496	Y

HNRNPUL1	Heterogeneous nuclear ribonucleoprotein U-like protein 1	4.1	0.0029	Y
FBN1	Fibrillin-1	4.0	0.0459	
HNRNPLL	Heterogeneous nuclear ribonucleoprotein L-like	3.9	0.0261	Y
CALML5	Calmodulin-like protein 5	3.9	0.0490	
RPL17	60S ribosomal protein L17	3.9	0.0066	
PUM3	Pumilio domain-containing protein KIAA0020	3.8	0.0397	
HNRNPUL2	Heterogeneous nuclear ribonucleoprotein U-like protein 2	3.8	0.0244	Y
HNRNPAB	Heterogeneous nuclear ribonucleoprotein A/B	3.8	0.0456	Y
DDX50	ATP-dependent RNA helicase DDX50	3.7	0.0307	Y
SRSF6	Serine/arginine-rich splicing factor 6	3.7	0.0211	Y
SRSF5	Serine/arginine-rich splicing factor 5	3.6	0.0221	Y
RRP9	U3 small nucleolar RNA-interacting protein 2	3.6	0.0083	
DDX18	ATP-dependent RNA helicase DDX18	3.5	0.0203	
DIMT1	Probable dimethyladenosine transferase	3.4	0.0153	
HNRNPD	Heterogeneous nuclear ribonucleoprotein D0	3.4	0.0043	Y
H1FX	Histone H1x	3.3	0.0381	Y
ZNF438	Zinc finger protein 438	3.3	0.0273	Y
RAVER1	Ribonucleoprotein PTB-binding 1	3.3	0.0228	Y
YLPM1	YLP motif-containing protein 1	3.3	0.0500	Y
RPL14	60S ribosomal protein L14	3.3	0.0329	Y
TBL3	Transducin beta-like protein 3	3.3	0.0139	Y
DDX21	Nucleolar RNA helicase 2	3.3	0.0151	Y
SAP18	Histone deacetylase complex subunit SAP18	3.3	0.0198	
NAT10	N-acetyltransferase 10	3.2	0.0007	
RRP1	Ribosomal RNA processing protein 1 homolog A	3.2	0.0391	
KIAA2026	Uncharacterized protein KIAA2026	3.2	0.0109	
NOC2L	Nucleolar complex protein 2 homolog	3.2	0.0388	
RBM15	Putative RNA-binding protein 15	3.1	0.0496	Y
FBLN5	Fibulin-5	3.1	0.0480	
DDX47	Probable ATP-dependent RNA helicase DDX47	3.1	0.0291	
RSL1D1	Ribosomal L1 domain-containing protein 1	3.1	0.0027	Y
LUC7L3	Luc7-like protein 3	3.0	0.0174	
XRN2	5-3 exoribonuclease 2	3.0	0.0030	Y
WDR43	WD repeat-containing protein 43	3.0	0.0192	Y
PSPC1	Paraspeckle component 1	3.0	0.0086	Y
HSPG2	Basement membrane-specific heparan sulfate proteoglycan core protein	2.9	0.0252	
DES	Desmin	2.9	0.0155	
RPL24	60S ribosomal protein L24	2.9	0.0074	Y
RPS6	40S ribosomal protein S6	2.8	0.0483	
RBM4	RNA-binding protein 4	2.8	0.0145	Y
RPL21	60S ribosomal protein L21	2.8	0.0141	
MBNL1	Muscleblind-like protein 1	2.8	0.0001	Y
ALKBH5	RNA demethylase ALKBH5	2.8	0.0411	
H2AFY	Core histone macro-H2A.1	2.8	0.0050	Y
ZFR	Zinc finger RNA-binding protein	2.8	0.0263	
SHROOM3	Protein Shroom3	2.8	0.0250	Y
NOP58	Nucleolar protein 58	2.7	0.0009	Y
LUC7L2	Putative RNA-binding protein Luc7-like 2	2.7	0.0297	
CPSF7	Cleavage and polyadenylation specificity factor subunit 7	2.7	0.0252	Y
FBL	rRNA 2-O-methyltransferase fibrillarin	2.7	0.0205	Y
U2AF1	Splicing factor U2AF 35 kDa subunit	2.7	0.0069	Y

INTS7	Integrator complex subunit 7	2.7	0.0134	
KHSRP	Far upstream element-binding protein 2	2.6	0.0075	Y
ELAVL1	ELAV-like protein 1	2.6	0.0500	Y
IMP3	U3 small nucleolar ribonucleoprotein protein IMP3	2.6	0.0067	
PTDSS1	Phosphatidylserine synthase 1	2.6	0.0126	Y
FUBP3	Far upstream element-binding protein 3	2.6	0.0127	Y
RPS4Y1	40S ribosomal protein S4, Y isoform 1	2.6	0.0024	Y
GAR1	H/ACA ribonucleoprotein complex subunit 1	2.5	0.0021	Y
WDR12	Ribosome biogenesis protein WDR12	2.5	0.0142	
CTBP1	C-terminal-binding protein 1	2.5	0.0046	
FUBP1	Far upstream element-binding protein 1	2.5	0.0183	Y
AAAS	Aladin	2.4	0.0295	
PRPSAP2	Phosphoribosyl pyrophosphate synthase-associated protein 2	2.4	0.0246	
EIF6	Eukaryotic translation initiation factor 6	2.4	0.0234	
NOL11	Nucleolar protein 11	2.4	0.0480	
NLE1	Notchless protein homolog 1	2.4	0.0430	
HEATR1	HEAT repeat-containing protein 1	2.3	0.0468	
U2AF2	Splicing factor U2AF 65 kDa subunit	2.3	0.0342	Y
CIRBP	Cold-inducible RNA-binding protein	2.3	0.0231	
RPL18A	60S ribosomal protein L18a	2.3	0.0346	
PNO1	RNA-binding protein PNO1	2.3	0.0079	
CPSF6	Cleavage and polyadenylation specificity factor subunit 6	2.3	0.0404	Y
LYAR	Cell growth-regulating nucleolar protein	2.2	0.0350	
MYBBP1A	Myb-binding protein 1A	2.2	0.0235	Y
RCC2	Protein RCC2	2.2	0.0025	Y
CYR61	Protein CYR61	2.2	0.0109	
TARDBP	TAR DNA-binding protein 43	2.2	0.0036	Y
PWP1	Periodic tryptophan protein 1 homolog	2.2	0.0020	
NOP56	Nucleolar protein 56	2.2	0.0347	Y
EMILIN1	EMILIN-1	2.1	0.0172	
CCAR2	Cell cycle and apoptosis regulator protein 2	2.1	0.0072	
SKIV2L2	Superkiller viralicidic activity 2-like 2	2.1	0.0156	Y
LUC7L	Putative RNA-binding protein Luc7-like 1	2.0	0.0295	
NEXN	Nexilin	2.0	0.0437	
TIA1	Nucleolysin TIA-1 isoform p40	2.0	0.0147	Y
MRTO4	mRNA turnover protein 4 homolog	2.0	0.0028	
SRRT	Serrate RNA effector molecule homolog	2.0	0.0397	
NOL9	Polynucleotide 5-hydroxyl-kinase NOL9	1.9	0.0465	
TSPO	Translocator protein	1.9	0.0383	
RBM22	Pre-mRNA-splicing factor RBM22	1.9	0.0494	
PTBP1	Polypyrimidine tract-binding protein 1	1.9	0.0283	Y
EXOSC3	Exosome complex component RRP40	1.9	0.0029	
TGFBI	Transforming growth factor-beta-induced protein ig-h3	1.9	0.0144	
PRPF40A	Pre-mRNA-processing factor 40 homolog A	1.9	0.0165	
SRI	Sorcin	1.9	0.0179	
RBM12	RNA-binding protein 12	1.9	0.0232	
RCC1	Regulator of chromosome condensation	1.8	0.0080	Y
ZNF207	BUB3-interacting and GLEBS motif-containing protein ZNF207	1.8	0.0202	
EXOSC9	Exosome complex component RRP45	1.8	0.0008	
TRIP6	Thyroid receptor-interacting protein 6	1.7	0.0374	

DHX15	Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	1.7	0.0377	Y
NUP160	Nuclear pore complex protein Nup160	1.7	0.0318	
ELOVL5	Elongation of very long chain fatty acids protein 5	1.7	0.0391	
CTGF	Connective tissue growth factor	1.7	0.0124	
ZC3H18	Zinc finger CCCH domain-containing protein 18	1.6	0.0077	
DNAJC8	DnaJ homolog subfamily C member 8	1.6	0.0204	
NHP2	H/ACA ribonucleoprotein complex subunit 2	1.6	0.0399	
EDF1	Endothelial differentiation-related factor 1	1.6	0.0398	
PRMT1	Protein arginine N-methyltransferase 1	1.6	0.0197	Y
TOP1	DNA topoisomerase 1	1.6	0.0064	Y
WDR18	WD repeat-containing protein 18	1.6	0.0394	
DDX3Y	ATP-dependent RNA helicase DDX3Y	1.6	0.0374	
DKC1	H/ACA ribonucleoprotein complex subunit 4	1.6	0.0223	Y
RRP12	RRP12-like protein	1.5	0.0448	
VPS26B	Vacuolar protein sorting-associated protein 26B	1.5	0.0366	
EXOSC6	Exosome complex component MTR3	1.5	0.0025	
SPARC	SPARC	1.5	0.0113	
CCDC50	Coiled-coil domain-containing protein 50	1.5	0.0178	
PLRG1	Pleiotropic regulator 1	1.5	0.0492	
SUPT16H	FACT complex subunit SPT16	1.4	0.0171	Y
LIMS1	LIM and senescent cell antigen-like-containing domain protein 1	1.4	0.0001	
SPATS2L	SPATS2-like protein	1.4	0.0411	
DHCR24	Delta(24)-sterol reductase	1.4	0.0111	
EXOSC2	Exosome complex component RRP4	1.3	0.0238	
METAP1	Methionine aminopeptidase 1	1.3	0.0136	
GATAD2A	Transcriptional repressor p66-alpha	1.3	0.0347	
CROCC	Rootletin	1.3	0.0206	
CCAR1	Cell division cycle and apoptosis regulator protein 1	1.3	0.0144	
SART1	U4/U6.U5 tri-snRNP-associated protein 1	1.3	0.0168	
RPA1	Replication protein A 70 kDa DNA-binding subunit	1.2	0.0151	
DDX1	ATP-dependent RNA helicase DDX1	1.2	0.0189	

Table 4-3. List of heterochromatin proteins.

Shown are the 172 proteins enriched in the strucHC + H3K9me3 IP sample, compared to the euchromatin fraction. The fold-enrichement column is computed by the relative change in protein rank between the samples.

gene symbol	protein	fold-enrichment in strucHC + K9me3 vs euchromatin	t-test
RBM14	RNA-binding protein 14	28.9	0.0273
RBMX	RNA-binding motif protein, X chromosome	19.6	0.0082
KHDRBS3	KH domain-containing, RNA-binding, signal transduction-associated protein 3	19.3	0.0399
VCAN	Versican core protein	18.7	0.0366
KHDRBS1	KH domain-containing, RNA-binding, signal transduction-associated protein 1	18.6	0.0000
NONO	Non-POU domain-containing octamer-binding protein	15.8	0.0143
DDX5	Probable ATP-dependent RNA helicase DDX5	15.5	0.0347
DCD	Dermcidin	15.1	0.0002
SFPQ	Splicing factor, proline- and glutamine-rich	14.8	0.0037
S100A9	Protein S100-A9	14.1	0.0406
DDX17	Probable ATP-dependent RNA helicase DDX17	14.0	0.0288
EWSR1	RNA-binding protein EWS	13.6	0.0005
HIST1H1A	Histone H1.1	13.4	0.0060
SRSF9	Serine/arginine-rich splicing factor 9	13.1	0.0371
FUS	RNA-binding protein FUS	12.8	0.0402
FYTTD1	UAP56-interacting factor	12.8	0.0420
TAF15	TATA-binding protein-associated factor 2N	12.2	0.0108
S100A8	Protein S100-A8	12.1	0.0406
SHROOM3	Protein Shroom3	11.9	0.0113
ZNF326	DBIRD complex subunit ZNF326	11.4	0.0429
PVRL2	Nectin-2	11.3	0.0430
HNRNPM	Heterogeneous nuclear ribonucleoprotein M	11.0	0.0013
HNRNPH3	Heterogeneous nuclear ribonucleoprotein H3	10.9	0.0004
POLDIP3	Polymerase delta-interacting protein 3	10.8	0.0158
JUP	Junction plakoglobin	10.3	0.0475
DSG1	Desmoglein-1	10.2	0.0436
NCOA5	Nuclear receptor coactivator 5	10.0	0.0438
HNRNPK	Heterogeneous nuclear ribonucleoprotein K	9.9	0.0108
BRIX1	Ribosome biogenesis protein BRX1 homolog	9.5	0.0165
EBNA1BP2	Probable rRNA-processing protein EBP2	9.5	0.0406
RPF2	Ribosome production factor 2 homolog	9.5	0.0213
HNRNPH2	Heterogeneous nuclear ribonucleoprotein H2	9.3	0.0471
SERPINB12	Serpin B12	9.2	0.0442
H1FX	Histone H1x	9.2	0.0370
HNRNPH1	Heterogeneous nuclear ribonucleoprotein H	9.1	0.0009
ZNF438	Zinc finger protein 438	9.0	0.0272
KPRP	Keratinocyte proline-rich protein	9.0	0.0447
HNRNPR	Heterogeneous nuclear ribonucleoprotein R	9.0	0.0322
PPAN	Suppressor of SWI4 1 homolog	8.3	0.0456

NOP2	Probable 28S rRNA (cytosine(4447)-C(5))-methyltransferase	8.3	0.0200
HNRNPL	Heterogeneous nuclear ribonucleoprotein L	8.2	0.0035
HNRNPUL1	Heterogeneous nuclear ribonucleoprotein U-like protein 1	8.2	0.0012
H1F0	Histone H1.0	8.2	0.0031
HNRNPA2B1	Heterogeneous nuclear ribonucleoproteins A2/B1	7.8	0.0242
SRSF10	Serine/arginine-rich splicing factor 10	7.6	0.0206
RRS1	Ribosome biogenesis regulatory protein homolog	7.6	0.0390
BLMH	Bleomycin hydrolase	7.5	0.0423
GNL3	Guanine nucleotide-binding protein-like 3	7.5	0.0152
MATR3	Matrin-3	7.4	0.0130
DDX54	ATP-dependent RNA helicase DDX54	7.4	0.0409
SAFB	Scaffold attachment factor B1	7.4	0.0200
PSPC1	Paraspeckle component 1	7.3	0.0089
RAVER1	Ribonucleoprotein PTB-binding 1	7.2	0.0219
GLTSCR2	Glioma tumor suppressor candidate region gene 2 protein	7.2	0.0458
GTPBP4	Nucleolar GTP-binding protein 1	7.1	0.0001
WDR46	WD repeat-containing protein 46	7.1	0.0479
RB1CC1	RB1-inducible coiled-coil protein 1	7.1	0.0480
SBSN	Suprabasin	7.0	0.0474
FTSJ3	pre-rRNA processing protein FTSJ3	7.0	0.0478
CASP14	Caspase-14	6.9	0.0465
YLPM1	YLP motif-containing protein 1	6.9	0.0411
HNRNPDL	Heterogeneous nuclear ribonucleoprotein D-like	6.9	0.0106
RALY	RNA-binding protein Raly	6.9	0.0001
HNRNPF	Heterogeneous nuclear ribonucleoprotein F	6.8	0.0041
TGM3	Protein-glutamine gamma-glutamyltransferase E	6.8	0.0484
RBM12B	RNA-binding protein 12B	6.6	0.0475
H2AFY	Core histone macro-H2A.1	6.6	0.0058
C3	Complement C3	6.5	0.0489
CMAS	N-acylneuraminate cytidylyltransferase	6.5	0.0181
SERPINB6	Serpin B6	6.4	0.0210
AKAP8	A-kinase anchor protein 8	6.4	0.0490
MYEF2	Myelin expression factor 2	6.4	0.0492
DDX27	Probable ATP-dependent RNA helicase DDX27	6.4	0.0313
WDR36	WD repeat-containing protein 36	6.4	0.0489
HNRNPU	Heterogeneous nuclear ribonucleoprotein U	6.2	0.0089
ZC3H14	Zinc finger CCCH domain-containing protein 14	6.1	0.0486
NOC3L	Nucleolar complex protein 3 homolog	6.1	0.0366
TRA2A	Transformer-2 protein homolog alpha	6.1	0.0014
CPSF7	Cleavage and polyadenylation specificity factor subunit 7	6.1	0.0211
RBM28	RNA-binding protein 28	6.0	0.0429
SRSF1	Serine/arginine-rich splicing factor 1	6.0	0.0048
KHSRP	Far upstream element-binding protein 2	6.0	0.0068
ZNF638	Zinc finger protein 638	5.9	0.0500
PWP2	Periodic tryptophan protein 2 homolog	5.9	0.0325
PGM5	Phosphoglucomutase-like protein 5	5.8	0.0338
TRA2B	Transformer-2 protein homolog beta	5.7	0.0008
DST	Dystonin	5.7	0.0143
HNRNPLL	Heterogeneous nuclear ribonucleoprotein L-like	5.5	0.0238
PTDSS1	Phosphatidylserine synthase 1	5.5	0.0139
DDX50	ATP-dependent RNA helicase DDX50	5.3	0.0233

CHTOP	Chromatin target of PRMT1 protein	5.2	0.0037
XRN2	5-3 exoribonuclease 2	5.2	0.0291
SRSF5	Serine/arginine-rich splicing factor 5	5.1	0.0218
SYNE1	Nesprin-1	5.1	0.0262
PES1	Pescadillo homolog	5.1	0.0005
RPL14	60S ribosomal protein L14	5.0	0.0195
HNRNPA0	Heterogeneous nuclear ribonucleoprotein A0	5.0	0.0012
HNRNPC	Heterogeneous nuclear ribonucleoproteins C1/C2	5.0	0.0262
HP1BP3	Heterochromatin protein 1-binding protein 3	5.0	0.0213
MBNL1	Muscleblind-like protein 1	4.8	0.0011
RBM15	Putative RNA-binding protein 15	4.8	0.0478
WDR43	WD repeat-containing protein 43	4.7	0.0165
RCC1	Regulator of chromosome condensation	4.6	0.0020
FUBP1	Far upstream element-binding protein 1	4.6	0.0181
GAR1	H/ACA ribonucleoprotein complex subunit 1	4.6	0.0001
HNRNPUL2	Heterogeneous nuclear ribonucleoprotein U-like protein 2	4.5	0.0279
TBL3	Transducin beta-like protein 3	4.4	0.0082
RBM4	RNA-binding protein 4	4.3	0.0025
DEK	Protein DEK	4.3	0.0292
RCC2	Protein RCC2	4.3	0.0001
RPS4Y1	40S ribosomal protein S4, Y isoform 1	4.1	0.0030
SRSF6	Serine/arginine-rich splicing factor 6	4.0	0.0228
SRSF7	Serine/arginine-rich splicing factor 7	4.0	0.0166
CPSF6	Cleavage and polyadenylation specificity factor subunit 6	4.0	0.0218
SUN1	SUN domain-containing protein 1	3.9	0.0187
ELAVL1	ELAV-like protein 1	3.8	0.0288
ALYREF	THO complex subunit 4	3.8	0.0026
TOR1AIP1	Torsin-1A-interacting protein 1	3.8	0.0062
DSP	Desmoplakin	3.7	0.0064
NEMF	Nuclear export mediator factor NEMF	3.7	0.0050
DDX21	Nucleolar RNA helicase 2	3.7	0.0094
HNRNPAB	Heterogeneous nuclear ribonucleoprotein A/B	3.4	0.0469
NOP58	Nucleolar protein 58	3.3	0.0045
TIA1	Nucleolysin TIA-1 isoform p40	3.3	0.0007
FUBP3	Far upstream element-binding protein 3	3.3	0.0064
TOP1	DNA topoisomerase 1	3.1	0.0007
RBM39	RNA-binding protein 39	3.0	0.0333
SKIV2L2	Superkiller viralicidic activity 2-like 2	3.0	0.0001
MAP1A	Microtubule-associated protein 1A	2.9	0.0170
LMNB1	Lamin-B1	2.9	0.0193
RSL1D1	Ribosomal L1 domain-containing protein 1	2.9	0.0022
ILF3	Interleukin enhancer-binding factor 3	2.9	0.0139
MYBBP1A	Myb-binding protein 1A	2.8	0.0258
ARG1	Arginase-1	2.8	0.0389
STT3B	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3B	2.8	0.0033
U2AF2	Splicing factor U2AF 65 kDa subunit	2.7	0.0171
U2AF1	Splicing factor U2AF 35 kDa subunit	2.7	0.0019
PARP1	Poly [ADP-ribose] polymerase 1	2.7	0.0009
SYNCRIP	Heterogeneous nuclear ribonucleoprotein Q	2.7	0.0151
NMT1	Glycylpeptide N-tetradecanoyltransferase 1	2.6	0.0053
NOP56	Nucleolar protein 56	2.6	0.0212
SF1	Splicing factor 1	2.6	0.0099

SUPT16H	FACT complex subunit SPT16	2.6	0.0011
NUMA1	Nuclear mitotic apparatus protein 1	2.5	0.0199
HNRNPD	Heterogeneous nuclear ribonucleoprotein D0	2.5	0.0224
DHX15	Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	2.5	0.0092
RPL24	60S ribosomal protein L24	2.5	0.0019
FBL	rRNA 2-O-methyltransferase fibrillarin	2.5	0.0408
MYH14	Myosin-14	2.4	0.0261
PTBP1	Polypyrimidine tract-binding protein 1	2.4	0.0318
RECQL	ATP-dependent DNA helicase Q1	2.4	0.0025
TARDBP	TAR DNA-binding protein 43	2.4	0.0022
PRPF8	Pre-mRNA-processing-splicing factor 8	2.3	0.0287
ADAR	Double-stranded RNA-specific adenosine deaminase	2.3	0.0175
HACD3	Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 3	2.3	0.0094
BCLAF1	Bcl-2-associated transcription factor 1	2.3	0.0406
ABCF1	ATP-binding cassette sub-family F member 1	2.2	0.0265
DKC1	H/ACA ribonucleoprotein complex subunit 4	2.1	0.0157
NUDT21	Cleavage and polyadenylation specificity factor subunit 5	2.0	0.0171
FAM120A	Constitutive coactivator of PPAR-gamma-like protein 1	2.0	0.0028
HIST1H4A	Histone H4	2.0	0.0474
HIST1H2BC	Histone H2B type 1-C/E/F/G/I	1.9	0.0031
THRAP3	Thyroid hormone receptor-associated protein 3	1.9	0.0284
PHB2	Prohibitin-2	1.8	0.0028
PRMT1	Protein arginine N-methyltransferase 1	1.7	0.0404
APEX1	DNA-(apurinic or apyrimidinic site) lyase	1.7	0.0433
IFI16	Gamma-interferon-inducible protein 16	1.7	0.0108
FXR1	Fragile X mental retardation syndrome-related protein 1	1.5	0.0495
NCL	Nucleolin	1.4	0.0265
RRBP1	Ribosome-binding protein 1	1.4	0.0489
RPS13	40S ribosomal protein S13	1.3	0.0348
HDAC2	Histone deacetylase 2	1.3	0.0200

Table 4-4. Hepatic genes in repressive fibroblast H3K9me3 domains.

List of 221 hg19 Refseq genes in repressive H3K9me3 domains (strucHC or intermediate subtypes) that are expressed in human hepatocytes over fibroblasts, based on available microarray data (Huang et al., 2014). The right-most column indicates how much the gene overlaps with the domains (union of the strucHC and intermediate subtypes of H3K9me3). Genes with at least 50% overlap qualified for this list. Select genes from this list are shown in Figure 4-6B.

RefSeq ID	gene symbol	name	fraction of gene in domain
NM_000014	**A2M**	alpha-2-macroglobulin	0.77
NM_003742	**ABCB11**	ATP binding cassette subfamily B member 11	0.50
NM_178559	**ABCB5**	ATP binding cassette subfamily B member 5	0.87
NR_003087	**ABCC13**	ATP binding cassette subfamily C member 13 (pseudogene)	0.92
NM_001097	**ACR**	acrosin	1.00
NM_052956	**ACSM1**	acyl-CoA synthetase medium-chain family member 1	0.92
NM_182617	**ACSM2B**	acyl-CoA synthetase medium-chain family member 2B	1.00
NM_017888	**ACSM5**	acyl-CoA synthetase medium-chain family member 5	1.00
NM_139057	**ADAMTS17**	ADAM metallopeptidase with thrombospondin type 1 motif 17	1.00
NM_207517	**ADAMTSL3**	ADAMTS like 3	0.98
NM_153838	**ADGRF4**	adhesion G protein-coupled receptor F4	0.80
NM_005989	**AKR1D1**	aldo-keto reductase family 1 member D1	0.79
NM_000032	**ALAS2**	5'-aminolevulinate synthase 2	0.77
NM_001012421	**ANKRD20A2**	ankyrin repeat domain 20 family member A2	0.99
NM_213599	**ANO5**	anoctamin 5	1.00
NM_001639	**APCS**	amyloid P component, serum	1.00
NM_015230	**ARAP2**	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 2	1.00
NM_001185	**AZGP1**	alpha-2-glycoprotein 1, zinc-binding	0.92
NM_054025	**B3GAT1**	beta-1,3-glucuronyltransferase 1	0.98
NM_000055	**BCHE**	butyrylcholinesterase	1.00
NM_001012978	**BEX5**	brain expressed X-linked 5	1.00
NM_001721	**BMX**	BMX non-receptor tyrosine kinase	0.56
NM_001002760	**BPY2B**	basic charge, Y-linked, 2B	1.00
NM_001104629	**C4orf19**	chromosome 4 open reading frame 19	0.78
NM_000065	**C6**	complement component 6	0.93
NM_002984	**CCL4**	C-C motif chemokine ligand 4	1.00
NM_033031	**CCNB3**	cyclin B3	1.00
NM_001766	**CD1D**	CD1d molecule	0.57
NM_001772	**CD33**	CD33 molecule	1.00
NM_001778	**CD48**	CD48 molecule	0.56

NM_003874	**CD84**	CD84 molecule	0.97
NM_001768	**CD8A**	CD8a molecule	0.61
NR_003595	**CDC14C**	cell division cycle 14C, pseudogene	1.00
NM_145024	**CES5A**	carboxylesterase 5A	0.52
NM_000744	**CHRNA4**	cholinergic receptor nicotinic alpha 4 subunit	0.64
NM_031422	**CHST9**	carbohydrate sulfotransferase 9	1.00
NM_001289	**CLIC2**	chloride intracellular channel 2	0.91
NM_024734	**CLMN**	calmin	0.51
NM_152311	**CLRN3**	clarin 3	0.52
NM_014141	**CNTNAP2**	contactin associated protein-like 2	0.95
NM_006651	**CPLX1**	complexin 1	0.59
NM_153634	**CPNE8**	copine 8	1.00
NM_000567	**CRP**	C-reactive protein, pentraxin-related	1.00
NM_172249	**CSF2RA**	colony stimulating factor 2 receptor alpha subunit	0.76
NR_024387	**CXADRP2**	coxsackie virus and adenovirus receptor pseudogene 2	1.00
NR_024076	**CXADRP3**	coxsackie virus and adenovirus receptor pseudogene 3	1.00
NM_000767	**CYP2B6**	cytochrome P450 family 2 subfamily B member 6	0.63
NM_000772	**CYP2C18**	cytochrome P450 family 2 subfamily C member 18	1.00
NM_000769	**CYP2C19**	cytochrome P450 family 2 subfamily C member 19	1.00
NM_000770	**CYP2C8**	cytochrome P450 family 2 subfamily C member 8	1.00
NM_000771	**CYP2C9**	cytochrome P450 family 2 subfamily C member 9	1.00
NM_000778	**CYP4A11**	cytochrome P450 family 4 subfamily A member 11	1.00
NM_021187	**CYP4F11**	cytochrome P450 family 4 subfamily F member 11	1.00
NM_023944	**CYP4F12**	cytochrome P450 family 4 subfamily F member 12	1.00
NM_001082	**CYP4F2**	cytochrome P450 family 4 subfamily F member 2	1.00
NM_007253	**CYP4F8**	cytochrome P450 family 4 subfamily F member 8	0.83
NM_178033	**CYP4X1**	cytochrome P450 family 4 subfamily X member 1	1.00
NM_178134	**CYP4Z1**	cytochrome P450 family 4 subfamily Z member 1	0.75
NM_005218	**DEFB1**	defensin beta 1	1.00
NM_001364	**DLG2**	discs large MAGUK scaffold protein 2	0.63
NM_004010	**DMD**	dystrophin	0.70
NM_138815	**DPPA2**	developmental pluripotency associated 2	1.00
NM_024422	**DSC2**	desmocollin 2	0.59
NM_022785	**EFCAB6**	EF-hand calcium binding domain 6	0.60
NM_014800	**ELMO1**	engulfment and cell motility 1	0.65
NM_000128	**F11**	coagulation factor XI	0.81
NM_174912	**FAAH2**	fatty acid amide hydrolase 2	1.00
NM_001077710	**FAM110C**	family with sequence similarity 110 member C	1.00
NM_015381	**FAM19A5**	family with sequence similarity 19 member A5, C-C motif chemokine like	0.97
NM_030764	**FCRL2**	Fc receptor like 2	1.00
NM_022970	**FGFR2**	fibroblast growth factor receptor 2	1.00
NM_002012	**FHIT**	fragile histidine triad	0.84
NM_001014986	**FOLH1**	folate hydrolase 1	1.00
NM_004476	**FOLH1**	folate hydrolase 1	1.00
NM_153696	**FOLH1B**	folate hydrolase 1B	0.91
NM_000814	**GABRB3**	gamma-aminobutyric acid type A receptor beta3 subunit	1.00
NM_021123	**GAGE7**	G antigen 7	1.00
NM_005256	**GAS2**	growth arrest specific 2	0.60
NM_020973	**GBA3**	glucosylceramidase beta 3 (gene/pseudogene)	0.75
NM_000583	**GC**	GC, vitamin D binding protein	0.80

NM_001491	**GCNT2**	glucosaminyl (N-acetyl) transferase 2, I-branching enzyme (I blood group)	0.90
NR_003267	**GGT3P**	gamma-glutamyltransferase 3 pseudogene	1.00
NR_003503	**GGT8P**	gamma-glutamyltransferase 8 pseudogene	0.69
NM_004122	**GHSR**	growth hormone secretagogue receptor	1.00
NM_130759	**GIMAP1**	GTPase, IMAP family member 1	1.00
NM_015660	**GIMAP2**	GTPase, IMAP family member 2	1.00
NM_018326	**GIMAP4**	GTPase, IMAP family member 4	1.00
NM_018384	**GIMAP5**	GTPase, IMAP family member 5	1.00
NM_024711	**GIMAP6**	GTPase, IMAP family member 6	1.00
NM_153236	**GIMAP7**	GTPase, IMAP family member 7	1.00
NM_175571	**GIMAP8**	GTPase, IMAP family member 8	0.82
NM_144669	**GLT1D1**	glycosyltransferase 1 domain containing 1	1.00
NM_001001413	**GOLGA6L1**	golgin A6 family-like 1	1.00
NM_022036	**GPRC5C**	G protein-coupled receptor class C group 5 member C	1.00
NM_000831	**GRIK3**	glutamate ionotropic receptor kainate type subunit 3	1.00
NM_000845	**GRM8**	glutamate metabotropic receptor 8	0.98
NM_001080476	**GRXCR1**	glutaredoxin and cysteine rich domain containing 1	1.00
NM_033423	**GZMH**	granzyme H	1.00
NM_004133	**HNF4G**	hepatocyte nuclear factor 4 gamma	1.00
NM_000860	**HPGD**	hydroxyprostaglandin dehydrogenase 15-(NAD)	0.72
NM_000415	**IAPP**	islet amyloid polypeptide	0.62
NM_001002915	**IGFL2**	IGF like family member 2	0.88
NM_017416	**IL1RAPL2**	interleukin 1 receptor accessory protein like 2	0.84
NM_173178	**IL36B**	interleukin 36, beta	0.71
NM_000632	**ITGAM**	integrin subunit alpha M	0.71
NM_138693	**KLF14**	Kruppel like factor 14	1.00
NM_057162	**KLHL4**	kelch like family member 4	0.97
NM_000892	**KLKB1**	kallikrein B1	0.87
NM_152643	**KNDC1**	kinase non-catalytic C-lobe domain containing 1	0.74
NM_000226	**KRT9**	keratin 9	0.85
NM_173464	**L3MBTL4**	l(3)mbt-like 4 (Drosophila)	0.50
NM_033277	**LACRT**	lacritin	0.84
NR_003061	**LILRP2**	leukocyte immunoglobulin-like receptor pseudogene 2	1.00
NR_024090	**LINC00320**	long intergenic non-protein coding RNA 320	0.63
NR_024160	**LINC00626**	long intergenic non-protein coding RNA 626	0.95
NM_005577	**LPA**	lipoprotein(a)	0.94
NM_052886	**MAL2**	mal, T-cell differentiation protein 2 (gene/pseudogene)	0.88
NM_006770	**MARCO**	macrophage receptor with collagenous structure	0.87
NM_004994	**MMP9**	matrix metallopeptidase 9	1.00
NM_001044370	**MPPED1**	metallophosphoesterase domain containing 1	0.65
NM_005797	**MPZL2**	myelin protein zero like 2	0.92
NM_144765	**MPZL2**	myelin protein zero like 2	1.00
NM_024021	**MS4A4A**	membrane spanning 4-domains A4A	1.00
NM_021201	**MS4A7**	membrane spanning 4-domains A7	1.00
NM_001033602	**MTUS2**	microtubule associated tumor suppressor candidate 2	0.59
NM_152423	**MUM1L1**	MUM1 like 1	0.96
NM_003970	**MYOM2**	myomesin 2	0.81
NM_021963	**NAP1L2**	nucleosome assembly protein 1 like 2	1.00
NM_002500	**NEUROD1**	neuronal differentiation 1	1.00

NM_145007	**NLRP11**	NLR family pyrin domain containing 11	1.00
NM_005123	**NR1H4**	nuclear receptor subfamily 1 group H member 4	0.58
NM_001105663	**NUDT7**	nudix hydrolase 7	0.64
NM_002534	**OAS1**	2'-5'-oligoadenylate synthetase 1	1.00
NM_003553	**OR1E1**	olfactory receptor family 1 subfamily E member 1	1.00
NM_001004739	**OR5L2**	olfactory receptor family 5 subfamily L member 2	1.00
NR_002171	**OR7E156P**	olfactory receptor family 7 subfamily E member 156 pseudogene	1.00
NM_001079935	**OR7E24**	olfactory receptor family 7 subfamily E member 24	1.00
NM_207320	**OTUD6A**	OTU deubiquitinase 6A	1.00
NM_178129	**P2RY8**	purinergic receptor P2Y8	0.64
NM_130467	**PAGE5**	PAGE family member 5	1.00
NM_000277	**PAH**	phenylalanine hydroxylase	0.88
NM_022843	**PCDH20**	protocadherin 20	1.00
NM_031856	**PCDHA8**	protocadherin alpha 8	1.00
NM_031882	**PCDHAC1**	protocadherin alpha subfamily C, 1	0.75
NM_001077358	**PDE11A**	phosphodiesterase 11A	0.54
NM_016953	**PDE11A**	phosphodiesterase 11A	0.60
NM_006210	**PEG3**	paternally expressed 3	1.00
NM_002625	**PFKFB1**	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 1	0.54
NM_016112	**PKD2L1**	polycystin 2 like 1, transient receptor potential cation channel	0.66
NM_004572	**PKP2**	plakophilin 2	0.89
NM_005084	**PLA2G7**	phospholipase A2 group VII	0.65
NM_024829	**PLBD1**	phospholipase B domain containing 1	0.67
NM_000301	**PLG**	plasminogen	1.00
NM_002763	**PROX1**	prospero homeobox 1	0.76
NM_002864	**PZP**	PZP, alpha-2-macroglobulin like	0.90
NM_138453	**RAB3C**	RAB3C, member RAS oncogene family	0.72
NM_006744	**RBP4**	retinol binding protein 4	0.79
NM_021026	**RFPL1**	ret finger protein like 1	1.00
NM_005615	**RNASE6**	ribonuclease A family member k6	1.00
NM_198085	**RNF148**	ring finger protein 148	1.00
NM_025236	**RNF39**	ring finger protein 39	0.65
NM_005621	**S100A12**	S100 calcium binding protein A12	0.57
NM_022136	**SAMSN1**	SAM domain, SH3 domain and nuclear localization signals 1	1.00
NM_145168	**SDR42E1**	short chain dehydrogenase/reductase family 42E, member 1	0.66
NM_001080451	**SERPINA11**	serpin family A member 11	1.00
NM_001085	**SERPINA3**	serpin family A member 3	0.89
NM_006215	**SERPINA4**	serpin family A member 4	1.00
NM_000354	**SERPINA7**	serpin family A member 7	1.00
NM_012435	**SHC2**	SHC adaptor protein 2	0.89
NM_001010846	**SHE**	Src homology 2 domain containing E	0.71
NM_001040153	**SLAIN1**	SLAIN motif family member 1	0.60
NM_003037	**SLAMF1**	signaling lymphocytic activation molecule family member 1	0.76
NM_021181	**SLAMF7**	SLAM family member 7	1.00
NM_005835	**SLC17A2**	solute carrier family 17 member 2	0.67
NM_001039752	**SLC22A10**	solute carrier family 22 member 10	0.91
NM_021977	**SLC22A3**	solute carrier family 22 member 3	0.52
NM_016615	**SLC6A13**	solute carrier family 6 member 13	0.60

NM_006446	**SLCO1B1**	solute carrier organic anion transporter family member 1B1	0.92
NM_019844	**SLCO1B3**	solute carrier organic anion transporter family member 1B3	0.91
NM_180991	**SLCO4C1**	solute carrier organic anion transporter family member 4C1	1.00
NM_001034852	**SMOC1**	SPARC related modular calcium binding 1	0.85
NM_152989	**SOX5**	SRY-box 5	0.97
NM_153189	**SPAM1**	sperm adhesion molecule 1	1.00
NM_003122	**SPINK1**	serine peptidase inhibitor, Kazal type 1	0.94
NM_021015	**SSX5**	SSX family member 5	0.51
NM_020225	**STOX2**	storkhead box 2	0.85
NM_014465	**SULT1B1**	sulfotransferase family 1B member 1	0.78
NM_016524	**SYT17**	synaptotagmin 17	0.77
NM_016945	**TAS2R16**	taste 2 receptor member 16	1.00
NM_003225	**TFF1**	trefoil factor 1	1.00
NM_138379	**TIMD4**	T-cell immunoglobulin and mucin domain containing 4	1.00
NM_018487	**TMEM176A**	transmembrane protein 176A	1.00
NM_014020	**TMEM176B**	transmembrane protein 176B	0.89
NM_001123376	**TMEM72**	transmembrane protein 72	0.55
NM_003701	**TNFSF11**	tumor necrosis factor superfamily member 11	0.55
NM_001080430	**TOX3**	TOX high mobility group box family member 3	0.97
NM_022445	**TPK1**	thiamin pyrophosphokinase 1	0.95
NM_012471	**TRPC5**	transient receptor potential cation channel subfamily C member 5	0.92
NM_173485	**TSHZ2**	teashirt zinc finger homeobox 2	0.63
NM_178562	**TSPAN33**	tetraspanin 33	0.80
NM_004616	**TSPAN8**	tetraspanin 8	0.95
NR_026547	**TSPEAR-AS2**	TSPEAR antisense RNA 2	1.00
NM_000371	**TTR**	transthyretin	1.00
NM_006001	**TUBA3C**	tubulin alpha 3c	1.00
NM_019076	**UGT1A8**	UDP glucuronosyltransferase family 1 member A8	0.61
NM_001075	**UGT2B10**	UDP glucuronosyltransferase family 2 member B10	0.87
NM_001073	**UGT2B11**	UDP glucuronosyltransferase family 2 member B11	1.00
NM_001076	**UGT2B15**	UDP glucuronosyltransferase family 2 member B15	1.00
NM_001077	**UGT2B17**	UDP glucuronosyltransferase family 2 member B17	1.00
NM_053039	**UGT2B28**	UDP glucuronosyltransferase family 2 member B28	1.00
NM_021139	**UGT2B4**	UDP glucuronosyltransferase family 2 member B4	1.00
NM_001074	**UGT2B7**	UDP glucuronosyltransferase family 2 member B7	0.75
NM_018974	**UNC93A**	unc-93 homolog A (C. elegans)	1.00
NM_206933	**USH2A**	usherin	0.77
NM_013452	**VCX**	variable charge, X-linked	1.00
NM_022479	**WBSCR17**	Williams-Beuren syndrome chromosome region 17	1.00
NM_182758	**WDR72**	WD repeat domain 72	1.00
NM_016373	**WWOX**	WW domain containing oxidoreductase	0.89
NM_052898	**XKR4**	XK related 4	0.99
NM_003446	**ZNF157**	zinc finger protein 157	0.51
NM_001037735	**ZNF630**	zinc finger protein 630	0.69

Table 4-5. Heterochromatic genes with enhanced activation after siRNA against SUV39H1 or RBMX/L1.

List of 364 genes in strucHC domains that are significantly upregulated (adjusted $p < 0.05$) by SUV39H1 siRNA and/or RBMX/L1 siRNA, in the presence of hiHep factors. Fold-changes relative to control siRNA are shown. The data here are used in Figure 4-7C.

gene symbol	Significantly upregulated in si-SUV39H1?	Significantly upregulated in si-RBMX/L1?	Fold-change: si-SUV39H1 vs control	Fold-change: si-RBMX/L1 vs control
NR1H4	signif in si-SUV39H1	signif in si-RBMX/L1	**17.1**	**15.2**
SPINK1	signif in si-SUV39H1	signif in si-RBMX/L1	**14.2**	**11.9**
PKIB	signif in si-SUV39H1	signif in si-RBMX/L1	**11.2**	**26.0**
TSPAN8	signif in si-SUV39H1	signif in si-RBMX/L1	**8.8**	**6.0**
MYO3B	signif in si-SUV39H1	signif in si-RBMX/L1	**8.2**	**10.0**
SLAMF7	signif in si-SUV39H1	signif in si-RBMX/L1	**7.6**	**12.6**
CRP	signif in si-SUV39H1	signif in si-RBMX/L1	**7.5**	**18.3**
CLIC5	signif in si-SUV39H1	signif in si-RBMX/L1	**6.4**	**5.6**
LOC101930452	signif in si-SUV39H1	signif in si-RBMX/L1	**5.8**	**7.3**
COL6A5	signif in si-SUV39H1	signif in si-RBMX/L1	**5.8**	**3.5**
SMIM10L2A	signif in si-SUV39H1	signif in si-RBMX/L1	**5.5**	**7.2**
UPP2	signif in si-SUV39H1	signif in si-RBMX/L1	**5.0**	**6.0**
FAT2	signif in si-SUV39H1	signif in si-RBMX/L1	**5.0**	**2.4**
BEND4	signif in si-SUV39H1	signif in si-RBMX/L1	**4.7**	**3.9**
DSG2	signif in si-SUV39H1	signif in si-RBMX/L1	**4.7**	**5.8**
GLP2R	signif in si-SUV39H1	signif in si-RBMX/L1	**4.6**	**3.7**
FAM134B	signif in si-SUV39H1	signif in si-RBMX/L1	**4.2**	**3.7**
IGSF1	signif in si-SUV39H1	signif in si-RBMX/L1	**4.2**	**4.0**
MAFB	signif in si-SUV39H1	signif in si-RBMX/L1	**4.1**	**3.4**
SLC17A3	signif in si-SUV39H1	signif in si-RBMX/L1	**4.1**	**4.3**
CNGA3	signif in si-SUV39H1	signif in si-RBMX/L1	**4.0**	**1.7**
GCNT2	signif in si-SUV39H1	signif in si-RBMX/L1	**4.0**	**1.6**
ACMSD	signif in si-SUV39H1	signif in si-RBMX/L1	**4.0**	**7.8**
NOX1	signif in si-SUV39H1	signif in si-RBMX/L1	**3.9**	**3.4**
NUGGC	signif in si-SUV39H1	signif in si-RBMX/L1	**3.8**	**2.4**
UCA1	signif in si-SUV39H1	signif in si-RBMX/L1	**3.6**	**4.9**
BRINP2	signif in si-SUV39H1	signif in si-RBMX/L1	**3.3**	**3.7**
MOB3B	signif in si-SUV39H1	signif in si-RBMX/L1	**3.3**	**4.2**
PADI2	signif in si-SUV39H1	signif in si-RBMX/L1	**3.3**	**3.9**
MIR3945HG	signif in si-SUV39H1	signif in si-RBMX/L1	**3.1**	**3.3**
DOCK8	signif in si-SUV39H1	signif in si-RBMX/L1	**3.0**	**2.2**
SERPINI1	signif in si-SUV39H1	signif in si-RBMX/L1	**3.0**	**1.6**
RAPGEF5	signif in si-SUV39H1	signif in si-RBMX/L1	**3.0**	**1.8**
NEBL	signif in si-SUV39H1	signif in si-RBMX/L1	**3.0**	**3.9**
VNN1	signif in si-SUV39H1	signif in si-RBMX/L1	**2.9**	**3.8**
GCNT3	signif in si-SUV39H1	signif in si-RBMX/L1	**2.9**	**2.4**
CCND2	signif in si-SUV39H1	signif in si-RBMX/L1	**2.8**	**5.4**
CHD7	signif in si-SUV39H1	signif in si-RBMX/L1	**2.8**	**2.2**
PIK3AP1	signif in si-SUV39H1	signif in si-RBMX/L1	**2.8**	**2.6**
ANO2	signif in si-SUV39H1	signif in si-RBMX/L1	**2.8**	**2.3**

AKR1C3	signif in si-SUV39H1	signif in si-RBMX/L1	**2.8**	**1.7**
LOC286297	signif in si-SUV39H1	signif in si-RBMX/L1	**2.7**	**2.7**
LIPH	signif in si-SUV39H1	signif in si-RBMX/L1	**2.7**	**1.8**
MPZL3	signif in si-SUV39H1	signif in si-RBMX/L1	**2.7**	**3.3**
LOC100133920	signif in si-SUV39H1	signif in si-RBMX/L1	**2.7**	**2.8**
PDZD2	signif in si-SUV39H1	signif in si-RBMX/L1	**2.6**	**1.7**
ELOVL7	signif in si-SUV39H1	signif in si-RBMX/L1	**2.6**	**3.4**
PLA1A	signif in si-SUV39H1	signif in si-RBMX/L1	**2.6**	**2.0**
GOLGA8M	signif in si-SUV39H1	signif in si-RBMX/L1	**2.6**	**1.6**
UGT1A7	signif in si-SUV39H1	signif in si-RBMX/L1	**2.6**	**1.8**
IQGAP2	signif in si-SUV39H1	signif in si-RBMX/L1	**2.6**	**3.2**
DDX11L16	signif in si-SUV39H1	signif in si-RBMX/L1	**2.6**	**3.8**
TMEM27	signif in si-SUV39H1	signif in si-RBMX/L1	**2.6**	**2.4**
FAM3D	signif in si-SUV39H1	signif in si-RBMX/L1	**2.5**	**2.3**
HRG	signif in si-SUV39H1	signif in si-RBMX/L1	**2.5**	**2.9**
ULBP3	signif in si-SUV39H1	signif in si-RBMX/L1	**2.5**	**1.8**
SORBS1	signif in si-SUV39H1	signif in si-RBMX/L1	**2.5**	**2.5**
WIPF3	signif in si-SUV39H1	signif in si-RBMX/L1	**2.4**	**3.1**
DCX	signif in si-SUV39H1	signif in si-RBMX/L1	**2.4**	**18.5**
FER1L6	signif in si-SUV39H1	signif in si-RBMX/L1	**2.4**	**1.8**
EFHD1	signif in si-SUV39H1	signif in si-RBMX/L1	**2.4**	**1.7**
UGT1A8	signif in si-SUV39H1	signif in si-RBMX/L1	**2.4**	**1.8**
UGT1A9	signif in si-SUV39H1	signif in si-RBMX/L1	**2.4**	**1.8**
DDX11L2	signif in si-SUV39H1	signif in si-RBMX/L1	**2.3**	**3.1**
TEKT4P2	signif in si-SUV39H1	signif in si-RBMX/L1	**2.3**	**2.7**
NLRC4	signif in si-SUV39H1	signif in si-RBMX/L1	**2.3**	**2.3**
GOLGA8J	signif in si-SUV39H1	signif in si-RBMX/L1	**2.3**	**1.6**
GOLGA8K	signif in si-SUV39H1	signif in si-RBMX/L1	**2.3**	**1.7**
SPTBN2	signif in si-SUV39H1	signif in si-RBMX/L1	**2.2**	**1.9**
PTGFRN	signif in si-SUV39H1	signif in si-RBMX/L1	**2.2**	**2.1**
ZFAND2A	signif in si-SUV39H1	signif in si-RBMX/L1	**2.2**	**1.4**
ABCC11	signif in si-SUV39H1	signif in si-RBMX/L1	**2.2**	**2.1**
PCSK5	signif in si-SUV39H1	signif in si-RBMX/L1	**2.1**	**2.0**
KLHL4	signif in si-SUV39H1	signif in si-RBMX/L1	**2.1**	**3.0**
TGM5	signif in si-SUV39H1	signif in si-RBMX/L1	**2.1**	**2.5**
PROS1	signif in si-SUV39H1	signif in si-RBMX/L1	**2.0**	**1.7**
APOD	signif in si-SUV39H1	signif in si-RBMX/L1	**2.0**	**2.3**
ALDOB	signif in si-SUV39H1	signif in si-RBMX/L1	**2.0**	**1.3**
HNF4A	signif in si-SUV39H1	signif in si-RBMX/L1	**2.0**	**1.4**
SULT1C2	signif in si-SUV39H1	signif in si-RBMX/L1	**1.9**	**1.7**
KIAA1161	signif in si-SUV39H1	signif in si-RBMX/L1	**1.9**	**1.7**
RNF44	signif in si-SUV39H1	signif in si-RBMX/L1	**1.9**	**1.4**
MFAP3L	signif in si-SUV39H1	signif in si-RBMX/L1	**1.9**	**1.5**
RAB6B	signif in si-SUV39H1	signif in si-RBMX/L1	**1.8**	**1.8**
FOXA3	signif in si-SUV39H1	signif in si-RBMX/L1	**1.8**	**1.4**
RASGRF2	signif in si-SUV39H1	signif in si-RBMX/L1	**1.8**	**1.5**
AQP7P1	signif in si-SUV39H1	signif in si-RBMX/L1	**1.8**	**1.6**
ACAA1	signif in si-SUV39H1	signif in si-RBMX/L1	**1.7**	**1.2**
HNF1A	signif in si-SUV39H1	signif in si-RBMX/L1	**1.7**	**1.4**
GLIDR	signif in si-SUV39H1	signif in si-RBMX/L1	**1.6**	**1.6**
PCDH1	signif in si-SUV39H1	signif in si-RBMX/L1	**1.6**	**2.4**
FUT6	signif in si-SUV39H1	signif in si-RBMX/L1	**1.6**	**2.0**
SLC6A20	signif in si-SUV39H1	signif in si-RBMX/L1	**1.6**	**2.1**
GOLGA8S	signif in si-SUV39H1	signif in si-RBMX/L1	**1.6**	**1.6**

ZNF578	signif in si-SUV39H1	signif in si-RBMX/L1	**1.6**	**1.5**
NATD1	signif in si-SUV39H1	signif in si-RBMX/L1	**1.5**	**1.5**
PRODH2	signif in si-SUV39H1	signif in si-RBMX/L1	**1.5**	**1.6**
SERPINC1	signif in si-SUV39H1	signif in si-RBMX/L1	**1.5**	**1.6**
DYSF	signif in si-SUV39H1	signif in si-RBMX/L1	**1.4**	**2.1**
ZBTB7B	signif in si-SUV39H1	signif in si-RBMX/L1	**1.4**	**1.5**
GPR108	signif in si-SUV39H1	signif in si-RBMX/L1	**1.3**	**1.4**
LINC00504	signif in si-SUV39H1	signif in si-RBMX/L1	**1.3**	**1.3**
TREML3P	signif in si-SUV39H1		**28.2**	**1.0**
STOX2	signif in si-SUV39H1		**10.1**	**7.4**
TESPA1	signif in si-SUV39H1		**9.9**	**24.5**
SYT3	signif in si-SUV39H1		**9.0**	**1.2**
TMEM71	signif in si-SUV39H1		**8.2**	**2.5**
TNMD	signif in si-SUV39H1		**8.1**	**3.7**
MYH16	signif in si-SUV39H1		**7.5**	**2.8**
LOC102723769	signif in si-SUV39H1		**7.1**	**6.7**
SLC30A2	signif in si-SUV39H1		**6.9**	**5.1**
RRAGD	signif in si-SUV39H1		**6.4**	**3.3**
NPPB	signif in si-SUV39H1		**6.0**	**1.2**
FATE1	signif in si-SUV39H1		**5.9**	**2.3**
FPR3	signif in si-SUV39H1		**5.6**	**3.0**
TRIM67	signif in si-SUV39H1		**5.6**	**2.5**
TNFRSF11A	signif in si-SUV39H1		**5.2**	**4.8**
PADI3	signif in si-SUV39H1		**5.0**	**2.3**
DHRS2	signif in si-SUV39H1		**4.9**	**0.5**
DRGX	signif in si-SUV39H1		**4.8**	**3.2**
LINC01234	signif in si-SUV39H1		**4.7**	**2.8**
CLSTN2	signif in si-SUV39H1		**4.6**	**4.2**
COBL	signif in si-SUV39H1		**4.6**	**1.9**
LRRC31	signif in si-SUV39H1		**4.5**	**1.6**
ATP8A1	signif in si-SUV39H1		**4.4**	**2.0**
ASPA	signif in si-SUV39H1		**4.2**	**1.3**
LYPD1	signif in si-SUV39H1		**3.9**	**1.5**
CYP2B6	signif in si-SUV39H1		**3.8**	**1.2**
CLMN	signif in si-SUV39H1		**3.8**	**1.3**
KLHL6	signif in si-SUV39H1		**3.8**	**2.8**
CHDH	signif in si-SUV39H1		**3.7**	**1.3**
ERICH2	signif in si-SUV39H1		**3.6**	**2.2**
AMN	signif in si-SUV39H1		**3.6**	**1.5**
CPEB1	signif in si-SUV39H1		**3.6**	**1.9**
AKR1C4	signif in si-SUV39H1		**3.4**	**2.5**
TBXAS1	signif in si-SUV39H1		**3.4**	**1.1**
HABP2	signif in si-SUV39H1		**3.4**	**1.2**
FGF14	signif in si-SUV39H1		**3.3**	**1.2**
PDGFB	signif in si-SUV39H1		**3.3**	**1.1**
CYP4F3	signif in si-SUV39H1		**3.2**	**1.6**
G6PC	signif in si-SUV39H1		**3.2**	**1.0**
LAMP3	signif in si-SUV39H1		**3.2**	**1.2**
GBA3	signif in si-SUV39H1		**3.1**	**1.5**
MYO1H	signif in si-SUV39H1		**3.1**	**2.5**
IL18R1	signif in si-SUV39H1		**3.1**	**2.1**
UNC13D	signif in si-SUV39H1		**3.0**	**1.1**
ACVR1C	signif in si-SUV39H1		**3.0**	**2.3**
AOC3	signif in si-SUV39H1		**3.0**	**1.4**

CRCT1	signif in si-SUV39H1		**3.0**	**2.3**
TSPAN33	signif in si-SUV39H1		**2.9**	**1.9**
ALDH1L1	signif in si-SUV39H1		**2.9**	**0.5**
CDH26	signif in si-SUV39H1		**2.9**	**1.2**
IFITM10	signif in si-SUV39H1		**2.9**	**1.2**
ATP1A2	signif in si-SUV39H1		**2.9**	**1.5**
KEL	signif in si-SUV39H1		**2.8**	**0.3**
SPIRE2	signif in si-SUV39H1		**2.8**	**1.1**
SLC7A9	signif in si-SUV39H1		**2.8**	**2.7**
PANX2	signif in si-SUV39H1		**2.8**	**0.4**
SDK1	signif in si-SUV39H1		**2.8**	**1.1**
NKAIN1	signif in si-SUV39H1		**2.7**	**1.5**
C1orf106	signif in si-SUV39H1		**2.6**	**1.6**
BMP8B	signif in si-SUV39H1		**2.6**	**1.5**
GOLGA8IP	signif in si-SUV39H1		**2.6**	**1.5**
NR1I3	signif in si-SUV39H1		**2.5**	**1.9**
BMP8A	signif in si-SUV39H1		**2.5**	**2.0**
MOGAT3	signif in si-SUV39H1		**2.5**	**1.5**
MTSS1	signif in si-SUV39H1		**2.5**	**1.6**
ZNF208	signif in si-SUV39H1		**2.5**	**1.2**
CYP4F12	signif in si-SUV39H1		**2.5**	**1.9**
PMEPA1	signif in si-SUV39H1		**2.5**	**1.4**
ZMYND15	signif in si-SUV39H1		**2.5**	**1.0**
RAET1K	signif in si-SUV39H1		**2.4**	**1.3**
IL17RB	signif in si-SUV39H1		**2.4**	**0.7**
CCDC64	signif in si-SUV39H1		**2.4**	**1.9**
DISP2	signif in si-SUV39H1		**2.4**	**1.5**
CACFD1	signif in si-SUV39H1		**2.4**	**1.2**
SERPINA6	signif in si-SUV39H1		**2.4**	**0.9**
GPX3	signif in si-SUV39H1		**2.4**	**1.0**
SPINT1	signif in si-SUV39H1		**2.3**	**1.0**
KIAA0040	signif in si-SUV39H1		**2.3**	**1.2**
AZGP1	signif in si-SUV39H1		**2.3**	**1.4**
URAHP	signif in si-SUV39H1		**2.3**	**1.6**
PAPPA2	signif in si-SUV39H1		**2.3**	**1.3**
TMEM37	signif in si-SUV39H1		**2.3**	**1.0**
ZBTB46	signif in si-SUV39H1		**2.3**	**1.3**
UGT1A6	signif in si-SUV39H1		**2.2**	**1.2**
GOLGA8T	signif in si-SUV39H1		**2.2**	**1.6**
SERPINA1	signif in si-SUV39H1		**2.2**	**1.1**
ADCK3	signif in si-SUV39H1		**2.2**	**1.3**
TRAPPC6A	signif in si-SUV39H1		**2.1**	**1.6**
CLU	signif in si-SUV39H1		**2.1**	**1.2**
ALS2CL	signif in si-SUV39H1		**2.1**	**1.6**
SLC6A13	signif in si-SUV39H1		**2.1**	**1.3**
GOLGA8H	signif in si-SUV39H1		**2.1**	**1.4**
PCSK6	signif in si-SUV39H1		**2.1**	**1.4**
TM6SF2	signif in si-SUV39H1		**2.1**	**1.1**
CTSD	signif in si-SUV39H1		**2.1**	**1.3**
SLC26A11	signif in si-SUV39H1		**2.1**	**1.1**
GLYAT	signif in si-SUV39H1		**2.0**	**2.0**
NEU1	signif in si-SUV39H1		**2.0**	**1.4**
TMEM133	signif in si-SUV39H1		**2.0**	**1.4**
GOLGA2P5	signif in si-SUV39H1		**2.0**	**0.7**

DNAH5	signif in si-SUV39H1		**2.0**	**1.0**
FBP1	signif in si-SUV39H1		**2.0**	**1.8**
BCL2	signif in si-SUV39H1		**2.0**	**1.3**
CAPN8	signif in si-SUV39H1		**1.9**	**0.6**
GOLGA8R	signif in si-SUV39H1		**1.9**	**1.3**
CTSF	signif in si-SUV39H1		**1.9**	**1.2**
LAD1	signif in si-SUV39H1		**1.9**	**1.4**
ZNF469	signif in si-SUV39H1		**1.9**	**1.1**
SH2D5	signif in si-SUV39H1		**1.9**	**1.5**
ABTB1	signif in si-SUV39H1		**1.9**	**1.0**
MELTF	signif in si-SUV39H1		**1.9**	**1.1**
SLC13A2	signif in si-SUV39H1		**1.9**	**1.4**
DAGLA	signif in si-SUV39H1		**1.9**	**1.2**
GOLGA8O	signif in si-SUV39H1		**1.9**	**1.3**
CAPN3	signif in si-SUV39H1		**1.8**	**0.7**
LONRF3	signif in si-SUV39H1		**1.8**	**1.6**
SULF2	signif in si-SUV39H1		**1.8**	**1.3**
LOC101928841	signif in si-SUV39H1		**1.8**	**1.7**
GOLGA8N	signif in si-SUV39H1		**1.8**	**1.3**
GSTZ1	signif in si-SUV39H1		**1.8**	**1.1**
AQP7P3	signif in si-SUV39H1		**1.8**	**1.2**
GALNT18	signif in si-SUV39H1		**1.8**	**1.0**
PDE11A	signif in si-SUV39H1		**1.8**	**1.1**
C1orf115	signif in si-SUV39H1		**1.8**	**1.5**
HAGH	signif in si-SUV39H1		**1.8**	**1.0**
CAMTA1	signif in si-SUV39H1		**1.8**	**1.3**
C12orf56	signif in si-SUV39H1		**1.7**	**1.4**
TBKBP1	signif in si-SUV39H1		**1.7**	**1.4**
ULBP1	signif in si-SUV39H1		**1.7**	**1.3**
FAM195B	signif in si-SUV39H1		**1.7**	**1.0**
RBPMS2	signif in si-SUV39H1		**1.7**	**1.2**
SERPINA3	signif in si-SUV39H1		**1.7**	**1.0**
IDUA	signif in si-SUV39H1		**1.7**	**0.9**
STARD8	signif in si-SUV39H1		**1.7**	**1.4**
SLC22A18	signif in si-SUV39H1		**1.7**	**1.1**
FGFR4	signif in si-SUV39H1		**1.7**	**1.4**
POMGNT2	signif in si-SUV39H1		**1.7**	**1.5**
OVGP1	signif in si-SUV39H1		**1.6**	**1.0**
HS1BP3	signif in si-SUV39H1		**1.6**	**1.3**
NAPRT	signif in si-SUV39H1		**1.6**	**1.0**
CLN3	signif in si-SUV39H1		**1.6**	**1.4**
PLEKHA7	signif in si-SUV39H1		**1.6**	**1.1**
GRN	signif in si-SUV39H1		**1.6**	**1.2**
PQLC2	signif in si-SUV39H1		**1.6**	**1.4**
FAM213A	signif in si-SUV39H1		**1.5**	**1.3**
ZNF512B	signif in si-SUV39H1		**1.5**	**1.1**
PRADC1	signif in si-SUV39H1		**1.5**	**1.0**
TRPV1	signif in si-SUV39H1		**1.5**	**0.8**
TMEM63B	signif in si-SUV39H1		**1.5**	**1.2**
GNPTG	signif in si-SUV39H1		**1.5**	**0.9**
UVSSA	signif in si-SUV39H1		**1.5**	**1.1**
CNNM3	signif in si-SUV39H1		**1.5**	**1.1**
PCK1	signif in si-SUV39H1		**1.5**	**0.5**
OSBPL2	signif in si-SUV39H1		**1.5**	**1.1**

ECI1	signif in si-SUV39H1		**1.5**	**1.2**
GAA	signif in si-SUV39H1		**1.5**	**1.0**
AGT	signif in si-SUV39H1		**1.5**	**1.3**
TOLLIP	signif in si-SUV39H1		**1.5**	**1.1**
TEX264	signif in si-SUV39H1		**1.4**	**1.1**
OTUB2	signif in si-SUV39H1		**1.4**	**1.0**
GET4	signif in si-SUV39H1		**1.4**	**1.0**
CLCN7	signif in si-SUV39H1		**1.4**	**1.0**
CLPTM1L	signif in si-SUV39H1		**1.4**	**1.0**
GOLGA6L5P	signif in si-SUV39H1		**1.4**	**1.0**
SLC26A6	signif in si-SUV39H1		**1.4**	**1.1**
CALCOCO1	signif in si-SUV39H1		**1.4**	**1.1**
RNF220	signif in si-SUV39H1		**1.3**	**1.1**
PLOD3	signif in si-SUV39H1		**1.3**	**1.0**
SLC38A10	signif in si-SUV39H1		**1.3**	**1.1**
LMAN2	signif in si-SUV39H1		**1.3**	**0.9**
WWP2	signif in si-SUV39H1		**1.3**	**1.0**
HECTD3	signif in si-SUV39H1		**1.3**	**1.0**
LRPAP1	signif in si-SUV39H1		**1.3**	**0.9**
IL17RA	signif in si-SUV39H1		**1.3**	**1.2**
TSPAN9	signif in si-SUV39H1		**1.3**	**1.0**
HDAC6	signif in si-SUV39H1		**1.3**	**1.1**
IQSEC1	signif in si-SUV39H1		**1.3**	**1.0**
TBC1D13	signif in si-SUV39H1		**1.2**	**1.1**
DLGAP4	signif in si-SUV39H1		**1.2**	**0.9**
S100A9		signif in si-RBMX/L1	**1.1**	**24.3**
LINC01559		signif in si-RBMX/L1	**4.5**	**17.2**
ADRA2C		signif in si-RBMX/L1	**2.4**	**11.0**
SPANXN3		signif in si-RBMX/L1	**3.9**	**10.9**
PRSS8		signif in si-RBMX/L1	**4.0**	**8.8**
DUOX1		signif in si-RBMX/L1	**0.7**	**8.3**
NEFL		signif in si-RBMX/L1	**2.3**	**8.2**
CES1		signif in si-RBMX/L1	**1.5**	**7.2**
PRODH		signif in si-RBMX/L1	**3.3**	**6.9**
SLC16A14		signif in si-RBMX/L1	**2.6**	**6.1**
SMIM2-AS1		signif in si-RBMX/L1	**1.6**	**5.5**
XIRP1		signif in si-RBMX/L1	**1.4**	**5.2**
GJA5		signif in si-RBMX/L1	**2.0**	**5.1**
ABCG1		signif in si-RBMX/L1	**0.8**	**5.1**
LRP2		signif in si-RBMX/L1	**1.6**	**5.0**
KSR2		signif in si-RBMX/L1	**1.6**	**5.0**
C2orf72		signif in si-RBMX/L1	**1.8**	**4.8**
DDX11L5		signif in si-RBMX/L1	**3.3**	**4.5**
ADGRG4		signif in si-RBMX/L1	**1.9**	**4.5**
KLHDC7A		signif in si-RBMX/L1	**3.4**	**4.2**
CALCA		signif in si-RBMX/L1	**1.4**	**4.1**
EBI3		signif in si-RBMX/L1	**2.0**	**3.9**
VSTM2L		signif in si-RBMX/L1	**0.9**	**3.9**
KCNQ3		signif in si-RBMX/L1	**1.7**	**3.9**
PEG3		signif in si-RBMX/L1	**1.5**	**3.8**
DNER		signif in si-RBMX/L1	**1.8**	**3.8**
AQP7		signif in si-RBMX/L1	**2.9**	**3.8**
MRO		signif in si-RBMX/L1	**1.5**	**3.6**
CLRN3		signif in si-RBMX/L1	**1.8**	**3.6**

MYH14		signif in si-RBMX/L1	**1.6**	**3.4**
TNFSF9		signif in si-RBMX/L1	**2.1**	**3.3**
RNF125		signif in si-RBMX/L1	**1.8**	**3.1**
FAM84B		signif in si-RBMX/L1	**2.1**	**3.0**
MYO5B		signif in si-RBMX/L1	**1.8**	**2.9**
TFCP2L1		signif in si-RBMX/L1	**1.8**	**2.9**
SCN3B		signif in si-RBMX/L1	**1.6**	**2.8**
SERPINF2		signif in si-RBMX/L1	**1.0**	**2.8**
TMEM178B		signif in si-RBMX/L1	**1.5**	**2.7**
SLC7A2		signif in si-RBMX/L1	**1.0**	**2.6**
FGB		signif in si-RBMX/L1	**0.4**	**2.4**
LAMB3		signif in si-RBMX/L1	**1.3**	**2.4**
NEB		signif in si-RBMX/L1	**1.6**	**2.4**
FRMD3		signif in si-RBMX/L1	**0.8**	**2.3**
FLVCR2		signif in si-RBMX/L1	**1.2**	**2.3**
TAT		signif in si-RBMX/L1	**1.3**	**2.2**
PRSS3		signif in si-RBMX/L1	**0.8**	**2.2**
PTPRH		signif in si-RBMX/L1	**1.4**	**2.2**
APOB		signif in si-RBMX/L1	**1.1**	**2.1**
GOLGA8DP		signif in si-RBMX/L1	**1.3**	**2.1**
GOLGA8EP		signif in si-RBMX/L1	**1.3**	**2.1**
PTPRD		signif in si-RBMX/L1	**1.2**	**2.1**
OLFM4		signif in si-RBMX/L1	**0.5**	**2.1**
GOLGA8CP		signif in si-RBMX/L1	**1.2**	**2.0**
AIF1L		signif in si-RBMX/L1	**0.5**	**2.0**
RBP4		signif in si-RBMX/L1	**0.8**	**2.0**
STMN3		signif in si-RBMX/L1	**0.4**	**2.0**
B3GNT5		signif in si-RBMX/L1	**1.2**	**2.0**
MUC17		signif in si-RBMX/L1	**0.6**	**1.9**
ZNF560		signif in si-RBMX/L1	**1.1**	**1.9**
TMEM86A		signif in si-RBMX/L1	**1.1**	**1.8**
GPR39		signif in si-RBMX/L1	**1.5**	**1.8**
EFNB1		signif in si-RBMX/L1	**1.2**	**1.8**
HSD3B7		signif in si-RBMX/L1	**1.4**	**1.7**
FGA		signif in si-RBMX/L1	**0.3**	**1.7**
MAPK13		signif in si-RBMX/L1	**1.5**	**1.7**
DGAT2		signif in si-RBMX/L1	**1.1**	**1.7**
SLC29A3		signif in si-RBMX/L1	**1.3**	**1.7**
ADCK2		signif in si-RBMX/L1	**1.4**	**1.6**
KIAA1671		signif in si-RBMX/L1	**1.5**	**1.6**
GGT1		signif in si-RBMX/L1	**1.1**	**1.6**
PTPRF		signif in si-RBMX/L1	**1.2**	**1.5**
ITPR1		signif in si-RBMX/L1	**1.0**	**1.5**
KIAA1217		signif in si-RBMX/L1	**1.3**	**1.5**
LOC101927374		signif in si-RBMX/L1	**1.3**	**1.5**
PNKD		signif in si-RBMX/L1	**1.1**	**1.5**
XPNPEP2		signif in si-RBMX/L1	**1.2**	**1.4**
APOA4		signif in si-RBMX/L1	**0.7**	**1.4**
FAM212B		signif in si-RBMX/L1	**1.1**	**1.4**
HDHD3		signif in si-RBMX/L1	**1.4**	**1.4**
SLC6A8		signif in si-RBMX/L1	**0.9**	**1.4**
LOC101929567		signif in si-RBMX/L1	**1.3**	**1.4**
H6PD		signif in si-RBMX/L1	**1.2**	**1.4**
ATP11A		signif in si-RBMX/L1	**1.2**	**1.3**

CHAPTER 5. PERSPECTIVES AND FUTURE DIRECTIONS

The packaging of DNA into repressive heterochromatin is a crucial process for the maintenance of genome stability (Peng and Karpen, 2008) and for lineage commitment in the early embryo (Fadloun et al., 2013a), and it has emerged as a major barrier to the reprogramming of differentiated cell identity (Soufi et al., 2012; Becker et al., 2016). Classically, heterochromatic structures were discerned on the basis of their physical compaction, such as by the density of DNA under the microscope (Heitz, 1928; Brown, 1966), and early biophysical studies reported that heterochromatin is resistant to mechanical shearing and sediments more rapidly during centrifugation (Frenster et al., 1963; Doenecke and McCarthy, 1975). Modern genome-wide mapping of histone modifications such as H3K9me3 and H3K27me3, which are enriched over known well-characterized heterochromatic regions such as pericentromeric repeats (Nakayama et al., 2001; Lehnertz et al., 2003) and the inactive X chromosome (Plath et al., 2003), have allowed global assessment of heterochromatic domain positions in the genome and their dynamics across development lineages (Hawkins et al., 2010; Zhu et al., 2013; Xu et al., 2014a). However, emerging evidence of H3K9me3 and H3K27me3 in active chromatin regions (Piacentini et al., 2003; Vakoc et al., 2005; Blahnik et al., 2011; Riddle et al., 2012) suggests that these marks may lack sufficient sensitivity and specificity to reveal domains that are physically condensed and transcriptionally silent. I developed a method, called Gradient-seq, to allow genome-wide mapping of heterochromatic domains on the basis of their compacted structure, as reflected in their resistance to fragmentation. This approach involves isolation of heterochromatic regions as intact chromatin, enabling our team to

characterize the proteins physically associated with these regions, including novel heterochromatin proteins we show to be functional regulators of gene silencing.

5.1 Gradient-seq as a method to map the human heterochromatin landscape

I observed that H3K9me3-marked DBR domains, previously mapped on the basis of an impediment to reprogramming factor binding (Soufi et al., 2012), were consistently depleted in input sequencing tracks for ChIP-seq experiments (Figure 4-1A), which suggested that resistance to sonication may be a widespread feature of heterochromatic regions. This agreed with previous work showing that heterochromatin is resistant to mechanical shearing (Frenster et al., 1963) and that highly sonication-sensitive regions map to active gene promoters (Auerbach et al., 2009), but the genome-wide correspondence between sonication resistance and heterochromatin was not initially clear. I found that isolation of sonication-resistant chromatin fragments was sufficient to enrich heterochromatic sequences in DBRs over euchromatic sequences by a magnitude comparable to conventional H3K9me3 ChIP (Supp. Fig. 4-1C). Using next-generation sequencing, I measured the relative representation of genomic sequences in the sonication-resistant material versus the sonication-sensitive material, a method that I termed Gradient-seq. The domains of sonication-resistant chromatin, which I refer to as structural heterochromatin (strucHC), show striking overlap with most H3K9me3 ChIP-seq domains (Figure 4-1F), contain genes with very low levels of gene expression by mRNA-seq (Figure 4-2D), cover nearly 100% of most heterochromatic satellite repeat elements (Figure 4-3F), show high rates of DNA methylation at CpG islands (Figure 4-2F), and are globally resistant to DNase I digestion (Figure 4-4D). Thus, the regions of the genome

purified by gradient sedimentation meet a variety of criteria traditionally associated with heterochromatin (Richards and Elgin, 2002).

This technique for physically enriching heterochromatin was informed by previous work using sucrose gradients to investigate chromatin structure (Gilbert et al., 2004; Ghirlando and Felsenfeld, 2008; Ishihara et al., 2010). Bickmore and colleagues, whose ultracentrifugation conditions were the basis for my own studies, showed gene-poor regions tend to sediment more rapidly after MNase digestion (Gilbert et al., 2004). Subsequent work showed that transcription of an inducible gene was inversely correlated with sedimentation rate for crosslinked, sonicated chromatin fragments (Ishihara et al., 2010). However, these studies detected DNA sequences using 1 MB interval microarrays (Gilbert et al., 2004) or quantitative PCR (Ishihara et al., 2010), or compared specific fragments excised by nucleases (Ghirlando and Felsenfeld, 2008), and thus rapidly sedimenting regions of chromatin have never been mapped with genomic resolution. More importantly, these studies all isolated fragments of the same size in order to investigate changes in chromatin buoyancy (Gilbert et al., 2004; Ghirlando and Felsenfeld, 2008; Ishihara et al., 2010), and thus their findings were unrelated to fragmentation resistance. Meanwhile, my observations using input tracks for ChIP-seq suggest that sonication resistance alone is a reliable indicator of heterochromatic regions. Furthermore, since our method for detecting heterochromatin avoids an electrophoresis-based size-selection step (Gilbert et al., 2004; Ishihara et al., 2010), it is compatible with subsequent analysis of proteins co-purifying with heterochromatin (Figure 4-5). For the purposes of genomic mapping of heterochromatin, however, future work should investigate whether isolation of large sonicated DNA fragments (by gel electrophoresis or bead selection), and thereby eliminating the effects of differential buoyancy related to chromatin shape, is sufficient to

produce genome-wide maps comparable to Gradient-seq. Such an adapted DNA-focused method might facilitate the generation of heterochromatin maps for diverse cell and tissue types, to gain insight into heterochromatin dynamics across development.

The finding that heterochromatic regions are depleted from highly sonicated chromatin highlights an important bias in some ChIP-seq studies. All ChIP-seq data presented here was normalized to a corresponding input sample on a bin-by-bin basis or per-basepair basis for track generation. However, normalization by input-subtraction or input-division is not a routine part of the bioinformatics workflow for public consortia such as Roadmap Epigenomics (Zhu et al., 2013; Roadmap Epigenomics Consortium et al., 2015) or ENCODE (ENCODE Project Consortium, 2012), which make ChIP-seq signal tracks available in unnormalized form. Studies based on these public resources, without normalizing for input, have claimed that large H3K9me3 domains in human cells are dependent upon serum used for cell culture (Zhu et al., 2013). However, my re-analysis of the same datasets revealed that H3K9me3 domains can be detected with comparable abundance across a range of cell and tissue types, *in vitro* and *in vivo*, but this detection requires input normalization (Supp. Fig. 4-1B). Thus, although it is possible that serum affects the magnitude of the H3K9me3 ChIP signal, neither the existence of large H3K9me3 domains nor the sonication resistance of those domains are culture artifacts. Recently, the Roadmap consortium published a chromatin state segmentation that consistently assigned small fractions of the genome to the “Heterochromatin” state (Roadmap Epigenomics Consortium et al., 2015). The algorithm used in this analysis did account for local enrichments in input but assumed a minimum background signal based on the genome average (Roadmap Epigenomics Consortium et al., 2015). My results suggest that these widely used genomic tools might warrant revision to account for regions

consistently depleted from input chromatin. Furthermore, in light of other recent studies mapping domains domains resistant to MNase digestion (Mieczkowski et al., 2016), which generally agree with the sonication-resistant strucHC domains (Supp. Fig. 4-4B,C) (exceptions discussed below), local normalization to input should be applied to ChIP-seq datasets produced by sonication and MNase digestion alike.

5.2 Heterochromatic versus euchromatic subtypes of H3K9me3 domains

Overall, the strucHC domains enriched by Gradient-seq show a high rate of agreement with conventional H3K9me3 ChIP-seq (Figure 4-1F, 4-2A), and 96.8% of the H3K9me3 domain coverage falls into either the strucHC or intermediate domain categories (Figure 4-2C), which have low rates of gene transcription (Figure 4-3B). However, the remaining 31 MB of H3K9me3-marked chromatin overlaps with structurally euchromatic domains, defined by a depletion of sequencing signal in the strucHC fraction compared to the euchromatin fraction (see plot in Figure 4-4A). This euchromatic subtype of H3K9me3 domains has remarkable selectivity for specific gene classes (Table 4-1), in particular KRAB-ZNF genes (Figure 4-3A,C), which form large gene clusters on chromosome 19 (Vogel et al., 2006). Earlier work shows that these genes are expressed from both alleles and are co-marked by H3K9me3 and the elongation-associated mark H3K36me3 (Blahnik et al., 2011), and similarly I find that the euchromatic H3K9me3 domains as a whole contain mostly active genes (Figure 4-3B) and are enriched for H3K36me3 (Supp. Fig. 4-3A). The fact that KRAB-ZNF gene clusters also form large domains of HP1 binding (Vogel et al., 2006) rules out the possibility that H3K9me3 detection in these regions is an artifact of poor antibody specificity. The sonication-sensitive structure of these regions, combined

with the fact that more than three-quarters of the genes they contain have mRNA-seq signal (Figure 4-3B), is arguably sufficient to justify labeling such regions of H3K9me3 as “euchromatic.” However, I note that the classification of these regions as distinct from heterochromatin is further supported by the hypomethylation of CpG islands in these regions (in stark contrast to strucHC, see Figure 4-4B), as well as the reduction in contact with the nuclear lamina (Figure 4-4C, Supp. Fig. 4-4D), which is a conserved property of heterochromatic regions (Towbin et al., 2013).

A previous analysis of three-dimensional genome contacts by Hi-C found that KRAB-ZNF genes have a unique pattern of interactions compared to other H3K9me3-marked regions, warranting their separate classification as subcompartment B4 (Rao et al., 2014) and supporting the idea that these regions are structurally distinct. However, this analysis was performed at low resolution (100-kb), and there was only sufficient sensitivity to detect subcompartment B4 when the analysis was performed on chromosome 19 alone (Rao et al., 2014), meaning that it was not possible to detect similar regions elsewhere in the genome. I have generated the first genome-wide map of H3K9me3 domains with these euchromatic properities, which can be found on all 24 chromosomes and have three-times more genomic coverage than was mapped for subcompartment B4 (Rao et al., 2014). I find that genes other than KRAB-ZNFs, such as the transcribed protocadherin gamma cluster that is involved in intercellular communication (Chen and Maniatis, 2013), are also found in this chromatin subtype (Figure 4-3C). It is notable that H3K9me3 overlapping euchromatin appears so selective for genes that exist as reiterated gene clusters, including HOX genes (Figure 4-3C). It is tempting to speculate H3K9me3 is recruited to these gene bodies, despite the fact that they are transcribed in this cell type, in order to prevent illicit recombination between the

similar sequences of these nearby genes, as has previously been suggested (Blahnik et al., 2011). Indeed, it was shown that an isolated 3' ZNF exon, a unit shared by many ZNF genes, is sufficient to recruit H3K9me3 when integrated elsewhere in the genome (Blahnik et al., 2011). However, this model should be further tested by investigating whether depletion of H3K9me3, HP1, or other heterochromatin components leads to increased recombination or mutagenesis at genes in euchromatic H3K9me3 domains. It will also be informative to compare genes and repeat classes in humans euchromatic H3K9me3 domains to those on Drosophila chromosome 4, which was also shown to have transcriptionally permissive H3K9me3 domains (Riddle et al., 2012), to see if any overarching principles can be discerned.

Although the H3K9me3 domains overlapping euchromatin represent a small fraction the total fibroblast H3K9me3 domains in the genome (31 of 968 MB), these regions have been instrumental to the study of H3K9me3 regulation, without widespread recognition of their unique properties. KRAB-ZNF gene clusters are the major sites where chromatin binding by KAP1/TRIM28 depends on its recruitment by KRAB-ZNF proteins (Iyengar et al., 2011), as part of a potential autoregulatory mechanism (O'Green et al., 2007; Frietze et al., 2010). In light of my findings, KAP1 recruitment by KRAB-ZNFs, which previously has been suggested as a model for heterochromatin domain formation (Groner et al., 2010), is compatible with both gene transcription and a sonication-sensitive chromatin structure, and is therefore not sufficient for heterochromatin establishment. The H3K9me3 signal over ZNF genes is in fact so prominent that a major study associated with the Roadmap Epigenomics consortium (Zhu et al., 2013) used these sites to standardize their H3K9me3 ChIP-seq datasets across cell types. The fact that sonication-sensitive H3K9me3 sites were used to define positive signal, combined with the lack of

normalization to an input control, may have suppressed the detection of sonication-resistant heterochromatic domains in this study (Zhu et al., 2013).

In addition to KAP1, our finding of H3K9me3 domains with euchromatic structure is also important for study of the three-protein HUSH complex, which regulates SETDB1 recruitment and H3K9me3 deposition (Tchasovnikarova et al., 2015). Although euchromatic H3K9me3 domains are comparatively rare, they account for a majority of reported HUSH-regulated sites in fibroblast H3K9me3 domains (Figure 4-3D), including the strongest sites (Supp. Fig. 4-3C) and sites highlighted in the original report (Tchasovnikarova et al., 2015) at ZNF and non-ZNF genes. My data suggest that HUSH complex binding promotes H3K9me3 deposition without significant chromatin compaction, challenging the use of the term “heterochromatin” to describe such sites (Timms et al., 2016). Nonetheless, my observations are compatible with published model that the HUSH complex “dims” the expression of transcribed genes but does not silence them fully (Tchasovnikarova et al., 2015; Timms et al., 2016). Future work should investigate the contribution of HUSH binding to chromatin accessibility as opposed to merely the readout of the H3K9me3 mark.

A major outstanding question is what chromatin components or mechanisms distinguish euchromatic from heterochromatic H3K9me3 domains. Although the HUSH complex seems to preferentially act in the euchromatic subtype (Figure 4-3D), the loss of this complex leads to increased gene expression and reduced H3K9me3 in these regions (Tchasovnikarova et al., 2015), meaning that HUSH is not converting regions that would otherwise be heterochromatic H3K9me3 to euchromatin. An attractive experimental strategy would be to perform H3K9me3-directed IP from of the euchromatic gradient fraction, followed by mass spectrometry, to identify more proteins associated with these

regions. I indeed performed such IPs using the euchromatin fraction (analysis of DNA in Supp. Fig. 4-1D, sequencing track in Figure 4-3A); however, I found that the amount of chromatin obtained was too low to yield informative proteomic data (data not shown). An additional challenge of such a purification strategy is that the second step, the H3K9me3 IP, also re-enriches the heterochromatic H3K9me3 domains (which are depleted <2-fold, but still present, in the euchromatic fraction – see Figure 4-1C), and the much greater genomic abundance of the heterochromatic regions would presumably cause them to overwhelm the signal from the rarer euchromatic regions. A more viable strategy might be to perform careful proteomic comparisons between a traditional H3K9me3 ChIP-MS and H3K9me3 ChIP from the strucHC fraction, seeking to identify proteins enriched in the former compared to the latter, which would be candidates for binders of euchromatic H3K9me3 chromatin. Alternatively, to directly purify regions of euchromatic H3K9me3, future work could use a protocol called “proteomics of isolated chromatin segments” (PICh) (Déjardin and Kingston, 2009), which uses nucleic acid probes to pull down specific loci in chromatin and has been used to identify major satellite-associated proteins (Saksouk et al., 2014), perhaps directed at an H3K9me3-marked KRAB-ZNF exon. Identifying proteins that preferentially promote a euchromatic or heterochromatic structure within H3K9me3 domains would be useful for efforts to increase chromatin accessibility during cell reprogramming. Such findings may also prove relevant to the chromatin state of pluripotent cells, which have H3K9me3-marked regions but lack compacted heterochromatin structures by electron spectroscopic imaging and electron microscopy (Meshorer and Misteli, 2006; Ahmed et al., 2010; Fussner et al., 2011; Underwood et al., 2016).

5.3 Diversity of chromatin structure among H3K27me3 domains

The literature has been conflicted about whether the term “heterochromatin” includes H3K27me3-marked regions, with some authors using the term synonymously with H3K9me3-marked and HP1-bound domains (Beisel and Paro, 2011). Indeed, analysis of select H3K27me3-marked sites reveals accessibility to binding by general transcription machinery (Breiling et al., 2001; Dellino et al., 2004), suggesting that such regions may constitute a “lesser” form of heterochromatin. Meanwhile, paradigmatic examples of facultative heterochromatin, such as Barr body of the inactive chromosome, are marked by H3K27me3 (Plath et al., 2003; Heard, 2005) and also have condensed ultrastructure (Rego et al., 2008). Yet H3K9me3-marked chromatin too is dynamic in development (Hawkins et al., 2010; Soufi et al., 2012; Becker et al., 2016; Figure 2-1), and thus the functional descriptor “facultative” is inadequate to distinguish between the chromatin states associated with H3K27me3 versus H3K9me3. Clearly, greater understanding is needed on the relationship between H3K27me3-marked domains and structurally compact heterochromatin.

I find that 49% of the H3K27me3 domains in human fibroblasts (327 of 672 MB; Figure 4-2C) fall into domains we map as strucHC by Gradient-seq. The quantitative enrichment of these regions in the strucHC fraction compared to the euchromatin fraction is actually similar to the H3K9me3-marked strucHC domains (Figure 4-4A, compare red plots between H3K9me3 and H3K27me3 subtypes). Thus, H3K9me3- and H3K27me3-marked forms of strucHC have a chromatin structure that is comparably resistant to sonication, supporting the notion that H3K27me3 is a bona fide heterochromatin mark (Trojer and Reinberg, 2007). However, compared to H3K9me3, H3K27me3 was less

predictive of a heterochromatic or intermediate state, as 29% of H3K27me3 domains (193 MB; Figure 4-2C) preferentially migrated to the euchromatin fraction (versus 3.2% or 31 MB of H3K9me3). The euchromatic nature of this H3K27me3 subtype is supported by a variety of metrics, including increased gene expression (Figure 4-3B), decreased DNA methylation at CpG islands (Figure 4-4B), and marked sensitivity to DNase I digestion (Figure 4-4D). The euchromatic subtype of H3K27me3 includes all four major HOX gene clusters (Figure 4-3C,E), which have been model loci for the study of Polycomb-regulated regions over decades (Lewis, 1978; Duncan, 1982; Xu et al., 2014b), without widespread acknowledgement that these regions have an atypical chromatin structure compared to more heterochromatic H3K27me3 regions. These findings highlight the challenges of generalizing from specific H3K27me3-marked loci to all PRC targets. Future work might investigate whether the accessibility of H3K27me3-marked chromatin to transcription factor binding (Breiling et al., 2001; Dellino et al., 2004) is different between the strucHC and euchromatic subtypes mapped in this study.

5.4 Relationship between chromatin subtypes and hallmark properties of repressive chromatin

A variety of proprieties have been associated with H3K9me3-marked chromatin or repressive regions of the genome–such as DNA methylation (Lehnertz et al., 2003), localization with the nuclear periphery (Csink and Henikoff, 1996; Guelen et al., 2008; Poleshko et al., 2013; Towbin et al., 2013), late replication (O'Keefe et al., 1992; Chandra et al., 2012), and resistance to nuclease digestion (Gottesfeld et al., 1975; Wallrath and Elgin, 1995; Hamid et al., 1996)—and they have generally have been studied separately. It has been unclear whether these properties are directly associated with H3K9me3

deposition, compacted chromatin structure, or both, and our identification of H3K9me3 domains inside and outside of strucHC (as well as strucHC regions lacking H3K9me3 or H3K27me3 enrichment) provided an opportunity to test these questions. We find that, within CpG islands, the frequency of DNA methylation per CpG shows remarkable correlation with the structural classifications from Gradient-seq (Figure 4-4B). Island CpGs are consistently methylated in strucHC, regardless of histone mark, and are consistently hypomethylated in euchromatin, with and without H3K9me3 or H3K27me3. The highest rates of methylation are, in fact, seen at strucHC domains with neither mark enriched. The nature of this "third" heterochromatin state, distinct from H3K9me3 and H3K27me3 domains, remains mysterious at present. Performing gradient sedimentation before and after 5-azacytosine treatment, to inhibit DNA methyltransferases, would allow determination of whether the observed hypermethylated state is required for heterochromatic compaction in these regions. Overall, the patterns of CpG island methylation, similar to gene transcription (Figure 4-3B), show a direct relationship with genome structure as measured by Gradient-seq. However, since, outside of CpG islands, DNA methylation is high in all chromatin categories (Supp. Fig. 4-4A), bisulfite sequencing cannot be used to map the full territory of heterochromatin regions.

In contrast to CpG island methylation, DNA replication timing is most accurately predicted by histone mark rather than extent of chromatin compaction. The strucHC and euchromatic subtypes of H3K9me3 replicate similarly late in S phase, while all forms of H3K27me3 replicate much earlier (Figure 4-4E). These striking associations for H3K9me3- and H3K27me3-marked chromatin were described in some previous studies (Chandra et al., 2012), while other groups report weak associations between replication timing and histone marks based on genome segmentation models (Dileep et al., 2015),

perhaps related to issues of input-normalization discussed above. My data suggest that late replication might be conferred by chromatin complexes that localize to H3K9me3-marked domains, rather than being a direct byproduct of a compacted chromatin structure, and that late replication and active gene transcription can co-occur (at euchromatic subtypes of H3K9me3). Meanwhile, association of chromatin with the nuclear lamina is predicted by specific combinations of structural and histone mark classifications (Figure 4-4C, Supp. Fig. 4-4D). H3K9me3 domains generally have higher levels of lamin B1 binding compared to H3K27me3 domains, but within both histone mark categories, strucHC and intermediate subtypes showed greater lamin association than euchromatic subtypes.

Resistance to nucleases like DNase I and MNase is typically interpreted as a measure of chromatin structure (Wallrath and Elgin, 1995; Hamid et al., 1996; Thurman et al., 2012). It is reassuring that that there is a strong overall concordance between the classifications from Gradient-seq and nuclease-based datasets, with euchromatin regions showing higher rates of DNase I cleavage (Figure 4-4D) and sensitivity to MNase (Supp. Fig. 4-4B). In particular, fibroblast strucHC domains overlap strikingly with domains of increased resistance to MNase, determined based on MNase titration in K562 leukemia cells (Mieczkowski et al., 2016), particularly among regions with similar H3K9me3 and H3K27me3 signals in both cell types. Thus, my findings can be generally corroborated by independent approaches not involving sonication or gradient sedimentation. However, a key distinction between these methods occurs at the regions of H3K9me3 that are euchromatic by Gradient-seq, but remain largely resistant to DNase I (Figure 4-4D) and MNase (Supp. Fig. 4-4B,C). Inspection of regions co-marked by H3K9me3 and H3K36me3 in both fibroblasts and K562 cells reveals a highly sonication-sensitive

chromatin structure by Gradient-seq but resistance to MNase (Supp. Fig. 4-4C). Since these two “structural” approaches are discordant regarding the accessibility of such regions, independent methods should be used to adjudicate the discrepancy, such as this late breaking paper (Risca et al., 2016) that uses correlations in radiation-induced DNA cleavage to investigate chromatin structure. However, in the meantime, given that euchromatic H3K9me3 domains (according to Gradient-seq) are permissive to gene transcription (Figure 4-3B) and DNA hypomethylation (Figure 4-4B), it is likely that there is at least some increased chromatin accessibility in these regions compared to other types of H3K9me3-marked chromatin. It is possible that something about the local environment of H3K9me3-marked chromatin or the chromatin complexes bound to these sites that inhibits DNA cleavage reactions, including at more euchromatic sites, and that the structure of such regions is better probed by a non-enzymatic method involving sound waves/cavitation. Although it remains to be established whether Gradient-seq is a superior readout of chromatin structure compared to nuclease-based methods, we note that only Gradient-seq allows separation of transcriptionally silent H3K9me3 domains from transcriptionally active ones prior to proteomics, which was a principal goal of our study.

5.5 Heterochromatin impedes direct conversion between differentiated lineages

The impediment of genes to be activated during reprogramming is a property, like lamin B1 binding, that also appears to be related to both structural state of chromatin and histone mark enrichment. Previous transcriptomic analysis showed that, after fibroblasts are converted to human induced hepatocytes (hiHeps), large numbers of genes expressed in

native liver cells remain aberrantly repressed in the reprogrammed hiHep cells (Huang et al., 2014). However, properties of chromatin that predict whether a hepatic gene will be properly induced were not identified. I found that, among hepatic genes that are initially silent in fibroblasts, H3K9me3-marked genes have a profound defect in activation (Supp. Fig. 4-1A), with the majority of genes showing little to no activation above fibroblast levels at the conclusion of hiHep reprogramming. This defect was not apparent in H3K27me3-marked chromatin (Supp. Fig. 4-1A). These findings are consistent with prior evidence that H3K9me3-marked chromatin, in particular, presents an impediment to reprogramming to pluripotency (Soufi et al., 2012; Matoba et al., 2014). Moreover, these results suggest that H3K9me3-marked chromatin poses a general barrier to cell fate reprogramming that is not contingent on specific factor combinations (such as OSKM) (Soufi et al., 2012, 2015) or the reprogramming being towards a state like pluripotency that has more limited heterochromatin (Meshorer and Misteli, 2006; Ahmed et al., 2010; Fussner et al., 2011; Underwood et al., 2016).

The differential effects of H3K9me3 versus H3K27me3 on hepatic gene activation during reprogramming persist even when the comparison is restricted to within strucHC (red plots in Figure 4-6A). Thus, even among regions with a similarly sonication-resistant chromatin structure, H3K9me3-marked chromatin is uniquely refractory to gene activation, revealing an important functional difference between these two heterochromatic marks. Moreover, this finding further justifies the study of H3K9me3-associated heterochromatin proteins as a means to understand barriers to cell conversion (Becker et al., 2016). However, among H3K27me3-marked regions, the strucHC subtype has the lowest levels of hepatic gene induction (Figure 4-6A), suggesting that sonication-resistant heterochromatic structure also impedes reprogramming independent from the H3K9me3

mark. Moreover, whereas the strucHC and intermediate subtypes of H3K9me3 domains are refractory to gene activation by hiHep factors, the euchromatic subtype of H3K9me3 is not (Figure 4-6A). Consequently, the depletion of such sonication-sensitive regions by gradient sedimentation is an important step in order to purify only H3K9me3-associated proteins that are relevant to reprogramming resistance.

In contrast to human iPS reprogramming, which proceeds at low efficiency (Hanna et al., 2009; Vierbuchen and Wernig, 2012; Papp and Plath, 2013), a high fraction of cells (10-15%) at the conclusion of hiHep reprogramming process express the hepatocyte marker genes albumin and alpha-1-antitrypsin (Huang et al., 2014) (Supp. Fig. 4-6A), which represents as much of half the cells infected with all three lentiviruses at MOI ~1 per virus. Thus, there is limited dynamic range to observe improvements of the efficiency of hiHep reprogramming, whereas the main concern is the low fidelity of the conversion. We show that knockdown of SUV39H1 or RBMX/L1 enhances the induction of a variety of hepatic genes, marked by H3K9me3 in the starting fibroblasts, both early in reprogramming (Figure 4-6D) and at the conclusion of the process (Supp. Fig. 4-6E). These findings confirm, first, that the resistance of these heterochromatic H3K9me3 regions to hiHep reprogramming is dependent on the H3K9me3 methyltransferase SUV39H1, similar to the impediment to iPS factor activity (Soufi et al., 2012). Second, this shows that siRNA against SUV39H1 or RBMX/L1 enhances the fidelity of conversion at the genes most refractory to reprogramming. Ongoing work in the laboratory is investigating whether these perturbations also enhance hiHep reprogramming fidelity from a functional standpoint, by testing whether the reprogrammed cells have a greater ability to engraft in and repopulate the livers of FAH-deficient mice (Azuma et al., 2007). Alternatively, one might consider testing whether the siRNA treatments affect the

performance of the reprogrammed cells in *in-vitro* assays of hepatocyte function, such as P450 drug metabolism, considering that such activity is typically limited in programmed hepatocytes (Vallier, 2014) and given that CYP450 genes are repressed in fibroblast H3K9me3 domains (Figure 4-6B).

We note that the activation of these heterochromatinized hepatocyte genes after SUV39H1 or RBMX/L1 knockdown is generally modest in comparison to native hepatocytes. This suggests that these regions of chromatin are still refractory to activation even in the presence of these perturbations, and it may be useful to try enhancing the extent of heterochromatin disruption. One strategy might be to use CRISPR-mediated gene deletion instead of RNAi, to produce cells with complete loss of these heterochromatin proteins. Alternatively, combinations of heterochromatin perturbations should be tested, such as RNAi/CRISPR against multiple H3K9me3-associated proteins, or combining SUV39H1 knockdown with HDAC inhibitions in order to simultaneously disrupt multiple mechanisms of chromatin compaction.

5.6 Proteomics reveals functional regulators of H3K9me3-marked heterochromatin

Our identification of 172 heterochromatin proteins by quantitative proteomics (Figure 4-5B) was based on analysis of H3K9me3 IPs performed using the strucHC fraction, compared to the euchromatin fraction (Figure 4-5A). Thus, our strategy resembles traditional ChIP-MS approaches in terms of its specificity for components of H3K9me3-marked chromatin (Soldi and Bonaldi, 2013; Engelen et al., 2015; Ji et al., 2015), but has the advantage of first depleting transcriptionally active H3K9me3 domains during the

gradient sedimentation step. As expected, the list of identified heterochromatin proteins includes known structural components of heterochromatin, such as linker histone and lamin B1, as well as a variety of known mediators of gene silencing (Figure 4-5D). Our results also overlap significantly, but not completely, with the only previous H3K9me3 ChIP-MS dataset performed for differentiated cells (Soldi and Bonaldi, 2013) (Figure 4-5E). Furthermore, the strong agreement between proteins enriched in the strucHC fraction and those enriched in the H3K9me3 IP off of the strucHC fraction (Figure 4-5C) is consistent with H3K9me3-marked chromatin being the major component of the strucHC fraction. The fact that 22 of the heterochromatin proteins have previously been shown to impede iPS reprogramming (Supp. Fig. 4-5B) suggests that this should be a rich dataset for revealing new factors involved in restricting cell fate conversion.

A surprising finding is the large number of heterochromatin proteins with RNA binding activity (119 of 172). However, this subset of heterochromatin proteins actually has an improved rate of overlap with other H3K9me3-related proteomic studies (Figure 4-5F) (Vermeulen et al., 2010; Soldi and Bonaldi, 2013), indicating that this observation is not particular to our approach or method of analysis. The presence of RNA-binding proteins in heterochromatin is consistent with previous evidence that RNA-binding by HP1 is required for its heterochromatin localization (Muchardt et al., 2002), that some constitutive heterochromatin regions generate RNA (Saksouk et al., 2015), and that particular noncoding RNAs are important for heterochromatic compaction after escape from pluripotency (Savić et al., 2014). Specific RNA-binding proteins like HNRNPK, which we also find in heterochromatin, have been shown to regulate H3K9me3 methyltransferases and reprogramming (Bao et al., 2015; Thompson et al., 2015). Nonetheless, the role of RNA in mammalian heterochromatin remains controversial, as

RNAi-dependent mechanisms for heterochromatin establishment discovered in yeast have not been verified in mammals (Saksouk et al., 2015). Since our two-step method for heterochromatin purification (gradient sedimentation followed by H3K9me3 IP) achieves high purity for heterochromatic regions (Supp. Fig. 4-1C), a major outstanding question is which RNAs are present in this sample. Indeed, we have performed RNA-seq from these samples, in collaboration with Roberto Bonasio's laboratory, and analysis of this data is in process.

We show that two RNA-binding proteins have functional roles in heterochromatic gene silencing. First, we show that TARDBP (also known as TDP-43), which is the least enriched among six ALS-related proteins in heterochromatin (Figure 4-5G) but the most comprehensively studied (Lee et al., 2012), suppresses expression of genes in strucHC domains (Supp. Fig. 4-5C,D), using microarray data form human patient-derived fibroblasts. This large number of strucHC genes shows modest but widespread upregulation in fibroblasts from TARDBP-mutant ALS patients, compared to healthy controls, but not in sporadic ALS cases with wild-type TARDBP (Supp. Fig. 4-5C,D). Second, we studied the role of RBMX (together with the related protein RBMXL1) in regulating gene expression genome-wide, with and without the presence of hiHep reprogramming factors. Indeed, RBMX/L1 knockdown led to significant upregulation of 67 strucHC genes in the absence of exogenous factors (Figure 4-7E, Supp. Fig. 4-7C). However, the most striking result was observed in the context of hiHep reprogramming, where the effects of SUV39H1 siRNA in enhancing activation of strucHC genes were closely mirrored by RBMX/L1 siRNA (Figure 4-7C; Table 4-5). Thus, RBMX/L1 contributes to the observed impediment in activating heterochromatic liver genes during hiHep

reprogramming, consistent with what was observed at candidate transcripts by RT-PCR (Figure 4-6D).

The mechanisms by which RBMX/L1 and TARDBP function in heterochromatic silencing is currently unclear. A crucial question is whether the RNA-binding activity of these proteins is required for the localization to or activity in strucHC. Previous work shows that the RNA-recognition motif (RRM) of RBMX is dispensable for both its binding to chromatin and its role in regulating centromeric cohesion (Matsunaga et al., 2012), a function that has been shown for other heterochromatin proteins like SUV39H1 (Peters et al., 2001). Thus, RBMX's role in strucHC may not require RNA interaction at all. This question should be investigated by testing if an RRM-deletion mutant of RBMX phenocopies RBMX deletion in terms of enhancing the activation of H3K9me3-marked hepatic genes during reprogramming. A similar strategy can be pursued for TARDBP and other RNA-binding proteins in our heterochromatin dataset, to investigate the wider functional relevance of the observed enrichment for RNA-binding domains among these proteins. In addition, to gain insight into the mechanism of RBMX function, it will be useful to identify RBMX-interacting proteins by IP and mass spectrometry, to see if RBMX binds HP1 or H3K9me3 methyltransferases, as was previously shown for HNRNPK (Bao et al., 2015; Thompson et al., 2015).

5.7 Monitoring reprogramming impediments to screen heterochromatin proteins

Our studies with the canonical H3K9me3 regulator SUV39H1 as well as RBMX illustrate that perturbation of heterochromatin does not immediately induce widespread de-

repression of silenced genes. This agrees with previous genetic studies of SUV39H1/H2 (Bulut-Karslioglu et al., 2014), it may indicate the existence of redundant silencing mechanisms or simply the requirement for targeted transcriptional activators to upregulate these cell-type specific genes (Figure 4-6D). In light of this data, testing putative heterochromatin regulators for whether they are directly required for gene silencing may be an inappropriate criterion for establishing functionality. The key functional property of heterochromatin is not merely transcriptional repression but dominance over transcriptional activating mechanisms, as seen for classic studies of position effect variegation (Wallrath and Elgin, 1995; Elgin and Reuter, 2013), as well as, more recently, resistance to reprogramming (Soufi et al., 2012). Our experiments in the context of hiHep conversion (Figure 4-6D) show that monitoring activation of sentinel heterochromatinized genes following ectopic addition of alternative-lineage transcription factors is a powerful assay for revealing the action of novel heterochromatin regulators.

Given that we've identified multiple hepatic genes that show an H3K9me3-dependent impediment to activation during hiHep reprogramming (Figure 4-7C,D), it is possible to leverage such sites in order to screen a wider set of proteins for functional roles in heterochromatin. Specifically, we are conducting an siRNA screen of approximately 100 heterochromatin proteins, from our proteomic analysis as well as select candidates from the literature. Similar to the experimental design in Figure 4-7A (top), we have transduced cells with hiHep transcription factors followed by independent siRNAs against each candidate heterochromatin protein. The endpoint of the assay is RT-PCR for hepatic genes in strucHC, such as *DSC2*, *CRP* and *NR1H4* (Figure 4-6D, 4-7D), with the goal of identifying proteins whose knockdown enhances the reprogramming of these genes, similar to RBMX. This approach can be extended beyond the hepatic lineage, by

using transcription factor combinations for iPS reprogramming (Takahashi and Yamanaka, 2006), conversion to neurons (Vierbuchen et al., 2010), or reprogramming to diverse other lineages. Our major goal is to identify additional components of the heterochromatin proteome that, like RBMX, oppose the induction of H3K9me3-marked genes for other fates. Of particular interest would be proteins that are involved in impeding conversion to specific lineages but not others. Such findings could help shed light on the cell type-specific mechanisms of heterochromatin establishment that help to guard against changes in cell identity, but that so far have remained elusive.

REFERENCES

Aagaard, L., Laible, G., Selenko, P., Schmid, M., Dorn, R., Schotta, G., Kuhfittig, S., Wolf, A., Lebersorger, A., Singh, P.B., et al. (1999). Functional mammalian homologues of the Drosophila PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31. EMBO J. *18*, 1923–1938.

Adam, R.C., and Fuchs, E. (2016). The Yin and Yang of Chromatin Dynamics In Stem Cell Fate Selection. Trends Genet. TIG *32*, 89–100.

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

Allan, R.S., Zueva, E., Cammas, F., Schreiber, H.A., Masson, V., Belz, G.T., Roche, D., Maison, C., Quivy, J.-P., Almouzni, G., et al. (2012). An epigenetic silencing pathway controlling T helper 2 cell lineage commitment. Nature *487*, 249–253.

Al-Sady, B., Madhani, H.D., and Narlikar, G.J. (2013). Division of labor between the chromodomains of HP1 and Suv39 methylase enables coordination of heterochromatin spread. Mol. Cell *51*, 80–91.

Amlie-Wolf, A., Ryvkin, P., Tong, R., Dragomir, I., Suh, E., Xu, Y., Van Deerlin, V.M., Gregory, B.D., Kwong, L.K., Trojanowski, J.Q., et al. (2015). Transcriptomic Changes Due to Cytoplasmic TDP-43 Expression Reveal Dysregulation of Histone Transcripts and Nuclear Chromatin. PloS One *10*, e0141836.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinforma. Oxf. Engl. *31*, 166–169.

Asgari, S., Moslem, M., Bagheri-Lankarani, K., Pournasr, B., Miryounesi, M., and Baharvand, H. (2013). Differentiation and transplantation of human induced pluripotent stem cell-derived hepatocyte-like cells. Stem Cell Rev. *9*, 493–504.

Aucott, R., Bullwinkel, J., Yu, Y., Shi, W., Billur, M., Brown, J.P., Menzel, U., Kioussis, D., Wang, G., Reisert, I., et al. (2008). HP1-beta is required for development of the cerebral neocortex and neuromuscular junctions. J. Cell Biol. *183*, 597–606.

Audergon, P.N.C.B., Catania, S., Kagansky, A., Tong, P., Shukla, M., Pidoux, A.L., and Allshire, R.C. (2015). Epigenetics. Restricted epigenetic inheritance of H3K9 methylation. Science *348*, 132–135.

Auerbach, R.K., Euskirchen, G., Rozowsky, J., Lamarre-Vincent, N., Moqtaderi, Z., Lefrançois, P., Struhl, K., Gerstein, M., and Snyder, M. (2009). Mapping accessible chromatin regions using Sono-Seq. Proc. Natl. Acad. Sci. U. S. A. *106*, 14926–14931.

Ayyanathan, K., Lechner, M.S., Bell, P., Maul, G.G., Schultz, D.C., Yamada, Y., Tanaka, K., Torigoe, K., and Rauscher, F.J. (2003). Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. Genes Dev. *17*, 1855–1869.

Azuma, H., Paulk, N., Ranade, A., Dorrell, C., Al-Dhalimy, M., Ellis, E., Strom, S., Kay, M.A., Finegold, M., and Grompe, M. (2007). Robust expansion of human hepatocytes in Fah-/-/Rag2-/-/Il2rg-/- mice. Nat. Biotechnol. *25*, 903–910.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120–124.

Bao, X., Wu, H., Zhu, X., Guo, X., Hutchins, A.P., Luo, Z., Song, H., Chen, Y., Lai, K., Yin, M., et al. (2015). The p53-induced lincRNA-p21 derails somatic cell reprogramming by sustaining H3K9me3 and CpG methylation at pluripotency gene promoters. Cell Res. *25*, 80–92.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823–837.

Bayne, E.H., White, S.A., Kagansky, A., Bijos, D.A., Sanchez-Pulido, L., Hoe, K.-L., Kim, D.-U., Park, H.-O., Ponting, C.P., Rappsilber, J., et al. (2010). Stc1: a critical link between RNAi and chromatin modification required for heterochromatin integrity. Cell *140*, 666–677.

Becker, J.S., Nicetto, D., and Zaret, K.S. (2016). H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. Trends Genet. TIG *32*, 29–41.

Beisel, C., and Paro, R. (2011). Silencing chromatin: comparing modes and mechanisms. Nat. Rev. Genet. *12*, 123–135.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315–326.

Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R., et al. (2010). The NIH Roadmap Epigenomics Mapping Consortium. Nat. Biotechnol. *28*, 1045–1048.

Bilodeau, S., Kagey, M.H., Frampton, G.M., Rahl, P.B., and Young, R.A. (2009). SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. Genes Dev. *23*, 2484–2489.

Black, J.C., Van Rechem, C., and Whetstine, J.R. (2012). Histone lysine methylation dynamics: establishment, regulation, and biological impact. Mol. Cell *48*, 491–507.

Blahnik, K.R., Dou, L., Echipare, L., Iyengar, S., O'Geen, H., Sanchez, E., Zhao, Y., Marra, M.A., Hirst, M., Costello, J.F., et al. (2011). Characterization of the contradictory chromatin signatures at the 3' exons of zinc finger genes. PloS One *6*, e17121.

Blelloch, R., Wang, Z., Meissner, A., Pollard, S., Smith, A., and Jaenisch, R. (2006). Reprogramming efficiency following somatic cell nuclear transfer is influenced by the differentiation and methylation state of the donor nucleus. Stem Cells Dayt. Ohio *24*, 2007–2013.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947–956.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349–353.

Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H., and Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev. *20*, 1123–1136.

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

Breiling, A., Turner, B.M., Bianchi, M.E., and Orlando, V. (2001). General transcription factors bind promoters repressed by Polycomb group proteins. Nature *412*, 651–655.

Brown, S.W. (1966). Heterochromatin. Science *151*, 417–425.

Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell *150*, 1209–1222.

Bühler, M., and Moazed, D. (2007). Transcription and RNAi in heterochromatic gene silencing. Nat. Struct. Mol. Biol. *14*, 1041–1048.

Bühler, M., Verdel, A., and Moazed, D. (2006). Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. Cell *125*, 873–886.

Bühler, M., Haas, W., Gygi, S.P., and Moazed, D. (2007). RNAi-dependent and -independent RNA turnover mechanisms contribute to heterochromatic gene silencing. Cell *129*, 707–721.

Bui, H.-T., Seo, H.-J., Park, M.-R., Park, J.-Y., Thuan, N.V., Wakayama, T., and Kim, J.-H. (2011). Histone deacetylase inhibition improves activation of ribosomal RNA genes and

embryonic nucleolar reprogramming in cloned mouse embryos. Biol. Reprod. *85*, 1048–1056.

Bulut-Karslioglu, A., Perrera, V., Scaranaro, M., de la Rosa-Velazquez, I.A., van de Nobelen, S., Shukeir, N., Popow, J., Gerle, B., Opravil, S., Pagani, M., et al. (2012). A transcription factor-based mechanism for mouse heterochromatin formation. Nat. Struct. Mol. Biol. *19*, 1023–1030.

Bulut-Karslioglu, A., De La Rosa-Velázquez, I.A., Ramirez, F., Barenboim, M., Onishi-Seebacher, M., Arand, J., Galán, C., Winter, G.E., Engist, B., Gerle, B., et al. (2014). Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. Mol. Cell *55*, 277–290.

Burton, A., and Torres-Padilla, M.-E. (2014). Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. Nat. Rev. Mol. Cell Biol. *15*, 723–734.

Cahan, P., Li, H., Morris, S.A., Lummertz da Rocha, E., Daley, G.Q., and Collins, J.J. (2014). CellNet: network biology applied to stem cell engineering. Cell *158*, 903–915.

Canzio, D., Chang, E.Y., Shankar, S., Kuchenbecker, K.M., Simon, M.D., Madhani, H.D., Narlikar, G.J., and Al-Sady, B. (2011). Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly. Mol. Cell *41*, 67–81.

Casanova, M., Pasternak, M., Marjou, F. El, Le Baccon, P., Probst, A.V., and Almouzni, G. (2013). Heterochromatin reorganization during early mouse development requires a single-stranded noncoding transcript. Cell Rep. *4*, 1156–1167.

Chadwick, B.P., and Willard, H.F. (2004). Multiple spatially distinct types of facultative heterochromatin on the human inactive X chromosome. Proc. Natl. Acad. Sci. U. S. A. *101*, 17450–17455.

Chandra, T., Kirschner, K., Thuret, J.-Y., Pope, B.D., Ryba, T., Newman, S., Ahmed, K., Samarajiwa, S.A., Salama, R., Carroll, T., et al. (2012). Independence of repressive histone marks and chromatin compaction during senescent heterochromatic layer formation. Mol. Cell *47*, 203–214.

Cheloufi, S., Elling, U., Hopfgartner, B., Jung, Y.L., Murn, J., Ninova, M., Hubmann, M., Badeaux, A.I., Euong Ang, C., Tenen, D., et al. (2015). The histone chaperone CAF-1 safeguards somatic cell identity. Nature *528*, 218–224.

Chen, W.V., and Maniatis, T. (2013). Clustered protocadherins. Dev. Camb. Engl. *140*, 3297–3302.

Chen, J., Liu, H., Liu, J., Qi, J., Wei, B., Yang, J., Liang, H., Chen, Y., Chen, J., Wu, Y., et al. (2013). H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. Nat. Genet. *45*, 34–42.

Chow, J.C., Ciaudo, C., Fazzari, M.J., Mise, N., Servant, N., Glass, J.L., Attreed, M., Avner, P., Wutz, A., Barillot, E., et al. (2010). LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. Cell *141*, 956–969.

Cirulli, E.T., Lasseigne, B.N., Petrovski, S., Sapp, P.C., Dion, P.A., Leblond, C.S., Couthouis, J., Lu, Y.-F., Wang, Q., Krueger, B.J., et al. (2015). Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. Science *347*, 1436–1441.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Csink, A.K., and Henikoff, S. (1996). Genetic modification of heterochromatic association and nuclear organization in Drosophila. Nature *381*, 529–531.

Czermin, B., Schotta, G., Hülsmann, B.B., Brehm, A., Becker, P.B., Reuter, G., and Imhof, A. (2001). Physical and functional association of SU(VAR)3-9 and HDAC1 in Drosophila. EMBO Rep. *2*, 915–919.

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., and Pirrotta, V. (2002). Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell *111*, 185–196.

Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987–1000.

De Cecco, M., Criscione, S.W., Peckham, E.J., Hillenmeyer, S., Hamm, E.A., Manivannan, J., Peterson, A.L., Kreiling, J.A., Neretti, N., and Sedivy, J.M. (2013). Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements. Aging Cell *12*, 247–256.

Déjardin, J., and Kingston, R.E. (2009). Purification of proteins associated with specific genomic Loci. Cell *136*, 175–186.

Dellino, G.I., Schwartz, Y.B., Farkas, G., McCabe, D., Elgin, S.C.R., and Pirrotta, V. (2004). Polycomb silencing blocks transcription initiation. Mol. Cell *13*, 887–893.

van Dijk, T.B., Gillemans, N., Pourfarzad, F., van Lom, K., von Lindern, M., Grosveld, F., and Philipsen, S. (2010). Fetal globin expression is regulated by Friend of Prmt1. Blood *116*, 4349–4352.

Dileep, V., Ay, F., Sima, J., Vera, D.L., Noble, W.S., and Gilbert, D.M. (2015). Topologically associating domains and their long-range contacts are established during early G1 coincident with the establishment of the replication-timing program. Genome Res. *25*, 1104–1113.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. Nature *489*, 101–108.

Djupedal, I., Portoso, M., Spåhr, H., Bonilla, C., Gustafsson, C.M., Allshire, R.C., and Ekwall, K. (2005). RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. Genes Dev. *19*, 2301–2306.

Dodge, J.E., Kang, Y.-K., Beppu, H., Lei, H., and Li, E. (2004). Histone H3-K9 methyltransferase ESET is essential for early development. Mol. Cell. Biol. *24*, 2478–2486.

Doenecke, D., and McCarthy, B.J. (1975). Protein content of chromatin fractions separated by sucrose gradient centrifugation. Biochemistry (Mosc.) *14*, 1366–1372.

Dong, X., Yu, C., Shynlova, O., Challis, J.R.G., Rennie, P.S., and Lye, S.J. (2009). p54nrb is a transcriptional corepressor of the progesterone receptor that modulates transcription of the labor-associated gene, connexin 43 (Gja1). Mol. Endocrinol. Baltim. Md *23*, 1147–1160.

Dou, Z., Xu, C., Donahue, G., Shimi, T., Pan, J.-A., Zhu, J., Ivanov, A., Capell, B.C., Drake, A.M., Shah, P.P., et al. (2015). Autophagy mediates degradation of nuclear lamina. Nature *527*, 105–109.

Du, Y., Wang, J., Jia, J., Song, N., Xiang, C., Xu, J., Hou, Z., Su, X., Liu, B., Jiang, T., et al. (2014). Human hepatocytes with drug metabolic function induced from fibroblasts by lineage reprogramming. Cell Stem Cell *14*, 394–403.

Duerksen, J.D., and McCarthy, B.J. (1971). Distribution of deoxyribonucleic acid sequences in fractionated chromatin. Biochemistry (Mosc.) *10*, 1471–1478.

Duncan, I.M. (1982). Polycomblike: a gene that appears to be required for the normal expression of the bithorax and antennapedia gene complexes of Drosophila melanogaster. Genetics *102*, 49–70.

Eberl, H.C., Spruijt, C.G., Kelstrup, C.D., Vermeulen, M., and Mann, M. (2013). A map of general and specialized chromatin readers in mouse tissues generated by label-free interaction proteomics. Mol. Cell *49*, 368–378.

Ecco, G., Cassano, M., Kauzlaric, A., Duc, J., Coluccio, A., Offner, S., Imbeault, M., Rowe, H.M., Turelli, P., and Trono, D. (2016). Transposable Elements and Their KRAB-ZFP Controllers Regulate Gene Expression in Adult Tissues. Dev. Cell *36*, 611–623.

Eden, A., Gaudet, F., Waghmare, A., and Jaenisch, R. (2003). Chromosomal instability and tumors promoted by DNA hypomethylation. Science *300*, 455.

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D.G., Buetow, K.H., et al. (2008). Global transcription in pluripotent embryonic stem cells. Cell Stem Cell *2*, 437–447.

Eissenberg, J.C., and Elgin, S.C.R. (2014). HP1a: a structural chromosomal protein regulating transcription. Trends Genet. TIG *30*, 103–110.

Eissenberg, J.C., James, T.C., Foster-Hartnett, D.M., Hartnett, T., Ngan, V., and Elgin, S.C. (1990). Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in Drosophila melanogaster. Proc. Natl. Acad. Sci. U. S. A. *87*, 9923–9927.

Elgin, S.C.R., and Reuter, G. (2013). Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila. Cold Spring Harb. Perspect. Biol. *5*, a017780.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Engelen, E., Brandsma, J.H., Moen, M.J., Signorile, L., Dekkers, D.H.W., Demmers, J., Kockx, C.E.M., Ozgür, Z., van IJcken, W.F.J., van den Berg, D.L.C., et al. (2015). Proteins that bind regulatory regions identified by histone modification chromatin immunoprecipitations and mass spectrometry. Nat. Commun. *6*, 7155.

Epsztejn-Litman, S., Feldman, N., Abu-Remaileh, M., Shufaro, Y., Gerson, A., Ueda, J., Deplus, R., Fuks, F., Shinkai, Y., Cedar, H., et al. (2008). De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. Nat. Struct. Mol. Biol. *15*, 1176–1183.

Ezhkova, E., Pasolli, H.A., Parker, J.S., Stokes, N., Su, I. -hsi., Hannon, G., Tarakhovsky, A., and Fuchs, E. (2009). Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. Cell *136*, 1122–1135.

Fadloun, A., Eid, A., and Torres-Padilla, M.-E. (2013a). Mechanisms and dynamics of heterochromatin formation during mammalian development: closed paths and open questions. Curr. Top. Dev. Biol. *104*, 1–45.

Fadloun, A., Le Gras, S., Jost, B., Ziegler-Birling, C., Takahashi, H., Gorab, E., Carninci, P., and Torres-Padilla, M.-E. (2013b). Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. Nat. Struct. Mol. Biol. *20*, 332–338.

Fanti, L., Giovinazzo, G., Berloco, M., and Pimpinelli, S. (1998). The heterochromatin protein 1 prevents telomere fusions in Drosophila. Mol. Cell *2*, 527–538.

Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., Cedar, H., and Bergman, Y. (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. Nat. Cell Biol. *8*, 188–194.

Ferrari, K.J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stützer, A., Fischle, W., Bonaldi, T., and Pasini, D. (2014). Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. Mol. Cell *53*, 49–62.

Filion, G.J., and van Steensel, B. (2010). Reassessing the abundance of H3K9me2 chromatin domains in embryonic stem cells. Nat. Genet. *42*, 4; author reply 5-6.

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

Fisher, A.G., and Merkenschlager, M. (2002). Gene silencing, cell fate and nuclear organisation. Curr. Opin. Genet. Dev. *12*, 193–197.

Fodor, B.D., Shukeir, N., Reuter, G., and Jenuwein, T. (2010). Mammalian Su(var) genes in chromatin control. Annu. Rev. Cell Dev. Biol. *26*, 471–501.

Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin compaction by a polycomb group protein complex. Science *306*, 1574–1577.

Frenster, J.H., Allfrey, V.G., and Mirsky, A.E. (1963). Repressed and active chromatin isolated from interphase lymphocytes. Proc. Natl. Acad. Sci. U. S. A. *50*, 1026–1032.

Friedman, J.R., Fredericks, W.J., Jensen, D.E., Speicher, D.W., Huang, X.P., Neilson, E.G., and Rauscher, F.J. (1996). KAP-1, a novel corepressor for the highly conserved KRAB repression domain. Genes Dev. *10*, 2067–2078.

Frietze, S., O'Geen, H., Blahnik, K.R., Jin, V.X., and Farnham, P.J. (2010). ZNF274 recruits the histone methyltransferase SETDB1 to the 3' ends of ZNF genes. PloS One *5*, e15082.

Fukagawa, T., Nogami, M., Yoshikawa, M., Ikeno, M., Okazaki, T., Takami, Y., Nakayama, T., and Oshimura, M. (2004). Dicer is essential for formation of the heterochromatin structure in vertebrate cells. Nat. Cell Biol. *6*, 784–791.

Fussner, E., Djuric, U., Strauss, M., Hotta, A., Perez-Iratxeta, C., Lanner, F., Dilworth, F.J., Ellis, J., and Bazett-Jones, D.P. (2011). Constitutive heterochromatin reorganization during somatic cell reprogramming. EMBO J. *30*, 1778–1789.

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

Gaspar-Maia, A., Qadeer, Z.A., Hasson, D., Ratnakumar, K., Leu, N.A., Leroy, G., Liu, S., Costanzi, C., Valle-Garcia, D., Schaniel, C., et al. (2013). MacroH2A histone variants act as a barrier upon reprogramming towards pluripotency. Nat. Commun. *4*, 1565.

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. Nat. Rev. Genet. *15*, 829–845.

Ghirlando, R., and Felsenfeld, G. (2008). Hydrodynamic studies on defined heterochromatin fragments support a 30-nm fiber having six nucleosomes per turn. J. Mol. Biol. *376*, 1417–1425.

Gilbert, N., and Allan, J. (2001). Distinctive higher-order chromatin structure at mammalian centromeres. Proc. Natl. Acad. Sci. U. S. A. *98*, 11949–11954.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647–661.

Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N.P., and Bickmore, W.A. (2004). Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. Cell *118*, 555–566.

Gottesfeld, J.M., Murphy, R.F., and Bonner, J. (1975). Structure of transcriptionally active chromatin. Proc. Natl. Acad. Sci. U. S. A. *72*, 4404–4408.

Graf, T., and Enver, T. (2009). Forcing cells to change lineages. Nature *462*, 587–594.

Grewal, S.I.S., and Elgin, S.C.R. (2007). Transcription and RNA interference in the formation of heterochromatin. Nature *447*, 399–406.

Groner, A.C., Meylan, S., Ciuffi, A., Zangger, N., Ambrosini, G., Dénervaud, N., Bucher, P., and Trono, D. (2010). KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. PLoS Genet. *6*, e1000869.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., et al. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature *453*, 948–951.

Guenatri, M., Bailly, D., Maison, C., and Almouzni, G. (2004). Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. J. Cell Biol. *166*, 493–505.

Gurdon, J.B., Elsdale, T.R., and Fischberg, M. (1958). Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei. Nature *182*, 64–65.

Hall, I.M., Shankaranarayana, G.D., Noma, K.-I., Ayoub, N., Cohen, A., and Grewal, S.I.S. (2002). Establishment and maintenance of a heterochromatin domain. Science *297*, 2232–2237.

Hamid, Q.A., Thanumalayan, S., and Parnaik, V.K. (1996). An improved method to distinguish micrococcal nuclease sensitivity of chromatin. J. Biochem. Biophys. Methods *33*, 59–64.

Han, H., Irimia, M., Ross, P.J., Sung, H.-K., Alipanahi, B., David, L., Golipour, A., Gabut, M., Michael, I.P., Nachman, E.N., et al. (2013). MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. Nature *498*, 241–245.

Hanna, J., Wernig, M., Markoulaki, S., Sun, C.-W., Meissner, A., Cassady, J.P., Beard, C., Brambrink, T., Wu, L.-C., Townes, T.M., et al. (2007). Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. Science *318*, 1920–1923.

Hanna, J., Saha, K., Pando, B., van Zon, J., Lengner, C.J., Creyghton, M.P., van Oudenaarden, A., and Jaenisch, R. (2009). Direct cell reprogramming is a stochastic process amenable to acceleration. Nature *462*, 595–601.

Haridass, D., Narain, N., and Ott, M. (2008). Hepatocyte transplantation: waiting for stem cells. Curr. Opin. Organ Transplant. *13*, 627–632.

Harr, J.C., Luperchio, T.R., Wong, X., Cohen, E., Wheelan, S.J., and Reddy, K.L. (2015). Directed targeting of chromatin to the nuclear lamina is mediated by chromatin state and A-type lamins. J. Cell Biol. *208*, 33–52.

Hathaway, N.A., Bell, O., Hodges, C., Miller, E.L., Neel, D.S., and Crabtree, G.R. (2012). Dynamics and memory of heterochromatin in living cells. Cell *149*, 1447–1460.

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

Hayashihara, K., Uchiyama, S., Shimamoto, S., Kobayashi, S., Tomschik, M., Wakamatsu, H., No, D., Sugahara, H., Hori, N., Noda, M., et al. (2010). The middle region of an HP1-binding protein, HP1-BP74, associates with linker DNA at the entry/exit site of nucleosomal DNA. J. Biol. Chem. *285*, 6498–6507.

Heard, E. (2005). Delving into the diversity of facultative heterochromatin: the epigenetics of the inactive X chromosome. Curr. Opin. Genet. Dev. *15*, 482–489.

Heitz, E. (1928). Das Heterochromatin der Moose (Bornträger).

Hemberger, M., Dean, W., and Reik, W. (2009). Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. Nat. Rev. Mol. Cell Biol. *10*, 526–537.

Highley, J.R., Kirby, J., Jansweijer, J.A., Webb, P.S., Hewamadduma, C.A., Heath, P.R., Higginbottom, A., Raman, R., Ferraiuolo, L., Cooper-Knock, J., et al. (2014). Loss of nuclear TDP-43 in amyotrophic lateral sclerosis (ALS) causes altered expression of splicing machinery and widespread dysregulation of RNA splicing in motor neurones. Neuropathol. Appl. Neurobiol. *40*, 670–685.

Hiratani, I., Ryba, T., Itoh, M., Rathjen, J., Kulik, M., Papp, B., Fussner, E., Bazett-Jones, D.P., Plath, K., Dalton, S., et al. (2010). Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis. Genome Res. *20*, 155–169.

Hodges, C., and Crabtree, G.R. (2012). Dynamics of inherently bounded histone modification domains. Proc. Natl. Acad. Sci. U. S. A. *109*, 13296–13301.

Huang, P., He, Z., Ji, S., Sun, H., Xiang, D., Liu, C., Hu, Y., Wang, X., and Hui, L. (2011). Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. Nature *475*, 386–389.

Huang, P., Zhang, L., Gao, Y., He, Z., Yao, D., Wu, Z., Cen, J., Chen, X., Liu, C., Hu, Y., et al. (2014). Direct reprogramming of human fibroblasts to functional and expandable hepatocytes. Cell Stem Cell *14*, 370–384.

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

Iager, A.E., Ragina, N.P., Ross, P.J., Beyhan, Z., Cunniff, K., Rodriguez, R.M., and Cibelli, J.B. (2008). Trichostatin A improves histone acetylation in bovine somatic cell nuclear transfer early embryos. Cloning Stem Cells *10*, 371–379.

Ieda, M., Fu, J.-D., Delgado-Olguin, P., Vedantham, V., Hayashi, Y., Bruneau, B.G., and Srivastava, D. (2010). Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. Cell *142*, 375–386.

Ishihara, S., Varma, R., and Schwartz, R.H. (2010). A new fractionation assay, based on the size of formaldehyde-crosslinked, mildly sheared chromatin, delineates the chromatin structure at promoter regions. Nucleic Acids Res. *38*, e124.

Iwafuchi-Doi, M., and Zaret, K.S. (2014). Pioneer transcription factors in cell reprogramming. Genes Dev. *28*, 2679–2692.

Iyengar, S., and Farnham, P.J. (2011). KAP1 protein: an enigmatic master regulator of the genome. J. Biol. Chem. *286*, 26267–26276.

Iyengar, S., Ivanov, A.V., Jin, V.X., Rauscher, F.J., and Farnham, P.J. (2011). Functional analysis of KAP1 genomic recruitment. Mol. Cell. Biol. *31*, 1833–1847.

Ji, S., Zhang, L., and Hui, L. (2013). Cell fate conversion: direct induction of hepatocyte-like cells from fibroblasts. J. Cell. Biochem. *114*, 256–265.

Ji, X., Dadon, D.B., Abraham, B.J., Lee, T.I., Jaenisch, R., Bradner, J.E., and Young, R.A. (2015). Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions. Proc. Natl. Acad. Sci. U. S. A. *112*, 3841–3846.

Jinn-Fei, Y., and el-Labban, N.G. (1986). An ultrastructural study of binucleate plasma cells. J. Oral Pathol. *15*, 118–121.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. *13*, 484–492.

Jullien, J., Pasque, V., Halley-Stott, R.P., Miyamoto, K., and Gurdon, J.B. (2011). Mechanisms of nuclear reprogramming by eggs and oocytes: a deterministic process? Nat. Rev. Mol. Cell Biol. *12*, 453–459.

Jullien, J., Miyamoto, K., Pasque, V., Allen, G.E., Bradshaw, C.R., Garrett, N.J., Halley-Stott, R.P., Kimura, H., Ohsumi, K., and Gurdon, J.B. (2014). Hierarchical molecular events driven by oocyte-specific factors lead to rapid and extensive reprogramming. Mol. Cell *55*, 524–536.

Kanellopoulou, C., Muljo, S.A., Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M., and Rajewsky, K. (2005). Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes Dev. *19*, 489–501.

Kanhere, A., Viiri, K., Araújo, C.C., Rasaiyaah, J., Bouwman, R.D., Whyte, W.A., Pereira, C.F., Brookes, E., Walker, K., Bell, G.W., et al. (2010). Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. Mol. Cell *38*, 675–688.

Kapopoulou, A., Mathew, L., Wong, A., Trono, D., and Jensen, J.D. (2016). The evolution of gene expression and binding specificity of the largest transcription factor family in primates. Evol. Int. J. Org. Evol. *70*, 167–180.

Kato, H., Goto, D.B., Martienssen, R.A., Urano, T., Furukawa, K., and Murakami, Y. (2005). RNA polymerase II is required for RNAi-dependent heterochromatin assembly. Science *309*, 467–469.

Kazazian, H.H. (2004). Mobile elements: drivers of genome evolution. Science *303*, 1626–1632.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

van Koningsbruggen, S., Gierlinski, M., Schofield, P., Martin, D., Barton, G.J., Ariyurek, Y., den Dunnen, J.T., and Lamond, A.I. (2010). High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. Mol. Biol. Cell *21*, 3735–3748.

Kurimoto, K., Yabuta, Y., Hayashi, K., Ohta, H., Kiyonari, H., Mitani, T., Moritoki, Y., Kohri, K., Kimura, H., Yamamoto, T., et al. (2015). Quantitative Dynamics of Chromatin Remodeling during Germ Cell Specification from Mouse Embryonic Stem Cells. Cell Stem Cell *16*, 517–532.

Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes Dev. *16*, 2893–2905.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116–120.

Ladewig, J., Koch, P., and Brüstle, O. (2013). Leveling Waddington: the emergence of direct programming and the loss of cell fate hierarchies. Nat. Rev. Mol. Cell Biol. *14*, 225–236.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature *499*, 214–218.

Lee, E.B., Lee, V.M.-Y., and Trojanowski, J.Q. (2012). Gains or losses: molecular mechanisms of TDP43-mediated neurodegeneration. Nat. Rev. Neurosci. *13*, 38–50.

Lee, M.G., Villa, R., Trojer, P., Norman, J., Yan, K.-P., Reinberg, D., Di Croce, L., and Shiekhattar, R. (2007). Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. Science *318*, 447–450.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., et al. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301–313.

Lehnertz, B., Ueda, Y., Derijck, A.A.H.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H.F.M. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr. Biol. CB *13*, 1192–1200.

Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565–570.

Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell *128*, 707–719.

Li, W., Jin, Y., Prazak, L., Hammell, M., and Dubnau, J. (2012). Transposable elements in TDP-43-mediated neurodegenerative disorders. PloS One *7*, e44099.

Liang, G., Taranova, O., Xia, K., and Zhang, Y. (2010). Butyrate promotes induced pluripotent stem cell generation. J. Biol. Chem. *285*, 25516–25521.

Lienert, F., Mohn, F., Tiwari, V.K., Baubec, T., Roloff, T.C., Gaidatzis, D., Stadler, M.B., and Schübeler, D. (2011). Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate pluripotent from terminally differentiated cells. PLoS Genet. *7*, e1002090.

Lister, R., Pelizzola, M., Kida, Y.S., Hawkins, R.D., Nery, J.R., Hon, G., Antosiewicz-Bourget, J., O'Malley, R., Castanon, R., Klugman, S., et al. (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature *471*, 68–73.

Liu, J., Magri, L., Zhang, F., Marsh, N.O., Albrecht, S., Huynh, J.L., Kaur, J., Kuhlmann, T., Zhang, W., Slesinger, P.A., et al. (2015a). Chromatin landscape defined by repressive histone methylation during oligodendrocyte differentiation. J. Neurosci. Off. J. Soc. Neurosci. *35*, 352–365.

Liu, N., Zhang, Z., Wu, H., Jiang, Y., Meng, L., Xiong, J., Zhao, Z., Zhou, X., Li, J., Li, H., et al. (2015b). Recognition of H3K9 methylation by GLP is required for efficient establishment of H3K9 methylation, rapid target gene repression, and mouse viability. Genes Dev. *29*, 379–393.

Loh, Y.-H., Zhang, W., Chen, X., George, J., and Ng, H.-H. (2007). Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. Genes Dev. *21*, 2545–2557.

Lohmann, F., Loureiro, J., Su, H., Fang, Q., Lei, H., Lewis, T., Yang, Y., Labow, M., Li, E., Chen, T., et al. (2010). KMT1E mediated H3K9 methylation is required for the maintenance of embryonic stem cells by repressing trophectoderm differentiation. Stem Cells Dayt. Ohio *28*, 201–212.

Lomberk, G., Bensi, D., Fernandez-Zapico, M.E., and Urrutia, R. (2006). Evidence for the existence of an HP1-mediated subcode within the histone code. Nat. Cell Biol. *8*, 407–415.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lu, J., and Gilbert, D.M. (2007). Proliferation-dependent and cell cycle regulated transcription of mouse pericentric heterochromatin. J. Cell Biol. *179*, 411–421.

Luo, M., Ling, T., Xie, W., Sun, H., Zhou, Y., Zhu, Q., Shen, M., Zong, L., Lyu, G., Zhao, Y., et al. (2013). NuRD blocks reprogramming of mouse somatic cells into pluripotent stem cells. Stem Cells Dayt. Ohio *31*, 1278–1286.

Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., et al. (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell *1*, 55–70.

Maison, C., Bailly, D., Peters, A.H.F.M., Quivy, J.-P., Roche, D., Taddei, A., Lachner, M., Jenuwein, T., and Almouzni, G. (2002). Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. Nat. Genet. *30*, 329–334.

Maison, C., Bailly, D., Roche, D., Montes de Oca, R., Probst, A.V., Vassias, I., Dingli, F., Lombard, B., Loew, D., Quivy, J.-P., et al. (2011). SUMOylation promotes de novo targeting of HP1α to pericentric heterochromatin. Nat. Genet. *43*, 220–227.

Mali, P., Chou, B.-K., Yen, J., Ye, Z., Zou, J., Dowey, S., Brodsky, R.A., Ohm, J.E., Yu, W., Baylin, S.B., et al. (2010). Butyrate greatly enhances derivation of human induced pluripotent stem cells by promoting epigenetic remodeling and the expression of pluripotency-associated genes. Stem Cells Dayt. Ohio *28*, 713–720.

Mansour, A.A., Gafni, O., Weinberger, L., Zviran, A., Ayyash, M., Rais, Y., Krupalnik, V., Zerbib, M., Amann-Zalcenstein, D., Maza, I., et al. (2012). The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. Nature *488*, 409–413.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343–349.

Marmorstein, R. (2003). Structure of SET domain proteins: a new twist on histone methylation. Trends Biochem. Sci. *28*, 59–62.

Marro, S., Pang, Z.P., Yang, N., Tsai, M.-C., Qu, K., Chang, H.Y., Südhof, T.C., and Wernig, M. (2011). Direct lineage conversion of terminally differentiated hepatocytes to functional neurons. Cell Stem Cell *9*, 374–382.

Martens, J.H.A., O'Sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. EMBO J. *24*, 800–812.

Mathur, M., Tucker, P.W., and Samuels, H.H. (2001). PSF is a novel corepressor that mediates its effect through Sin3A and the DNA binding domain of nuclear hormone receptors. Mol. Cell. Biol. *21*, 2298–2311.

Matoba, S., Liu, Y., Lu, F., Iwabuchi, K.A., Shen, L., Inoue, A., and Zhang, Y. (2014). Embryonic development following somatic cell nuclear transfer impeded by persisting histone methylation. Cell *159*, 884–895.

Matsumura, Y., Nakaki, R., Inagaki, T., Yoshida, A., Kano, Y., Kimura, H., Tanaka, T., Tsutsumi, S., Nakao, M., Doi, T., et al. (2015). H3K4/H3K9me3 Bivalent Chromatin Domains Targeted by Lineage-Specific DNA Methylation Pauses Adipocyte Differentiation. Mol. Cell *60*, 584–596.

Matsunaga, S., Takata, H., Morimoto, A., Hayashihara, K., Higashi, T., Akatsuchi, K., Mizusawa, E., Yamakawa, M., Ashida, M., Matsunaga, T.M., et al. (2012). RBMX: a

regulator for maintenance and centromeric protection of sister chromatid cohesion. Cell Rep. *1*, 299–308.

Meister, P., Mango, S.E., and Gasser, S.M. (2011). Locking the genome: nuclear organization and cell fate. Curr. Opin. Genet. Dev. *21*, 167–174.

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

Mieczkowski, J., Cook, A., Bowman, S.K., Mueller, B., Alver, B.H., Kundu, S., Deaton, A.M., Urban, J.A., Larschan, E., Park, P.J., et al. (2016). MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. Nat. Commun. *7*, 11485.

Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. Nature *454*, 49–55.

Millanes-Romero, A., Herranz, N., Perrera, V., Iturbide, A., Loubat-Casanovas, J., Gil, J., Jenuwein, T., García de Herreros, A., and Peiró, S. (2013). Regulation of heterochromatin transcription by Snail1/LOXL2 during epithelial-to-mesenchymal transition. Mol. Cell *52*, 746–757.

Montoya-Durango, D.E., Liu, Y., Teneng, I., Kalbfleisch, T., Lacy, M.E., Steffen, M.C., and Ramos, K.S. (2009). Epigenetic control of mammalian LINE-1 retrotransposon by retinoblastoma proteins. Mutat. Res. *665*, 20–28.

Moretti, A., Bellin, M., Welling, A., Jung, C.B., Lam, J.T., Bott-Flügel, L., Dorn, T., Goedel, A., Höhnke, C., Hofmann, F., et al. (2010). Patient-specific induced pluripotent stem-cell models for long-QT syndrome. N. Engl. J. Med. *363*, 1397–1409.

Morris, S.A., Cahan, P., Li, H., Zhao, A.M., San Roman, A.K., Shivdasani, R.A., Collins, J.J., and Daley, G.Q. (2014). Dissecting engineered cell types and enhancing cell fate conversion via CellNet. Cell *158*, 889–902.

Muchardt, C., Guilleme, M., Seeler, J.-S., Trouche, D., Dejean, A., and Yaniv, M. (2002). Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1alpha. EMBO Rep. *3*, 975–981.

Müller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J.A. (2002). Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell *111*, 197–208.

Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science *292*, 110–113.

Nielsen, S.J., Schneider, R., Bauer, U.M., Bannister, A.J., Morrison, A., O'Carroll, D., Firestein, R., Cleary, M., Jenuwein, T., Herrera, R.E., et al. (2001). Rb targets histone H3 methylation and HP1 to promoters. Nature *412*, 561–565.

Nozawa, R.-S., Nagao, K., Igami, K.-T., Shibata, S., Shirai, N., Nozaki, N., Sado, T., Kimura, H., and Obuse, C. (2013). Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway. Nat. Struct. Mol. Biol. *20*, 566–573.

van Oevelen, C., Kallin, E.M., and Graf, T. (2013). Transcription factor-induced enhancer modulations during cell fate conversions. Curr. Opin. Genet. Dev. *23*, 562–567.

O'Geen, H., Squazzo, S.L., Iyengar, S., Blahnik, K., Rinn, J.L., Chang, H.Y., Green, R., and Farnham, P.J. (2007). Genome-wide analysis of KAP1 binding suggests autoregulation of KRAB-ZNFs. PLoS Genet. *3*, e89.

O'Keefe, R.T., Henderson, S.C., and Spector, D.L. (1992). Dynamic organization of DNA replication in mammalian cell nuclei: spatially and temporally defined replication of chromosome-specific alpha-satellite DNA sequences. J. Cell Biol. *116*, 1095–1110.

Onder, T.T., Kara, N., Cherry, A., Sinha, A.U., Zhu, N., Bernt, K.M., Cahan, P., Marcarci, B.O., Unternaehrer, J., Gupta, P.B., et al. (2012). Chromatin-modifying enzymes as modulators of reprogramming. Nature *483*, 598–602.

Papp, B., and Plath, K. (2013). Epigenetics of reprogramming to induced pluripotency. Cell *152*, 1324–1343.

Pardue, M.L., and Gall, J.G. (1970). Chromosomal localization of mouse satellite DNA. Science *168*, 1356–1358.

Pasque, V., Jullien, J., Miyamoto, K., Halley-Stott, R.P., and Gurdon, J.B. (2011a). Epigenetic factors influencing resistance to nuclear reprogramming. Trends Genet. TIG *27*, 516–525.

Pasque, V., Gillich, A., Garrett, N., and Gurdon, J.B. (2011b). Histone variant macroH2A confers resistance to nuclear reprogramming. EMBO J. *30*, 2373–2387.

Pasque, V., Radzisheuskaya, A., Gillich, A., Halley-Stott, R.P., Panamarova, M., Zernicka-Goetz, M., Surani, M.A., and Silva, J.C.R. (2012). Histone variant macroH2A marks embryonic differentiation in vivo and acts as an epigenetic barrier to induced pluripotency. J. Cell Sci. *125*, 6094–6104.

Pauler, F.M., Sloane, M.A., Huang, R., Regha, K., Koerner, M.V., Tamir, I., Sommer, A., Aszodi, A., Jenuwein, T., and Barlow, D.P. (2009). H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. Genome Res. *19*, 221–233.

Peng, J.C., and Karpen, G.H. (2008). Epigenetic regulation of heterochromatic DNA stability. Curr. Opin. Genet. Dev. *18*, 204–211.

Pereira, C.F., Piccolo, F.M., Tsubouchi, T., Sauer, S., Ryan, N.K., Bruno, L., Landeira, D., Santos, J., Banito, A., Gil, J., et al. (2010). ESCs require PRC2 to direct the successful reprogramming of differentiated cells toward pluripotency. Cell Stem Cell *6*, 547–556.

Peters, A.H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schöfer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A., et al. (2001). Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell *107*, 323–337.

Peters, A.H.F.M., Mermoud, J.E., O'Carroll, D., Pagani, M., Schweizer, D., Brockdorff, N., and Jenuwein, T. (2002). Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. Nat. Genet. *30*, 77–80.

Peters, A.H.F.M., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A.H.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., et al. (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Mol. Cell *12*, 1577–1589.

Piacentini, L., Fanti, L., Berloco, M., Perrini, B., and Pimpinelli, S. (2003). Heterochromatin protein 1 (HP1) is associated with induced gene expression in Drosophila euchromatin. J. Cell Biol. *161*, 707–714.

Piacentini, L., Fanti, L., Negri, R., Del Vescovo, V., Fatica, A., Altieri, F., and Pimpinelli, S. (2009). Heterochromatin protein 1 (HP1a) positively regulates euchromatic gene expression through RNA transcript association and interaction with hnRNPs in Drosophila. PLoS Genet. *5*, e1000670.

Pimpinelli, S., Berloco, M., Fanti, L., Dimitri, P., Bonaccorsi, S., Marchetti, E., Caizzi, R., Caggese, C., and Gatti, M. (1995). Transposable elements are stable structural components of Drosophila melanogaster heterochromatin. Proc. Natl. Acad. Sci. U. S. A. *92*, 3804–3808.

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. Science *300*, 131–135.

Poleshko, A., Mansfield, K.M., Burlingame, C.C., Andrake, M.D., Shah, N.R., and Katz, R.A. (2013). The human protein PRR14 tethers heterochromatin to the nuclear lamina during interphase and mitotic exit. Cell Rep. *5*, 292–301.

Politz, J.C.R., Scalzo, D., and Groudine, M. (2013). Something silent this way forms: the functional organization of the repressive nuclear compartment. Annu. Rev. Cell Dev. Biol. *29*, 241–270.

Polo, J.M., Anderssen, E., Walsh, R.M., Schwarz, B.A., Nefzger, C.M., Lim, S.M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J., et al. (2012). A molecular roadmap of reprogramming somatic cells into iPS cells. Cell *151*, 1617–1632.

Pope, B.D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D.L., Wang, Y., Hansen, R.S., Canfield, T.K., et al. (2014). Topologically associating domains are stable units of replication-timing regulation. Nature *515*, 402–405.

Probst, A.V., and Almouzni, G. (2008). Pericentric heterochromatin: dynamic organization during early development in mammals. Differ. Res. Biol. Divers. *76*, 15–23.

Probst, A.V., Okamoto, I., Casanova, M., Marjou, F. El, Le Baccon, P., and Almouzni, G. (2010). A strand-specific burst in transcription of pericentric satellites is required for chromocenter formation and early mouse development. Dev. Cell *19*, 625–638.

Qin, H., Diaz, A., Blouin, L., Lebbink, R.J., Patena, W., Tanbun, P., LeProust, E.M., McManus, M.T., Song, J.S., and Ramalho-Santos, M. (2014). Systematic identification of barriers to human iPSC generation. Cell *158*, 449–461.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinforma. Oxf. Engl. *26*, 841–842.

Ragunathan, K., Jih, G., and Moazed, D. (2015). Epigenetics. Epigenetic inheritance uncoupled from sequence-specific recruitment. Science *348*, 1258699.

Rais, Y., Zviran, A., Geula, S., Gafni, O., Chomsky, E., Viukov, S., Mansour, A.A., Caspi, I., Krupalnik, V., Zerbib, M., et al. (2013). Deterministic direct reprogramming of somatic cells to pluripotency. Nature *502*, 65–70.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., et al. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593–599.

Rego, A., Sinclair, P.B., Tao, W., Kireev, I., and Belmont, A.S. (2008). The facultative heterochromatin of the inactive X chromosome has a distinctive condensed ultrastructure. J. Cell Sci. *121*, 1119–1127.

Reyes-Turcu, F.E., Zhang, K., Zofall, M., Chen, E., and Grewal, S.I.S. (2011). Defects in RNA quality control factors reveal RNAi-independent nucleation of heterochromatin. Nat. Struct. Mol. Biol. *18*, 1132–1138.

Richards, E.J., and Elgin, S.C.R. (2002). Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. Cell *108*, 489–500.

Riddle, N.C., Jung, Y.L., Gu, T., Alekseyenko, A.A., Asker, D., Gui, H., Kharchenko, P.V., Minoda, A., Plachetka, A., Schwartz, Y.B., et al. (2012). Enrichment of HP1a on

Drosophila chromosome 4 genes creates an alternate chromatin structure critical for regulation in this heterochromatic domain. PLoS Genet. *8*, e1002954.

Risca, V.I., Denny, S.K., Straight, A.F., and Greenleaf, W.J. (2016). Variable chromatin structure revealed by in situ spatially correlated DNA cleavage mapping. Nature.

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.

Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P.V., Layard-Liesching, H., Verp, S., Marquis, J., et al. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. Nature *463*, 237–240.

Ryan, R.F., Schultz, D.C., Ayyanathan, K., Singh, P.B., Friedman, J.R., Fredericks, W.J., and Rauscher, F.J. (1999). KAP-1 corepressor protein interacts and colocalizes with heterochromatic and euchromatic HP1 proteins: a potential role for Krüppel-associated box-zinc finger proteins in heterochromatin-mediated gene silencing. Mol. Cell. Biol. *19*, 4366–4378.

Sadoni, N., Langer, S., Fauth, C., Bernardi, G., Cremer, T., Turner, B.M., and Zink, D. (1999). Nuclear organization of mammalian genomes. Polar chromosome territories build up functionally distinct higher order compartments. J. Cell Biol. *146*, 1211–1226.

Sahi, J., Grepper, S., and Smith, C. (2010). Hepatocytes as a tool in drug metabolism, transport and safety evaluations in drug discovery. Curr. Drug Discov. Technol. *7*, 188–198.

Saksouk, N., Barth, T.K., Ziegler-Birling, C., Olova, N., Nowak, A., Rey, E., Mateos-Langerak, J., Urbach, S., Reik, W., Torres-Padilla, M.-E., et al. (2014). Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation. Mol. Cell *56*, 580–594.

Saksouk, N., Simboeck, E., and Déjardin, J. (2015). Constitutive heterochromatin formation and transcription in mammals. Epigenetics Chromatin *8*, 3.

Santenard, A., Ziegler-Birling, C., Koch, M., Tora, L., Bannister, A.J., and Torres-Padilla, M.-E. (2010). Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3. Nat. Cell Biol. *12*, 853–862.

Santoro, R., Li, J., and Grummt, I. (2002). The nucleolar remodeling complex NoRC mediates heterochromatin formation and silencing of ribosomal gene transcription. Nat. Genet. *32*, 393–396.

Savić, N., Bär, D., Leone, S., Frommel, S.C., Weber, F.A., Vollenweider, E., Ferrari, E., Ziegler, U., Kaech, A., Shakhova, O., et al. (2014). lncRNA maturation to initiate heterochromatin formation in the nucleolus is required for exit from pluripotency in ESCs. Cell Stem Cell *15*, 720–734.

Schotta, G., Ebert, A., Krauss, V., Fischer, A., Hoffmann, J., Rea, S., Jenuwein, T., Dorn, R., and Reuter, G. (2002). Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. EMBO J. *21*, 1121–1131.

Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., and Rauscher, F.J. (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. Genes Dev. *16*, 919–932.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342.

Schwartz, Y.B., Kahn, T.G., Nix, D.A., Li, X.-Y., Bourgon, R., Biggin, M., and Pirrotta, V. (2006). Genome-wide analysis of Polycomb targets in Drosophila melanogaster. Nat. Genet. *38*, 700–705.

Seki, Y., Kurisaki, A., Watanabe-Susaki, K., Nakajima, Y., Nakanishi, M., Arai, Y., Shiota, K., Sugino, H., and Asashima, M. (2010). TIF1beta regulates the pluripotency of embryonic stem cells in a phosphorylation-dependent manner. Proc. Natl. Acad. Sci. U. S. A. *107*, 10926–10931.

Sekiya, S., and Suzuki, A. (2011). Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. Nature *475*, 390–393.

Shi, Y., Do, J.T., Desponts, C., Hahm, H.S., Schöler, H.R., and Ding, S. (2008). A combined chemical and genetic approach for the generation of induced pluripotent stem cells. Cell Stem Cell *2*, 525–528.

Shu, J., Wu, C., Wu, Y., Li, Z., Shao, S., Zhao, W., Tang, X., Yang, H., Shen, L., Zuo, X., et al. (2013). Induction of pluripotency in mouse somatic cells with lineage specifiers. Cell *153*, 963–975.

Slotkin, R.K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. Nat. Rev. Genet. *8*, 272–285.

Snitow, M., Lu, M., Cheng, L., Zhou, S., and Morrisey, E.E. (2016). Ezh2 restricts the smooth muscle lineage during mouse lung mesothelial development. Dev. Camb. Engl. *143*, 3733–3741.

Soldi, M., and Bonaldi, T. (2013). The proteomic investigation of chromatin functional domains reveals novel synergisms among distinct heterochromatin components. Mol. Cell. Proteomics MCP *12*, 764–780.

Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell *151*, 994–1004.

Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M., and Zaret, K.S. (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell *161*, 555–568.

Sridharan, R., Tchieu, J., Mason, M.J., Yachechko, R., Kuoy, E., Horvath, S., Zhou, Q., and Plath, K. (2009). Role of the murine reprogramming factors in the induction of pluripotency. Cell *136*, 364–377.

Sridharan, R., Gonzales-Cope, M., Chronis, C., Bonora, G., McKee, R., Huang, C., Patel, S., Lopez, D., Mishra, N., Pellegrini, M., et al. (2013). Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1γ in reprogramming to pluripotency. Nat. Cell Biol. *15*, 872–882.

Stadtfeld, M., Maherali, N., Breault, D.T., and Hochedlinger, K. (2008). Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. Cell Stem Cell *2*, 230–240.

Stock, J.K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A.G., and Pombo, A. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nat. Cell Biol. *9*, 1428–1435.

Tachibana, M., Sugimoto, K., Fukushima, T., and Shinkai, Y. (2001). Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. J. Biol. Chem. *276*, 25309–25317.

Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H., et al. (2002). G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes Dev. *16*, 1779–1791.

Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T., Iwanari, H., Sakihama, T., Kodama, T., Hamakubo, T., and Shinkai, Y. (2005). Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. Genes Dev. *19*, 815–826.

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

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861–872.

Tanabe, K., Nakamura, M., Narita, M., Takahashi, K., and Yamanaka, S. (2013). Maturation, not initiation, is the major roadblock during reprogramming toward pluripotency from human fibroblasts. Proc. Natl. Acad. Sci. U. S. A. *110*, 12172–12179.

Tartof, K.D., Hobbs, C., and Jones, M. (1984). A structural basis for variegating position effects. Cell *37*, 869–878.

Tchasovnikarova, I.A., Timms, R.T., Matheson, N.J., Wals, K., Antrobus, R., Göttgens, B., Dougan, G., Dawson, M.A., and Lehner, P.J. (2015). GENE SILENCING. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. Science *348*, 1481–1485.

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

Thompson, P.J., Dulberg, V., Moon, K.-M., Foster, L.J., Chen, C., Karimi, M.M., and Lorincz, M.C. (2015). hnRNP K coordinates transcriptional silencing by SETDB1 in embryonic stem cells. PLoS Genet. *11*, e1004933.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.

Timms, R.T., Tchasovnikarova, I.A., Antrobus, R., Dougan, G., and Lehner, P.J. (2016). ATF7IP-Mediated Stabilization of the Histone Methyltransferase SETDB1 Is Essential for Heterochromatin Formation by the HUSH Complex. Cell Rep. *17*, 653–659.

Ting, D.T., Lipson, D., Paul, S., Brannigan, B.W., Akhavanfard, S., Coffman, E.J., Contino, G., Deshpande, V., Iafrate, A.J., Letovsky, S., et al. (2011). Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. Science *331*, 593–596.

Toh, C.-X.D., Chan, J.-W., Chong, Z.-S., Wang, H.F., Guo, H.C., Satapathy, S., Ma, D., Goh, G.Y.L., Khattar, E., Yang, L., et al. (2016). RNAi Reveals Phase-Specific Global Regulators of Human Somatic Cell Reprogramming. Cell Rep. *15*, 2597–2607.

Toma, J.S., Shettar, B.C., Chipman, P.H., Pinto, D.M., Borowska, J.P., Ichida, J.K., Fawcett, J.P., Zhang, Y., Eggan, K., and Rafuse, V.F. (2015). Motoneurons derived from induced pluripotent stem cells develop mature phenotypes typical of endogenous spinal motoneurons. J. Neurosci. Off. J. Soc. Neurosci. *35*, 1291–1306.

Towbin, B.D., Gonzalez-Sandoval, A., and Gasser, S.M. (2013). Mechanisms of heterochromatin subnuclear localization. Trends Biochem. Sci. *38*, 356–363.

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

Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G., and Reuter, G. (1994). The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. EMBO J. *13*, 3822–3831.

Underwood, J.M., Becker, K., Stein, G.S., and Nickerson, J.A. (2016). The Ultrastructural Signature of Human Embryonic Stem Cells. J. Cell. Biochem.

Vakoc, C.R., Mandat, S.A., Olenchock, B.A., and Blobel, G.A. (2005). Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. Mol. Cell *19*, 381–391.

Vakoc, C.R., Sachdeva, M.M., Wang, H., and Blobel, G.A. (2006). Profile of histone lysine methylation across transcribed mammalian chromatin. Mol. Cell. Biol. *26*, 9185–9195.

Vallier, L. (2014). Heps with pep: direct reprogramming into human hepatocytes. Cell Stem Cell *14*, 267–269.

Vassen, L., Fiolka, K., and Möröy, T. (2006). Gfi1b alters histone methylation at target gene promoters and sites of gamma-satellite containing heterochromatin. EMBO J. *25*, 2409–2419.

Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J.V., Hyman, A.A., Stunnenberg, H.G., et al. (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell *142*, 967–980.

Vierbuchen, T., and Wernig, M. (2012). Molecular roadblocks for cellular reprogramming. Mol. Cell *47*, 827–838.

Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Südhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. Nature *463*, 1035–1041.

Vogel, M.J., Guelen, L., de Wit, E., Peric-Hupkes, D., Lodén, M., Talhout, W., Feenstra, M., Abbas, B., Classen, A.-K., and van Steensel, B. (2006). Human heterochromatin proteins form large domains containing KRAB-ZNF genes. Genome Res. *16*, 1493–1504.

Voigt, P., LeRoy, G., Drury, W.J., Zee, B.M., Son, J., Beck, D.B., Young, N.L., Garcia, B.A., and Reinberg, D. (2012). Asymmetrically modified nucleosomes. Cell *151*, 181–193.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I.S., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833–1837.

Wakayama, T., Perry, A.C., Zuccotti, M., Johnson, K.R., and Yanagimachi, R. (1998). Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. Nature *394*, 369–374.

Wallrath, L.L., and Elgin, S.C. (1995). Position effect variegation in Drosophila is associated with an altered chromatin structure. Genes Dev. *9*, 1263–1277.

Wang, S., Zou, C., Fu, L., Wang, B., An, J., Song, G., Wu, J., Tang, X., Li, M., Zhang, J., et al. (2015). Autologous iPSC-derived dopamine neuron transplantation in a nonhuman primate Parkinson's disease model. Cell Discov. *1*, 15012.

Wen, B., Wu, H., Shinkai, Y., Irizarry, R.A., and Feinberg, A.P. (2009). Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. Nat. Genet. *41*, 246–250.

Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B.E., and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature *448*, 318–324.

Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., and Campbell, K.H. (1997). Viable offspring derived from fetal and adult mammalian cells. Nature *385*, 810–813.

Workman, J.L., and Kingston, R.E. (1998). Alteration of nucleosome structure as a mechanism of transcriptional regulation. Annu. Rev. Biochem. *67*, 545–579.

Wu, S.M., and Hochedlinger, K. (2011). Harnessing the potential of induced pluripotent stem cells for regenerative medicine. Nat. Cell Biol. *13*, 497–505.

Xie, H., Ye, M., Feng, R., and Graf, T. (2004). Stepwise reprogramming of B cells into macrophages. Cell *117*, 663–676.

Xu, C.-R., Li, L.-C., Donahue, G., Ying, L., Zhang, Y.-W., Gadue, P., and Zaret, K.S. (2014a). Dynamics of genomic H3K27me3 domains and role of EZH2 during pancreatic endocrine specification. EMBO J. *33*, 2157–2170.

Xu, M., Zhao, G.-N., Lv, X., Liu, G., Wang, L.Y., Hao, D.-L., Wang, J., Liu, D.-P., and Liang, C.-C. (2014b). CTCF controls HOXA cluster silencing and mediates PRC2-repressive higher-order chromatin structure in NT2/D1 cells. Mol. Cell. Biol. *34*, 3867–3879.

Yamanaka, S., and Blau, H.M. (2010). Nuclear reprogramming to a pluripotent state by three approaches. Nature *465*, 704–712.

Yamashita, K., Sato, A., Asashima, M., Wang, P.-C., and Nishinakamura, R. (2007). Mouse homolog of SALL1, a causative gene for Townes-Brocks syndrome, binds to A/T-rich sequences in pericentric heterochromatin via its C-terminal zinc finger domains. Genes Cells Devoted Mol. Cell. Mech. *12*, 171–182.

Yeap, L.-S., Hayashi, K., and Surani, M.A. (2009). ERG-associated protein with SET domain (ESET)-Oct4 interaction regulates pluripotency and represses the trophectoderm lineage. Epigenetics Chromatin *2*, 12.

Yuan, P., Han, J., Guo, G., Orlov, Y.L., Huss, M., Loh, Y.-H., Yaw, L.-P., Robson, P., Lim, B., and Ng, H.-H. (2009). Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. Genes Dev. *23*, 2507–2520.

Zaret, K.S. (2008). Genetic programming of liver and pancreas progenitors: lessons for stem-cell differentiation. Nat. Rev. Genet. *9*, 329–340.

Zhang, K., Mosch, K., Fischle, W., and Grewal, S.I.S. (2008). Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. Nat. Struct. Mol. Biol. *15*, 381–388.

Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J., and Melton, D.A. (2008). In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. Nature *455*, 627–632.

Zhu, J., Adli, M., Zou, J.Y., Verstappen, G., Coyne, M., Zhang, X., Durham, T., Miri, M., Deshpande, V., De Jager, P.L., et al. (2013). Genome-wide chromatin state transitions associated with developmental and environmental cues. Cell *152*, 642–654.

Ziv, Y., Bielopolski, D., Galanty, Y., Lukas, C., Taya, Y., Schultz, D.C., Lukas, J., Bekker-Jensen, S., Bartek, J., and Shiloh, Y. (2006). Chromatin relaxation in response to DNA double-strand breaks is modulated by a novel ATM- and KAP-1 dependent pathway. Nat. Cell Biol. *8*, 870–876.

Appendix A. Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes

This appendix contains a manuscript to which I contributed during my graduate work, which was published in *Genes and Development* (Vol. 27(3), p. 251-260), and in which I am listed as the third author. The large majority of the work in this manuscript was performed by Juanma Caravaca, Ph.D., under the guidance of Dr. Zaret. I performed the RNA polymerase II immunofluorescence experiments as well as the ethynyl uridine RNA labeling experiments that are presented in Supplementary Figure 5, and which supported the choice of RT-PCR timepoints used in Figure 6. The manuscript was written by Dr. Caravaca and Dr. Zaret, and I assisted in editing the manuscript.

Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes

Juan Manuel Caravaca,[1] Greg Donahue,[1] Justin S. Becker,[1] Ximiao He,[2] Charles Vinson,[2] and Kenneth S. Zaret[1,3]

[1]Epigenetics Program, Department of Cell and Developmental Biology, University of Pennsylvania, Perelman School of Medicine, Smilow Center for Translation Research, Philadelphia, Pennsylvania 19104, USA; [2]National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

While most transcription factors exit the chromatin during mitosis and the genome becomes silent, a subset of factors remains and "bookmarks" genes for rapid reactivation as cells progress through the cell cycle. However, it is unknown whether such bookmarking factors bind to chromatin similarly in mitosis and how different binding capacities among them relate to function. We compared a diverse set of transcription factors involved in liver differentiation and found markedly different extents of mitotic chromosome binding. Among them, the pioneer factor FoxA1 exhibits the greatest extent of mitotic chromosome binding. Genomically, ~15% of the FoxA1 interphase target sites are bound in mitosis, including at genes that are important for liver differentiation. Biophysical, genome mapping, and mutagenesis studies of FoxA1 reveals two different modes of binding to mitotic chromatin. Specific binding in mitosis occurs at sites that continue to be bound from interphase. Nonspecific binding in mitosis occurs across the chromosome due to the intrinsic chromatin affinity of FoxA1. Both specific and nonspecific binding contribute to timely reactivation of target genes post-mitosis. These studies reveal an unexpected diversity in the mechanisms by which transcription factors help retain cell identity during mitosis.

[*Keywords*: mitosis; bookmarking; pioneer factor; FoxA1; chromatin; chromosomes]

Supplemental material is available for this article.

Received September 21, 2012; revised version accepted December 26, 2012.

When cells enter mitosis, chromosomes condense (Caravaca et al. 2005) and the genome becomes silent (Prescott and Bender 1962; Spencer et al. 2000). Only a fraction of transcription factors are retained on mitotic chromosomes (Martinez-Balbas et al. 1995; Michelotti et al. 1997; Burke et al. 2005; Yan et al. 2006; Egli et al. 2008), and a subset of these facilitate rapid gene reactivation post-mitosis (Young et al. 2007; Blobel et al. 2009; Dey et al. 2009; Kadauke et al. 2012). The basis for the marked differential in transcription factors' binding to mitotic chromatin and how it reflects the factors' roles in genome reactivation is not clear.

In liver development, binding sites for FoxA and GATA factors are occupied on the silent liver gene *alb1* in undifferentiated embryonic endoderm (Gualdi et al. 1996; Bossard and Zaret 1998). Upon hepatic induction, nearby binding sites for NF-1, C/EBP, and other factors become occupied, and the liver gene is activated. Among the factors that promote liver development, only FoxA proteins can bind their target sites on nucleosomes and enable the other factors to bind (Cirillo and Zaret 1999; Cirillo et al. 2002); hence, FoxA factors have been called "pioneer factors" (Cirillo et al. 2002; Zaret and Carroll 2011). While GATA4 is dependent on FoxA for binding nucleosomes (Cirillo and Zaret 1999), it can bind to compacted chromatin that is inaccessible to the other factors (Cirillo et al. 2002) and hence can be considered a subordinate pioneer factor. The structure of the DNA-binding domain (DBD) of FoxA resembles that of linker histone (Clark et al. 1993; Ramakrishnan et al. 1993), and the FoxA C-terminal domain, which is unlike that of linker histone, interacts with core histones and promotes local chromatin opening (Cirillo et al. 2002). The highly related FoxA1 and FoxA2 are encoded by unlinked genes and both are necessary for the activation of the hepatic program (Lee et al. 2005); FoxA1 has been shown to remain bound to mitotic chromosomes in adult liver cells (Zaret et al. 2011). We therefore sought to investigate the mechanism and

[3]**Corresponding author**
E-mail zaret@upenn.edu
Article published online ahead of print. Article and publication date are online at http://www.genesdev.org/cgi/doi/10.1101/gad.206458.112.

role of FoxA binding to the mitotic genome in hepatic cells.

Results

Pioneer factors bind more strongly than other factors to mitotic chromatin

We previously assessed the interphase chromatin binding and mobility of GFP-tagged versions of FoxA1, GATA4, C/EBPα, NF-1, and other proteins that are expressed in the liver and contain different DBD structures (HMGB1, c-Myc, and linker histone H1o). Notably, FoxA1 moves much more slowly in chromatin than the other factors, correlating with its high nucleosomal binding ability, although not as slow as H1o (Sekiya et al. 2009). Here, we expressed the constructs in HUH7 adult human hepatoma cells that had been blocked in mitosis with nocodazole and visualized GFP fluorescence in live cells by high-resolution deconvolution microscopy (Agard 1984). GFP-FoxA1 was seen almost exclusively bound to chromosomes in the metaphase-arrested cells as well as in drug-free control cells passing through mitosis, mimicking the pattern of GFP-H1o (Fig. 1A). We estimate that the GFP transfected cells expressed ~10-fold more of the respective amounts of the transcription factor than the endogenous protein (data not shown). When we used 20-fold lower amounts of transfected GFP-FoxA1 plasmid DNA, we observed much fainter signals but still marked binding to mitotic chromosomes (Supplemental Fig. 1a). GFP-GATA4 and GFP-HMGB1 fluorescence was seen both on the mitotic chromosomes and in the cytoplasm (Fig. 1B), while GFP-C/EBPα gave fainter signals on mitotic chromosomes than the other factors. Western blotting of endogenous C/EBPα showed it to be several-fold less stable in mitotic hepatoma cells, whereas FoxA1 was equal in abundance in mitotic and asynchronous cells (Supplemental Fig. 1b). GFP-c-Myc and GFP-NF1 exhibited background fluorescence on the mitotic chromosomes, reflecting factor exclusion (Fig. 1B). Cells released from the metaphase mitotic block and fixed at anaphase and telophase showed that GFP-FoxA1 remained bound to chromosomes throughout mitosis, while a GFP protein fused to a nuclear localization sequence was excluded (Supplemental Fig. 2a). Endogenous FoxA1 exhibited similar properties but with a more diffuse signal that is typical of fixed cells, compared with that seen when live cells are imaged (Supplemental Fig. 2b). GFP fused to the FoxA1 DBD was sufficient to bind mitotic chromosomes (Fig. 1A).

Thus, within this group of regulatory factors for the liver lineage, the earliest pioneer factor in development is the most strongly bound to mitotic chromatin; the subordinate pioneer factor GATA4 is partially bound, similar to that seen for GATA1 elsewhere (Kadauke et al. 2012); and later developmental factors are either partially bound, bound yet unstable in mitosis, or excluded from the mitotic chromosomes altogether.

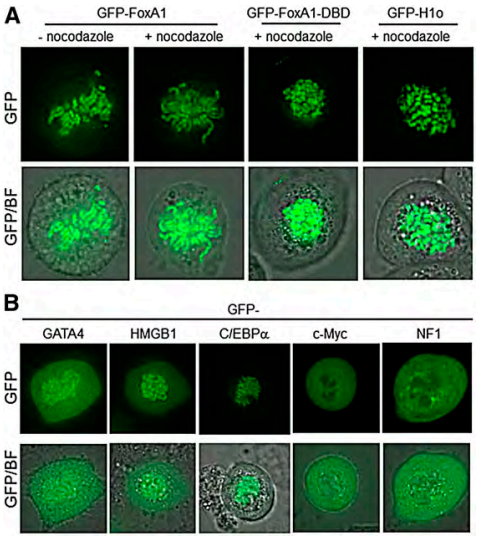


Figure 1. Diverse modes of hepatic transcription factor binding to mitotic chromosomes. (*A*,*B*) GFP fluorescence in live HUH7 hepatoma cells visualized by deconvolution microscopy with or without nocodazole for mitotic arrest. (*A*) GFP and GFP/bright-field (BF) views showing that GFP-FoxA1, GFP-FoxA1 DBD, and GFP-H1o remain quantitatively bound to mitotic chromosomes. (*B*) A portion of the cellular GATA4 and HMGB1 is released from mitotic chromosomes in nocodazole-treated cells; C/EBPα becomes unstable, yet a portion remains bound; and c-Myc and NF1 are excluded from the mitotic chromosomes.

Highly transcribed gene targets of FoxA1 remain specifically bound in mitosis

To assess where endogenous FoxA1 binds the genome in mitotic cells compared with asynchronously cycling cells, we performed chromatin immunoprecipitation (ChIP) coupled with deep sequencing (ChIP-seq) on triplicate cell populations, with and without nocodazole treatment (Supplemental Fig. 3a,b). More than 94% of the nocodazole-arrested cells were in mitosis, as assessed by cell morphology and H3Ser10 phosphorylation, while <2% of the asynchronously cycling wells were in mitosis (Supplemental Fig. 3a). Using model-based analysis of ChIP-seq (MACS) (Zhang et al. 2008) to assign peaks, we discovered 546 FoxA1-bound sites in mitotic cells and 3509 sites in asynchronous cells (Fig. 2A). Eighty-seven of the called FoxA1 peaks were specific to mitotic cells, but visual inspection of the unique sequence reads and quantitative PCR (qPCR) analysis of independent chromatin preparations indicated that the FoxA1-bound sites in mitotic cells are also bound in asynchronous cells (e.g., Fig. 2B [intergenic site on ch. 5 and *MIPEP* site at +92.7 kb, red arrowheads], B [ChIP-qPCR validations are of mitotic and asynchronous cell chromatin]). Certain weak FoxA1 peaks in mitosis that were not called by MACS were significant by ChIP-qPCR (Fig. 2B, *AFP* −4.1- and −2-kb sites and *TTR* promoter), similar to the weaker mitotic site at the *HNF4a* −7-kb site and unlike a negative control site on ch. 13 (Fig. 2B). ChIP for histone H3 at

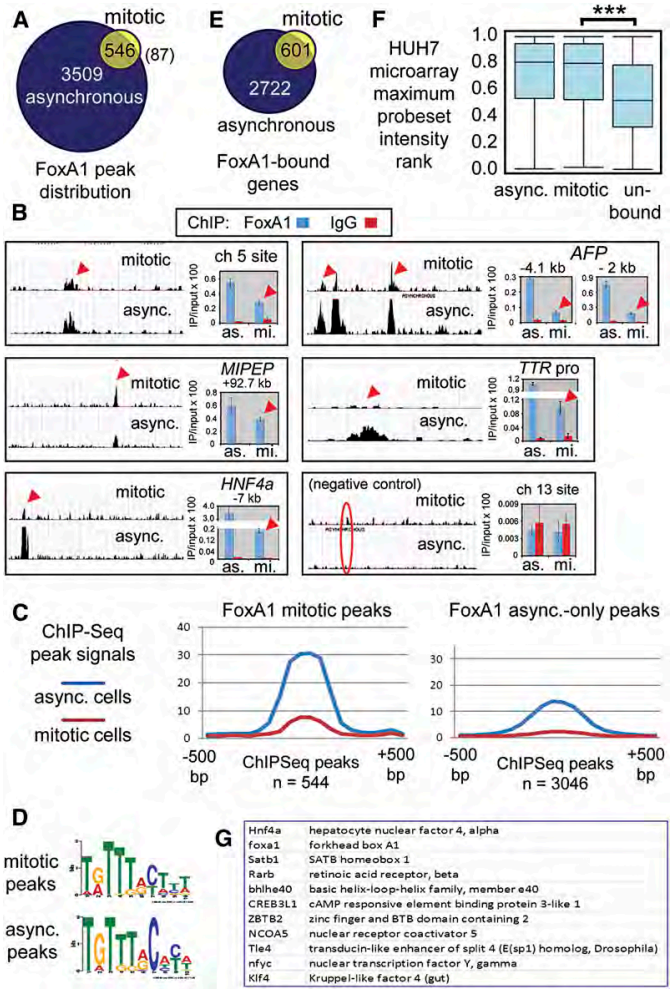


Figure 2. FoxA1 in mitosis occupies the most strongly expressed and strongly bound genes in asynchronous HUH7 cells. (*A*) Peaks were pooled from three replicate ChIP-seq samples, revealing 3509 asynchronous FoxA1-binding sites, of which 546 are also bound in mitosis. (*B*) FoxA1 ChIP-seq data tracks and ChIP-qPCR confirmation in independent chromatin samples with signals normalized to input and per million sequence reads. Red arrowheads depict sites of mitotic binding, and those that were verified by qPCR are shown on the *right*. The circle indicates the negative control site. The break in the *HNF4a* and TTR bar graphs is to accommodate different scales for the asynchronous (as.) and mitotic (mi.) data. (*C*) FoxA1 ChIP-seq signals over all sites bound in mitotic and asynchronous cells compared with all sites bound only in asynchronous cells; the signal is normalized to input DNA quantity and the total number of aligned sequence reads. FoxA1 binding is much stronger to sites in mitotic and asynchronous cells than to sites bound in asynchronous cells only. (*D*) Common FoxA-binding motif at sites bound in mitosis versus asynchronous only. (*E*) Nearly all of the FoxA1-bound sites in mitosis are associated with genes either within the transcribed region or <20 kb upstream. (*F*) Box and whisker plots showing that genes bound by FoxA1 in mitotic and asynchronous cells correspond to those more highly expressed genes in hepatoma cells; (***) $P < 10^{-15}$. (*G*) FoxA1 remains bound to hepatic transcription factor genes in mitosis.

these sites showed comparable signals between the asynchronous and the mitotic chromatin (Supplemental Fig. 3c), demonstrating that differences seen for FoxA1 were specific to the factor and not the preparations of chromatin.

FoxA1 peaks that occurred in both mitotic and asynchronous cells were, on average, markedly stronger in asynchronous cells than the FoxA1 peaks that occurred only in asynchronous cells (Fig. 2C, where overlapping peaks were merged). De novo motif analysis of FoxA1 peaks in mitosis revealed a Fox consensus sequence that was essentially the same as that seen at asynchronous-only sites (Fig. 2D). We found 601 genes bound by FoxA1 in both asynchronous and mitotic cells and 2722 genes bound by FoxA1 only in asynchronous cells (Fig. 2E). Triplicate microarray analyses of HUH7 cells showed that mitotic and interphase FoxA1 target genes represent those that are among the most highly expressed in interphase (Fig. 2F). Mitotic FoxA1 targets include transcription factor genes that are important for hepatic differentiation, such as *HNF4a* and *FOXA1* itself (Fig. 2G), and genes for kinase signaling pathways (Supplemental Fig. 3d). In summary, we estimate that ~15% of the FoxA1 sites that are bound in interphase cells are also bound in mitosis, corresponding to the strongest bound sites at the more highly expressed genes in interphase.

FoxA1 target sites that remain bound in mitosis have a higher intrinsic nucleosome occupancy score (INOS) than sites bound only in asynchronous cells

To further characterize potential differences between target sites for FoxA1 that remain bound in mitosis compared with sites bound only in asynchronous cells, we first compared the binding events with available histone modification data in the human ENCODE database. No substantial differences were observed to be centered over the FoxA1-binding sites (data not shown). Given the extensive prior data documenting the ability

of FoxA1 to bind nucleosomal DNA (Cirillo et al. 1998; Cirillo and Zaret 1999; Chaya et al. 2001), we compared the FoxA1 mitotic and asynchronous binding data with a computational model that predicts how well a nucleosome could form at underlying 147-base-pair (bp) lengths of DNA (Kaplan et al. 2009; Tillo and Hughes 2009; Tillo et al. 2010). Figure 3 shows the average INOS for FoxA1 ChIP-seq peaks in mitotic and asynchronous cells across 1500 bp spanning each binding event. In both cases, there is a peak in INOS near the center of DNA binding. Peaks in INOS of ~300 bp have been observed at binding events for other transcription factors (Tillo et al. 2010). A difference here is that the absolute value of INOS for FoxA1 is low, indicating that the protein does not generally bind to CG-rich regions of the genome. Indeed, the target sequence for FoxA1 is relatively AT-rich (Fig. 2D).

Interestingly, the INOS for FoxA1 in mitotic cells is significantly higher than that in asynchronous cells ($P = 9.5 \times 10^{-6}$). The higher INOS for mitotic peaks suggests that the underlying nucleosome is typically more stable at chromosomal sites where FoxA1 remains bound in mitotic cells compared with that seen at sites where FoxA1 binds only in asynchronous cells. This possibility is consistent with the aforementioned features of FoxA1 binding its target site on nucleosomes in vitro and in vivo and suggests that more stable nucleosomal targets help predict mitotic chromosome binding.

Substantial nonspecific mitotic chromosome binding by FoxA1

We next addressed the apparent conundrum that FoxA1 is stable in mitosis (Supplemental Fig. 1b) and retained globally on the chromosomes (Fig. 1A) yet dissociates from many specific interphase binding sites (Fig. 2A). To investigate changes in the chromatin-binding properties of FoxA1, we performed fluorescence recovery after photobleaching (FRAP) assays in unfixed cells on GFP-FoxA1, with GFP-H1o as a control. GFP-H1o in HUH7 cells exhibited a threefold increase in FRAP half-time in metaphase compared with interphase (Fig. 4A); a similar increase was seen for H1c-GFP in other cells (Chen et al. 2005). Thus, linker histone moves more slowly in metaphase chromatin, consistent with the chromosomes' high degree of compaction. In marked contrast, GFP-FoxA1 exhibited a more than twofold decrease in FRAP half-time in metaphase (Fig. 4A). Thus, the dramatic compaction of chromatin in mitosis is associated with an increase in the mobility of FoxA1 compared with interphase.

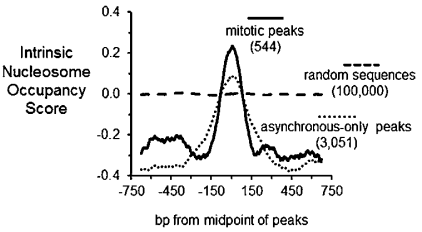


Figure 3. FoxA1-bound sites in mitotic cells have higher predicted nucleosome occupancy compared with sites bound only in asynchronous cells. Average INOS for FoxA1 3509 ChIP-seq peaks seen only in asynchronous HuH7 cells and 544 peaks seen in mitotic cells within ±750 bp from the center of the peak; also shown are 100,000 sequences selected at random from human genome (hg18). The average INOS profiling of FoxA1 in mitotic HuH7 cells is higher than seen in asynchronous cells (*t*-test, $P = 9.5 \times 10^{-6}$).

We questioned whether the increased mobility of FoxA1 in mitosis could be visualized in the genomic ChIP-seq data as an increase in nonspecific DNA-binding signals. Indeed, when viewed at the chromosomal level, there was a greater background of nonpeak FoxA1 ChIP-seq reads in mitotic cells compared with asynchronous cells (Fig. 4B, blue lines below red arrowheads; Supplemental Fig. 4). Thus, the increased mobility of FoxA1 in mitosis is associated with significant amounts of FoxA redistributing from specific sites to the flanking chromosomal domains.

We questioned whether the nonspecific binding at the chromosomal level in mitotic cells was dependent on FoxA1 interactions with DNA. To address this, we transfected HUH7 cells with GFP-tagged variants of FoxA1 that perturb nonspecific DNA binding (FoxA1-RR) or specific DNA binding (FoxA1-NH) (Sekiya et al. 2009). FoxA1-RR has alanine substitutions at two residues that make phosphate contacts with the DNA backbone (Fig. 5A). The FoxA1-RR mutant exhibits a marked decrease in overall and nonspecific DNA binding but still recognizes FoxA target sites, albeit more weakly than wild type (Sekiya et al. 2009). FoxA1-NH has alanine substitutions at two residues that make base contacts with DNA (Fig. 5A). The FoxA1-NH mutant exhibits normal nonspecific DNA binding but no longer recognizes FoxA target sites (Sekiya et al. 2009). Strikingly, when observed in live mitotic cells, much of the GFP-FoxA1-RR mutant was cytoplasmic (Fig. 5B). In contrast, the GFP-FoxA1-NH mutant was largely retained on the mitotic chromosomes (Fig. 5B), demonstrating that specific binding is not necessary for bulk mitotic retention. From these data, we conclude that much of the FoxA1 in mitotic chromatin is bound nonspecifically to the DNA.

To assess the contributions of specific and nonspecific binding to target site binding in mitosis, we performed ChIP for the GFP tag in cells transfected with GFP-FoxA1 and the variants (Fig. 5C; Supplemental Fig. 5a). As expected, wild-type GFP-FoxA1 bound to endogenous FoxA1 targets in both asynchronous and mitotic cells (Supplemental Fig. 5b). Of three target sites that possess a FoxA-binding motif, binding by the GFP-FoxA1-RR mutant at the *AFP* −4.1-kb site and the *TTR* promoter was diminished, but still significant, in comparison with the IgG ChIP control (Fig. 5C, red arrowheads). In contrast, the GFP-FoxA-NH mutant exhibited a loss of binding to all of these sites (Fig. 5C). Therefore, despite a loss in nonspecific DNA binding and significant release from chromosomes, the FoxA-RR mutant can still bind in mitosis to sites with FoxA1 motifs. We conclude that binding by FoxA1 in mitotic chromosomes can take two forms: nonspecific binding that is not reflected in specific peaks in the ChIP-seq data and specific binding at a subset of asynchronous cell FoxA1 target sites.

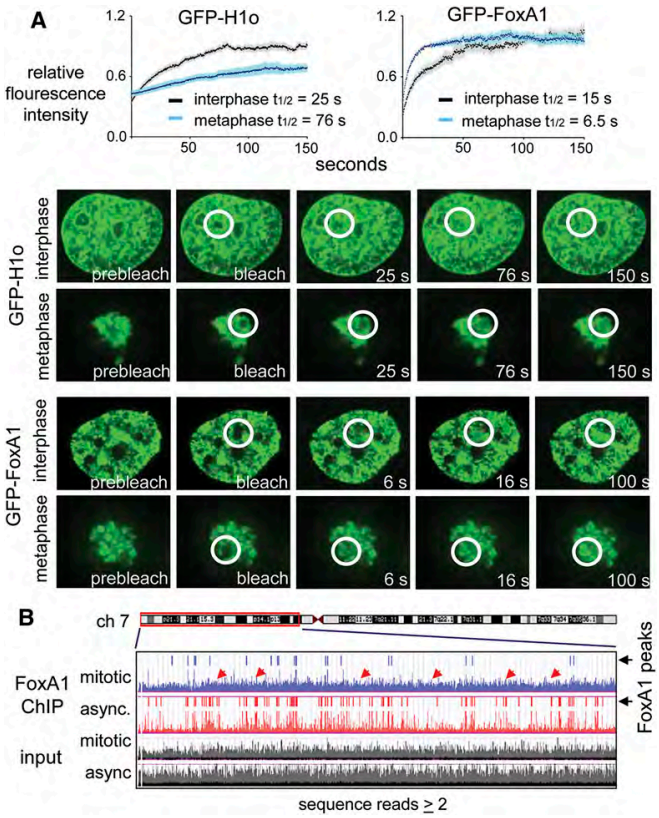


Figure 4. Increased mobility and nonspecific binding of FoxA1 in mitotic cells. (*A*) Relative fluorescence intensity (RFI) analysis showing that while GFP-H1o moves threefold more slowly in mitotic chromatin compared with interphase nuclei, GFP-FoxA1 moves 2.5-fold more quickly. Error bars denote standard error of the mean (SEM). Primary FRAP data for RFI analysis. White circles indicate the bleached area. (*B*) Unique FoxA1 ChIP-seq signals from two to 10 reads per million per 25-bp interval mapped to the left arm of human ch. 7 depicting higher nonspecific, background binding in mitotic cells (red arrows) and many more peaks in asynchronous cells. Input DNA is shown at two to 20 reads per 25-bp interval.

Specific and nonspecific mitotic binding promotes target gene reactivation post-mitosis

Previous knockdown studies of Brd4 (Dey et al. 2009), MLL (Blobel et al. 2009), and GATA1 (Kadauke et al. 2012) found that these mitotic bookmarking factors are required for the timely reactivation of genes to which they bind in mitosis and are not required for the initial reactivation of genes to which they are bound only in interphase. To assess FoxA1 in this context, we compared the target gene reactivation after mitotic release in cells transfected for 2 d with a *FOXA1* knockdown siRNA versus a control siRNA (Supplemental Fig. 5c,d). The relatively short time period is sufficient for determining effects on nascent transcript induction (see below) but, in our hands, not long enough for marked effects on steady-state levels of mRNA. By immunofluorescence, RNA polymerase II was absent from metaphase chromosomes in blocked HUH7 cells and then rebound chromatin 80 min after release in some cells (Supplemental Fig. 5e). Incorporation of ethynyl uridine (EU) (Jao and Salic 2008) into nascent RNA was undetectable in arrested cells; sparse staining appeared in some cells by 80 min after release and was more uniform across the nucleus by 90–100 min post-release in late telophase (Supplemental Fig. 5e), similar to other cell types (Prasanth et al. 2003). Based on these findings, we collected triplicate RNA time points before and after transcriptional reactivation and analyzed genes by RT-qPCR with primer sets that span intron–exon junctions to exclusively detect nascent mRNA. Many genes exhibited a continual increase in nascent transcripts for hours after mitotic release, while others exhibited an initial burst of expression (Fig. 6A; Supplemental Fig. 6).

Regardless of the initial re-expression pattern, most of the genes that were bound by FoxA1 in mitosis exhibited a statistically significant dependence on FoxA1 for their initial activation in late telophase (Fig. 6A; Supplemental Fig. 6, top rows), demonstrating that FoxA1 functions as a bookmarking transcription factor. None of the genes that were not bound by FoxA1 in mitosis or interphase were dependent on FoxA1 for early reactivation, demonstrating selectivity for FoxA1 target genes (Fig. 6A; Supplemental Fig. 6, bottom rows). Interestingly, many of the genes that were bound by FoxA1 in interphase but not in mitosis were also dependent on FoxA1 for their initial activation post-mitosis, although with a wider variation in response (Fig. 6A; Supplemental Fig. 6, middle rows). When we grouped all genes assayed in each category at the earliest time point and assessed the average fold induction in the presence of *FoxA1* siRNA over that with the control siRNA, both the mitotic-bound

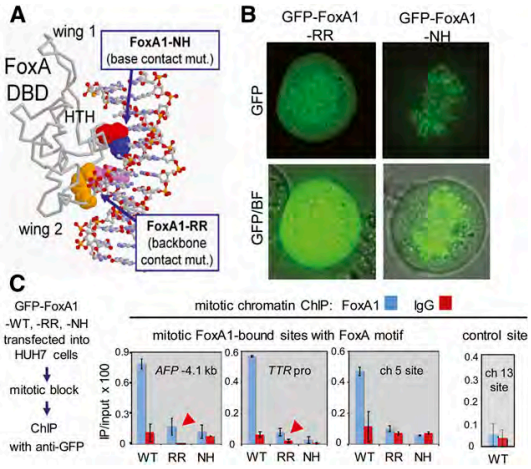


Figure 5. FoxA1 mutants that perturb specific or nonspecific DNA binding reveal significant nonspecific binding to mitotic chromatin. (*A*) Crystal structure view of the FoxA DBD depicting residues mutated to perturb specific (NH) and nonspecific (RR) DNA binding (Sekiya et al. 2009). (*B*) The FoxA1 mutant with impaired nonspecific DNA binding is partially released from mitotic chromosomes, while the FoxA1 mutant with impaired specific DNA binding is mostly retained. (*C*) ChIP-qPCR assays on transfected HUH7 cells in mitosis showing that the FoxA1-RR nonspecific binding mutant can still recognize the *Afp* −4.1-kb and *Ttr* promoter target sites that contain FoxA-binding motifs (shown in the *top* row, red arrowheads), whereas the FoxA1-NH specific binding mutant cannot. Thus, even when loss of nonspecific binding results in most of the FoxA1 being lost from mitotic chromosomes (*B*), FoxA1 can bind specifically to target sites in mitotic cell chromatin (*C*).

and asynchronous-only-bound FoxA1 target genes exhibited a significant difference from the genes not bound by FoxA (Fig. 6B). Similar results were obtained with a less extensive study using a different siRNA to knock down *FoxA1* (Supplemental Fig. 7). We conclude that FoxA1's increased mobility in mitotic chromatin (Fig. 4A) yet high nonspecific chromatin-binding capacity (Figs. 5, 6) keep the factor on chromatin and facilitate rapid reactivation during mitotic exit. The most highly expressed FoxA1 targets retain FoxA1 binding in mitosis, whereas most other FoxA1 targets appear to be dependent on nonspecific binding to mitotic chromatin. Importantly, FoxA1 does not indirectly enhance the reactivation of genes to which it is not bound in interphase. We conclude that the pioneer factor FoxA1 has mitotic chromatin-binding features that distinguish its activity from other bookmarking factors that have been characterized.

Discussion

In this study, we demonstrated that FoxA1, a pioneer factor involved in early steps of liver development, also bookmarks targets genes during mitosis. FoxA1 employs two different modes of mitotic binding that contributes to gene reactivation during exit from mitosis. The specific binding mode is responsible for mitotic marking of genes that are highly expressed in interphase, including important liver genes, and the nonspecific binding mode keeps FoxA1 in the vicinity of other target genes by random site retention on mitotic chromosomes. The nonspecific chromatin binding occurs via the high intrinsic affinity of FoxA1 for nucleosomal DNA (Cirillo and Zaret 1999; Sekiya et al. 2009) and the factor's increased chromatin mobility in mitosis (Fig. 4). By retaining FoxA1 on the DNA in mitosis, albeit nonspecifically, the factor has initial access to its specific target sites that would precede what would be observed for factors that do not bind mitotic chromatin and thus could facilitate early gene reactivation during mitotic exit (Fig. 7).

What determines whether a transcription factor will remain bound to a particular site in mitosis? The answer to this question has eluded the field, and there may be different mechanisms employed by different factors. As seen by other factors (Kadauke et al. 2012), the binding sequence motif for FoxA1 is essentially the same at mitotic-bound sites compared with sites bound only in interphase (Fig. 2D). We failed to discover histone modifications at FoxA1 target sites in asynchronous cells that would predict mitotic binding. However, this negative result could be due to a limitation of the available chromatin modifications currently in the online databases or to a lack of correspondence between the cell lines that have been assayed and HUH7 cells in which we performed our experiments. Importantly, we were able to discern that predicted INOSs (Kaplan et al. 2009) are sufficient to distinguish the subset of sites that remain bound by FoxA1 in mitosis; i.e., the sites that retain FoxA1 binding in mitosis had significantly higher INOSs. While nucleosome mapping in vivo will be required to assess the validity of these predictions, they are striking in light of FoxA1 having a markedly lower off rate for its target sites on nucleosomes compared with free DNA (Cirillo and Zaret 1999). Furthermore, sequential ChIP studies showed that FoxA1 binds a nucleosomal target site in vivo (Chaya et al. 2001). Because many, but not all, other transcription factors lack the high intrinsic affinity for nucleosomal DNA, it seems unlikely that the INOS will be a general predictor of mitotic chromosome binding. However, for this class of protein, it could be useful.

How is nonspecific DNA binding by FoxA1 more prominent in mitosis than in interphase? Our FRAP experiments revealed that despite linker histone moving more slowly in mitotic chromatin, in agreement with prior studies (Chen et al. 2005), the mobility of FoxA1 in chromatin is increased compared with interphase. This observation, coupled with the high intrinsic nucleosome and chromatin-binding capacity of FoxA1, could be sufficient to explain a decrement in specific target residence time with a concomitant increase in nonspecific chromatin binding. As for what causes an increased mobility of FoxA1 in mitotic chromatin, we speculate that a mitosis-specific modification of the protein could play a role, the overall compaction of chromatin could help exclude many specific sites, and/or an altered modification of the chromatin itself could contribute.

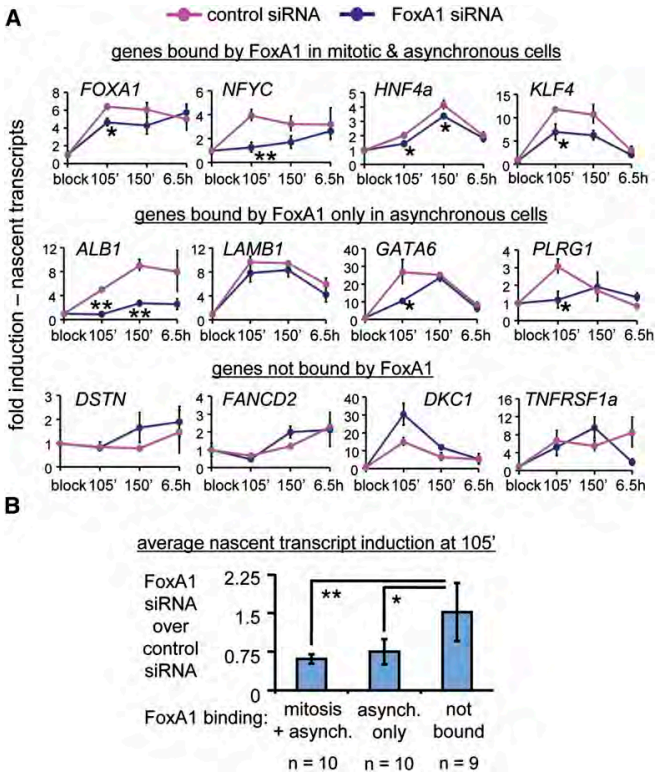


Figure 6. FoxA1 is necessary for timely reactivation of target genes as cells exit mitosis. (*A*) RT-qPCR analysis of primary transcripts in HUH7 cells treated with siRNAs for FoxA1 (si6689) or a control siRNA blocked in mitosis and released for various time points. Data are normalized to *Gapdh* and to the "block" time point prior to mitotic release. The first two genes in each row exhibited a net increase in synthesis over time, while the others exhibited an initial burst of synthesis. The *top* row depicts genes that are bound by FoxA1 in mitosis; these are dependent on FoxA1 for late telophase reactivation. The *middle* row depicts genes that are bound by FoxA1 only in asynchronous cells, many of which depend on FoxA1 for its initial activation. The *bottom* row depicts genes that are not bound by FoxA1 and are independent of FoxA1 for early reactivation. See Supplemental Figure 5 for more genes of each type. Error bars denote SEM; asterisks indicate significance by a one-tailed Student's *t*-test: (*) $P < 0.05$; (**) $P < 0.01$. (*B*) The average nascent transcript induction at 105 min post-mitotic block release is shown as a ratio of that in the presence of the FoxA1 siRNA over that for the control siRNA for genes in the three categories in *A* and Supplemental Figure 5. Error bars denote SEM, and asterisks indicate significance by a one-tailed Student's *t*-test: (*) $P < 0.05$; (**) $P < 0.01$. A separate Mann-Whitney test, not shown, revealed that the FoxA siRNA selectively perturbed mitotic FoxA1-bound versus unbound to $P < 0.00013$ and asynchronous-only bound versus unbound to $P < 0.022$. The data show that FoxA1 facilitates the early reactivation of genes bound in mitosis as well as genes bound solely in interphase cells and does not indirectly enhance the reactivation of genes to which it does not bind.

By assaying in parallel diverse types of factors critical for a single cell lineage, we discovered a diversity of mitotic chromatin-binding types, including total occupancy, specific and nonspecific binding, partial or total chromosome exclusion, and differential mitotic stability. We note that the earliest transcription factors in liver development, including FoxA1 and HNF1β (Lokmane et al. 2008), are necessary in the endoderm for hepatic induction and exhibit high-level binding to mitotic chromosomes (Verdeguer et al. 2010). In contrast, the other hepatic transcription factors tested exhibit successively less mitotic binding capacity. This raises the possibility that the diverse modes of transcription factor binding in mitosis may mimic binding hierarchies in development. The mitotic binding hierarchy could be a way to ensure the initial exclusion of factors that would otherwise promote an alternate cell fate when a given cell type exits mitosis.

Materials and methods

Mitotic arrest of HUH7 cells

HUH7 cells were plated in fresh DMEM High Glucose with L-glutamine (Invitrogen, 11965), 10% FBS (HyClone), and 1% penicillin–streptomycin (Invitrogen) and grown overnight to 75% confluence. Cells were blocked in S phase by addition of fresh medium containing 2 mM thymidine (Sigma, T1895). After 24 h, cells were washed three times with PBS and released into medium containing 0.06 μg/mL fresh nocodazole (Sigma, M1404). After 18 h, ~94% of the cells were blocked in metaphase (Supplemental Fig. 3a) and used for imaging or ChIP. Drug washout experiments showed that the cells remained viable and proliferative. For mitotic block release studies, arrested cells were either (1) harvested by gentle shaking and replated into fresh medium or (2) washed three times with PBS, and fresh medium was added to the cells. Cells were allowed to proceed for the periods indicated.

Fluorescence imaging of HUH7 cells

HUH7 cells were plated in glass-bottomed microwell dishes (Mat Tek Corporation) and transfected on the next day with

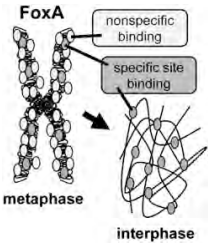


Figure 7. Specific and nonspecific modes of FoxA1 binding to the mitotic genome allow rapid reactivation of FoxA target genes during mitotic exit.

GFP-tagged fusion proteins (Sekiya et al. 2009) with FuGENE 6 (Roche). After 48 h of mitotic arrest, asynchronous and nocodazole-blocked cells were photographed at 100× using a Deltavision Core Deconvolution microscope (Applied Precision) from an Olympus IX70 microscope equipped with a Photometrics CoolSNAP HQ 12-bit monochrome cooled CCD camera. We used a 100× 1.4 NA oil immersion PlanSApo lens (Olympus, #UPLSAPO 100XO) objective with epi-illumination provided by a 300 W xenon arc lamp. The resulting images were deconvoluted using the constrained iterative algorithm with softWoRx (Applied Precision) acquisition software.

For conventional fluorescence microscopy, cells were fixed with 3.7% formaldehyde (Fisher Scientific) for 20 min, washed twice with PBS, permeabilized for 10 min in methanol (Fisher Scientific) at −20°C, and then rinsed three times with PBS. Cells were stained with 4,6-diamidino-2-phenylindole (DAPI) for 1 min. After three washes with PBS, the samples were imaged at 40× using a Nikon Optiphot microscope with an Optronics CCD camera. Alternatively, after the first PBS washes, cells were incubated with a monoclonal antibody against H3phospho-S10 (Abcam, ab14955) diluted 1/375 in 3% FBS, 5 mM KCl, 100 mM NaCl, 0.2% Triton X-100%, 1% BSA, and 10 mM Tris-HCl (pH 7.5) for 2 h at 4°C, then rocked 3× with PBS for 10 min at room temperature and incubated with a secondary antibody conjugated to Alexafluor 488 (Invitrogen; A21202) at 1/500 in dilution buffer for 1 h at room temperature. After rinsing twice with PBS, cells were stained with DAPI and imaged as above. For RNA polymerase II, cells were fixed for 30 min in PBS containing 4% paraformaldehyde and permeabilized for 20 min in PBS containing 0.5% Triton X-100. Blocking was in PBS, 0.1% Triton X-100 (PBS-T), and 10% goat serum (Sigma, G9023) for 1 h. 8WG16 (Abcam, ab817) was diluted 1:1000 in PBS-T and applied overnight at 4°C. Cells were washed three times with PBS for 10 min each and then treated with Alexa Fluor 488 goat anti-rabbit IgG (1:10,000 in PBS-T; Invitrogen, A11001) for 1 h at room temperature. Cells were washed three times with PBS. DAPI (Sigma, D9542) at 0.5 μg/mL PBS was applied to the cells for 2 min. Cells were washed with PBS and imaged at 60× using an Eclipse TE2000-U inverted microscope (Nikon) and CoolSNAP EZ camera (Photometrics).

For immunostaining of endogenous FoxA1, HuH7 cells were washed twice with PBS before fixation in PBS with 4% paraformaldehyde for 12 min at room temperature and then rehydrated in cold PBS, 0.5% Triton X-100 for 10 min. After rinsing twice with PBS, cells were treated with blocking solution containing PBS-T, 4% FBS, 10% glycerol, and 0.1 M glycine for 1 h at room temperature. Primary incubation was done with a goat antibody against FoxA1 (Abcam, ab5089) diluted 1/200 in blocking solution for 1 h at room temperature. Cells then were rocked three timeswith PBS-T for 5 min and incubated with Alexa Fluor 488 anti-goat (1/1000 in blocking solution; Invitrogen, A11055) for 1 h at room temperature. Finally, after rocking the cells three times with PBS-T for 10 min, cells were stained with DAPI and imaged as for H3phospho-S10.

To visualize new RNA synthesis after release from a mitotic block, HUH7 cells were synchronized by thymidine–nocodazole block and released into fresh medium. One hour before the desired time point for fixation, EU was added to 0.5 mM (Jao and Salic 2008). After this 1-h pulse, the cells were fixed and permeabilized as for RNA polymerase II. Cells were washed once with PBS, and the 30-min "click" reaction to Alexa Fluor 594-azide was initiated and later quenched as specified (Click-It RNA Alexa Fluor 594 imaging kit; Invitrogen, C10330). Cells were then washed three times with PBS while protected from light and counterstained with DAPI, washed with PBS, and imaged at 60×.

ChIP

Asynchronous and nocodazole/thymidine-blocked cells were cross-linked with 1% formaldehyde for 20 min at room temperature followed by addition of glycine to 125 mM. After two washes with PBS, cells were collected and frozen. Cell pellets were lysed, and genomic chromatin was sonicated to 200–500 bp (Diagenode Bioruptor). For ChIP-qPCR, 25 μg (for GFP-ChIP) or 35 μg (for FoxA1-ChIP) were precleared with salmon sperm DNA and protein A agarose (Millipore, #16-157). Samples were split, the first aliquot was incubated with 2 μg of rabbit IgG (Abcam, ab46540), and the second aliquot was incubated with 2 μg of rabbit polyclonal to GFP (Abcam, ab290), FoxA1 (Abcam, ab23738), or Histone H3 (Abcam, ab1791). Samples were rotated overnight at 4°C and immunoprecipitated with salmon sperm DNA–protein A agarose and low- and high-salt wash steps. Cross-linked products were reversed, RNase-treated, and DNA-purified. Samples were analyzed on an iCycler iQ multicolor real-time PCR (Bio-Rad) using SYBR Green (Bio-Rad). PCR primer sets are listed below. We used two biological replicates for the GFP-ChIP and three independent replicates for the FoxA1 and H3-ChIP.

ChIP-seq

ChIP was performed on triplicate chromatin samples of 335 μg divided for FoxA1 and IgG immunoprecipitations, and products were used to generate libraries (Illumina, IP-102-1001) of 100–200 bp (Bioanalyzer, Agilent) on an Illumina Sequencer. Additionally, one DNA input library each was made from mitotic and asynchronous chromatin. Sequence reads were aligned to the human genome (NCBI Build 36) with ELAND using default parameters. We uniquely mapped 40,002,286 and 43,176,234 reads for mitotic and asynchronous cells, respectively. Peaks were called for each lane separately with MACS (Mfold parameter 16; false discovery rate [FDR] 5% for mitotic peaks and 0.05% for asynchronous peaks) using the input lanes as background. A more lenient FDR was used to assess mitotic peaks because the majority of reads fell into regions of nonspecific binding (see Fig. 3C), making peak identification for this ChIP more difficult. The trends observed in Figure 2, C and F, are also observed when both peak sets are filtered at 5% FDR (data not shown). Peak sets from each replicate were intersected, and peaks that share ≥50% sequence were merged to define distinct binding sets. RefSeq transcripts were classified as targets if a FOXA1 peak was present in the gene body or within 20 kb upstream of the transcription start site. Data track images of peaks were constructed by pooling sequence reads from all replicate lanes, assessing tag counts at each position, normalizing the count per million aligned reads, and subtracting input. These counts were written into a wiggle file and uploaded to the University of California at Santa Cruz Genome Browser. Whole-chromosome views were created similarly, except that sequence tags were binned at 25-bp intervals prior to normalization. Motifs were discovered using MEME and TOMTOM. The presence or absence of the human FoxA1-binding sequence was assessed using the available site matrix from JASPAR (MA0148.1). ChIP-seq and gene expression array data have been uploaded to Gene Expression Omnibus.

Intrinsic nucleosome occupancy calculation

The INOS based on the Lasso algorithm (Kaplan et al. 2009; Tillo and Hughes 2009; Tillo et al. 2010) was calculated for evaluation of intrinsic nucleosome occupancy. For each 1500-bp sequence, we calculated the INOSs for each slide window of 147 bp and moved the window base by base to get the profiling of INOSs.

The control set was 100,000 sequences randomly selected from hg18.

Gene expression microarrays

Total RNA was collected from three different plates of asynchronous Huh7 cells using the RNeasy minikit (Qiagen). Expression microarrays were performed with a Human Gene 1.OST array (Affymetrix) at the University of Pennsylvania Microarray Core Facility and were evaluated using Partek.

Western blot analysis

Primary antibodies used were as follows: FoxA1 (0.001 mg/mL; Abcam, ab23738), C/EBPa (1:400; Santa Cruz Biotechnology, sc-61), and vinculin antibodies (1:5000; Sigma-Aldrich, V-9131). Secondary antibodies used were as follows: goat anti-rabbit IgG (H+L)-HRP (1:5000; Bio-Rad, 1706515) and goat anti-mouse IgG-HRP (1:5000; Santa Cruz Biotechnology, sc-2005).

FRAP assays

Cells cultured in glass-bottomed microwell dishes were transfected with GFP-tagged proteins and arrested in mitosis. The dishes were mounted onto a spinning-disk confocal microscope with a Yokogawa CSU X1 scan head and an Olympus IX 81 microscope. The cells were kept at 37°C using an Okolab Uno incubator. Acquisition and hardware were controlled by MetaMorph, version 7.7 (Molecular Devices). An Andor iXon3 897 EMCCD camera (Andor Technology) was used for image capture. Solid-state lasers for excitation (488 nm for GFP) were housed in a launch constructed by Spectral Applied Research. An Olympus 100×, 1.4 NA UPlanSApo oil immersion objective was used for all experiments.

FRAP was performed using the iLas^2 system (Roper Scientific), using a 50-mW diode-pumped crystal laser at 405 nm (Crysta-Laser, model DL405-050-O) controlled by MetaMorph. Laser power for bleaching was attenuated to 7.5%. For each experiment, four to six single-prebleach images were acquired before an area of 1.5 μm^2 was bleached. Images were collected over 150 sec (FoxA1) or 227 sec (H1) every 0.3 sec. At least nine cells were analyzed for each time.

To calculate the relative fluorescence intensity (RFI) in the bleached area at time t, we used the equation $RFI\,(t) = [I_B\,(t) - I_{BG}\,(t)/I_B\,(t_0) - I_{BG}\,(t_0)]\,[I_U\,(t_0) - I_{BG}\,(t_0)/I_U\,(t) - I_{BG}\,(t)]$. $I_B\,(t)$ is the fluorescence intensity in the bleached region at time t, $I_{BG}\,(t)$ is the fluorescence intensity in an area containing no cells (background) at time t, $I_U\,(t)$ is the fluorescence intensity in a nonbleached region in the same cell at time t, $I_B\,(t_0)$ is the fluorescence intensity of the bleached region before bleaching, $I_{BG}\,(t_0)$ is the fluorescence intensity in the region containing no cells before bleaching, and $I_U\,(t_0)$ is the fluorescence intensity in the nonbleached area before bleaching. Curve fitting was performed using Prism 5 (Graphpad Software). Error bars from individual time points represent standard error of the mean (SEM).

FoxA1 knockdown

Huh7 cells at 30% confluence were reverse-transfected with 3 nM siRNA (FoxA1 ID: si6689) (Fig. 6; Supplemental Figs. 5, 6) or s6687 (Supplemental Fig. 7) and negative control #1 Silencer Select siRNAs (Ambion) using RNAiMAX (Invitogen, 13778-075). After 58 h, mitotic arrested cells were harvested by shake-off and plated with fresh medium. Total RNA was collected at the time points indicated using the RNaEasy microkit (Qiagen). Nascent mRNA of target genes were analyzed by real-time PCR using SYBR Green (Applied Biosystems). Primer sets (see below) were designed to detect primary transcript, nonspliced mRNA by spanning intron–exon junctions and the adjacent exon. Expression levels were normalized to the levels of spliced GAPDH. The data are represented as fold induction over time 0 h: $2^{\Delta\Delta Ct} = 2^{Ctgene\_t\ -\ CtGapdh\_t}/2^{Ctgene\_t0\ -\ CtGapdh\_t0}$. Error bars indicate SEM. One asterisk denotes $P < 0.05$ and two asterisks represent $P < 0.01$, calculated using a one-tailed Student's t-test.

Acknowledgments

We thank Andrea Stout (CDB Microscopy Core) for help with deconvolution microscopy, the IDOM Functional Genomics Core (P30DK19525) for Illumina sequencing, and the University of Pennsylvania Microarray Core Facility. We thank Gerd Blobel, Stephan Kadauke, Shelley Berger, Kimberly Blahnik, Hua-Ying Fan, Dan Simola, Abdenour Soufi, and Chengran Xu for comments, and Eileen Hulme for help in preparing the manuscript. The research was supported by NIH Epigenomics grant R01DK082623 to K.S.Z.

References

Agard DA. 1984. Optical sectioning microscopy: Cellular architecture in three dimensions. *Annu Rev Biophys Bioeng* **13:** 191–219.

Blobel GA, Kadauke S, Wang E, Lau AW, Zuber J, Chou MM, Vakoc CR. 2009. A reconfigured pattern of MLL occupancy within mitotic chromatin promotes rapid transcriptional reactivation following mitotic exit. *Mol Cell* **36:** 970–983.

Bossard P, Zaret KS. 1998. GATA transcription factors as potentiators of gut endoderm differentiation. *Development* **125:** 4909–4917.

Burke LJ, Zhang R, Bartkuhn M, Tiwari VK, Tavoosidana G, Kurukuti S, Weth C, Leers J, Galjart N, Ohlsson R, et al. 2005. CTCF binding and higher order chromatin structure of the H19 locus are maintained in mitotic chromatin. *EMBO J* **24:** 3291–3300.

Caravaca JM, Cano S, Gallego I, Daban JR. 2005. Structural elements of bulk chromatin within metaphase chromosomes. *Chromosome Res* **13:** 725–743.

Chaya D, Hayamizu T, Bustin M, Zaret KS. 2001. Transcription factor FoxA (HNF3) on a nucleosome at an enhancer complex in liver chromatin. *J Biol Chem* **276:** 44385–44389.

Chen D, Dundr M, Wang C, Leung A, Lamond A, Misteli T, Huang S. 2005. Condensed mitotic chromatin is accessible to transcription factors and chromatin structural proteins. *J Cell Biol* **168:** 41–54.

Cirillo LA, Zaret KS. 1999. An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Mol Cell* **4:** 961–969.

Cirillo LA, McPherson CE, Bossard P, Stevens K, Cherian S, Shim E-Y, Clark EA, Burley SK, Zaret KS. 1998. Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *EMBO J* **17:** 244–254.

Cirillo L, Lin FR, Cuesta I, Jarnik M, Friedman D, Zaret K. 2002. Opening of compacted chromatin by early developmental transcription factors HNF3 (FOXA) and GATA-4. *Mol Cell* **9:** 279–289.

Clark KL, Halay ED, Lai E, Burley SK. 1993. Co-crystal structure of the HNF3/*fork head* DNA recognition motif resembles histone H5. *Nature* **364:** 412–420.

Dey A, Nishiyama A, Karpova T, McNally J, Ozato K. 2009. Brd4 marks select genes on mitotic chromatin and directs postmitotic transcription. *Mol Biol Cell* **20:** 4899–4909.

Egli D, Birkhoff G, Eggan K. 2008. Mediators of reprogramming: Transcription factors and transitions through mitosis. *Nat Rev Mol Cell Biol* **9:** 505–516.

Gualdi R, Bossard P, Zheng M, Hamada Y, Coleman JR, Zaret KS. 1996. Hepatic specification of the gut endoderm in vitro: Cell signaling and transcriptional control. *Genes Dev* **10:** 1670–1682.

Jao CY, Salic A. 2008. Exploring RNA transcription and turnover in vivo by using click chemistry. *Proc Natl Acad Sci* **105:** 15779–15784.

Kadauke S, Udugama MI, Pawlicki JM, Achtman JC, Jain DP, Cheng Y, Hardison RC, Blobel GA. 2012. Tissue-specific mitotic bookmarking by hematopoietic transcription factor GATA1. *Cell* **150:** 725–737.

Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J, et al. 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458:** 362–366.

Lee CS, Friedman JR, Fulmer JT, Kaestner KH. 2005. The initiation of liver development is dependent on Foxa transcription factors. *Nature* **435:** 944–947.

Lokmane L, Haumaitre C, Garcia-Villalba P, Anselme I, Schneider-Maunoury S, Cereghini S. 2008. Crucial role of vHNF1 in vertebrate hepatic specification. *Development* **135:** 2777–2786.

Martinez-Balbas MA, Dey A, Rabindran SK, Ozato K, Wu C. 1995. Displacement of sequence-specific transcription factors from mitotic chromatin. *Cell* **83:** 29–38.

Michelotti EF, Sanford S, Levens D. 1997. Marking of active genes on mitotic chromosomes. *Nature* **388:** 895–899.

Prasanth KV, Sacco-Bubulya PA, Prasanth SG, Spector DL. 2003. Sequential entry of components of the gene expression machinery into daughter nuclei. *Mol Biol Cell* **14:** 1043–1057.

Prescott DM, Bender MA. 1962. Synthesis of RNA and protein during mitosis in mammalian tissue culture cells. *Exp Cell Res* **26:** 260–268.

Ramakrishnan V, Finch JT, Graziano V, Lee PL, Sweet RM. 1993. Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature* **362:** 219–224.

Sekiya T, Muthurajan UM, Luger K, Tulin AV, Zaret KS. 2009. Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. *Genes Dev* **23:** 804–809.

Spencer CA, Kruhlak MJ, Jenkins HL, Sun X, Bazett-Jones DP. 2000. Mitotic transcription repression in vivo in the absence of nucleosomal chromatin condensation. *J Cell Biol* **150:** 13–26.

Tillo D, Hughes TR. 2009. G+C content dominates intrinsic nucleosome occupancy. *BMC Bioinformatics* **10:** 442.

Tillo D, Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Field Y, Lieb JD, Widom J, Segal E, Hughes TR. 2010. High nucleosome occupancy is encoded at human regulatory sequences. *PLoS ONE* **5:** e9129.

Verdeguer F, Le Corre S, Fischer E, Callens C, Garbay S, Doyen A, Igarashi P, Terzi F, Pontoglio M. 2010. A mitotic transcriptional switch in polycystic kidney disease. *Nat Med* **16:** 106–110.

Yan J, Xu L, Crawford G, Wang Z, Burgess SM. 2006. The forkhead transcription factor FoxI1 remains bound to condensed mitotic chromosomes and stably remodels chromatin structure. *Mol Cell Biol* **26:** 155–168.

Young DW, Hassan MQ, Pratap J, Galindo M, Zaidi SK, Lee SH, Yang X, Xie R, Javed A, Underwood JM, et al. 2007. Mitotic occupancy and lineage-specific transcriptional control of rRNA genes by Runx2. *Nature* **445:** 442–446.

Zaret KS, Carroll JS. 2011. Pioneer transcription factors: Establishing competence for gene expression. *Genes Dev* **25:** 2227–2241.

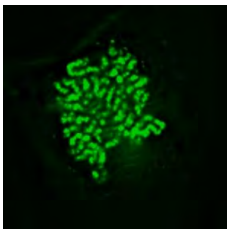
Zaret KS, Caravaca JM, Tulin A, Sekiya T. 2011. Nuclear mobility and mitotic chromosome binding similarities between pioneer transcription factor FoxA and linker histone H1. *Cold Spring Harb Symp Quant Biol* **75:** 219–226.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-seq (MACS). *Genome Biol* **9:** R137.

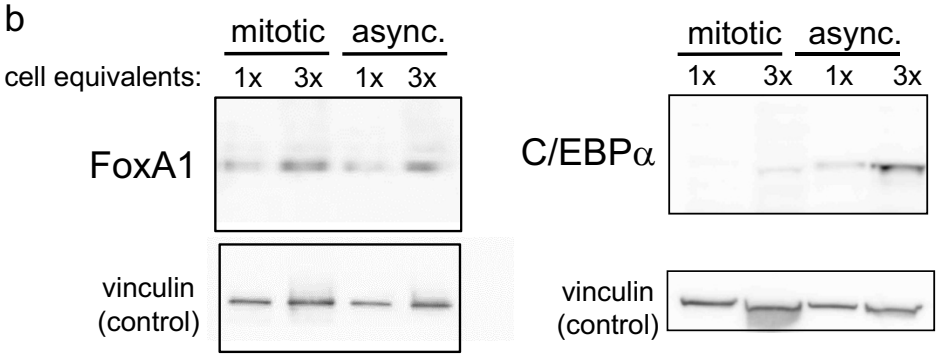
Caravaca / Zaret Supplementary Fig. 1

a

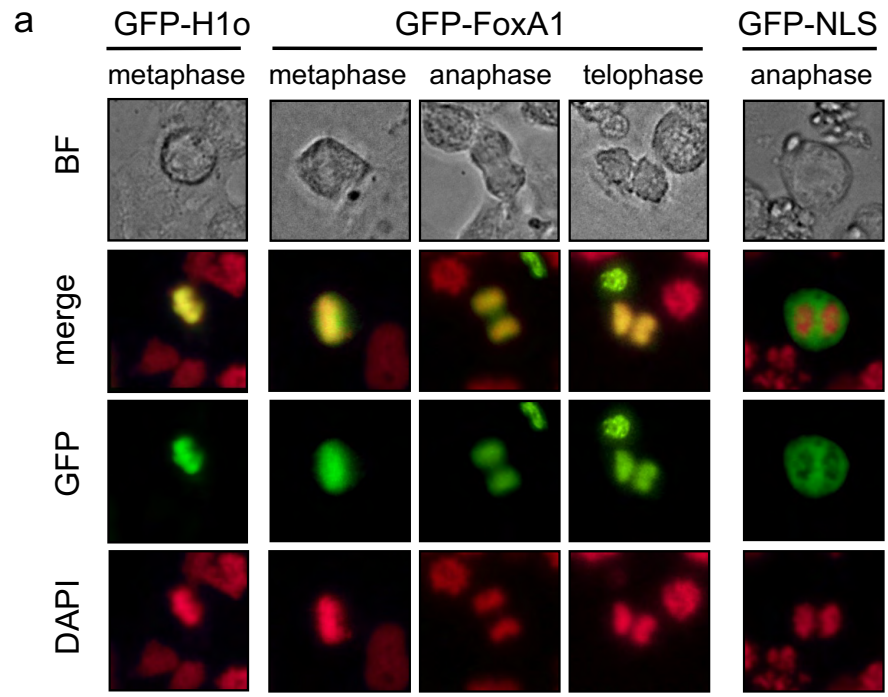
transfection with 0.05x amount of GFP-FoxA1 DNA compared to Fig. 1

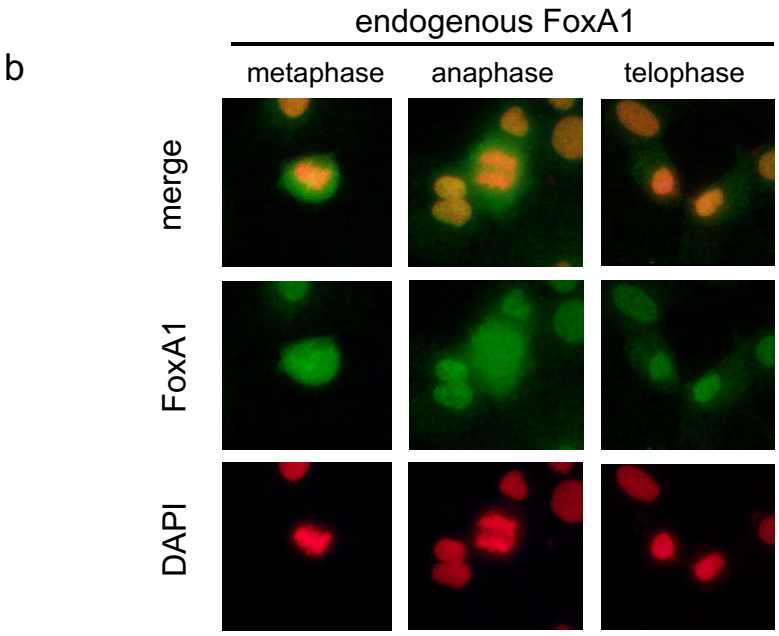


b

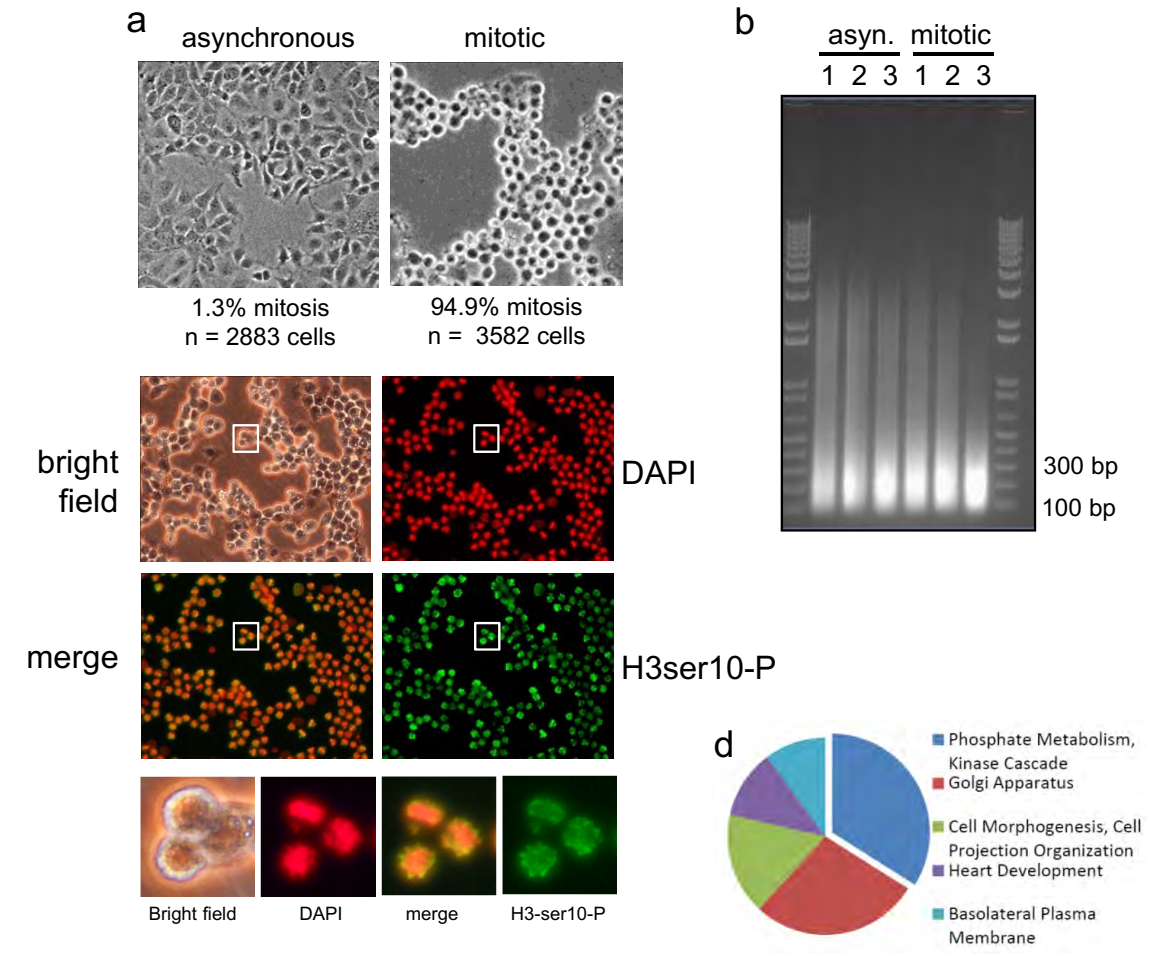


Caravaca / Zaret Supplementary Fig. 2

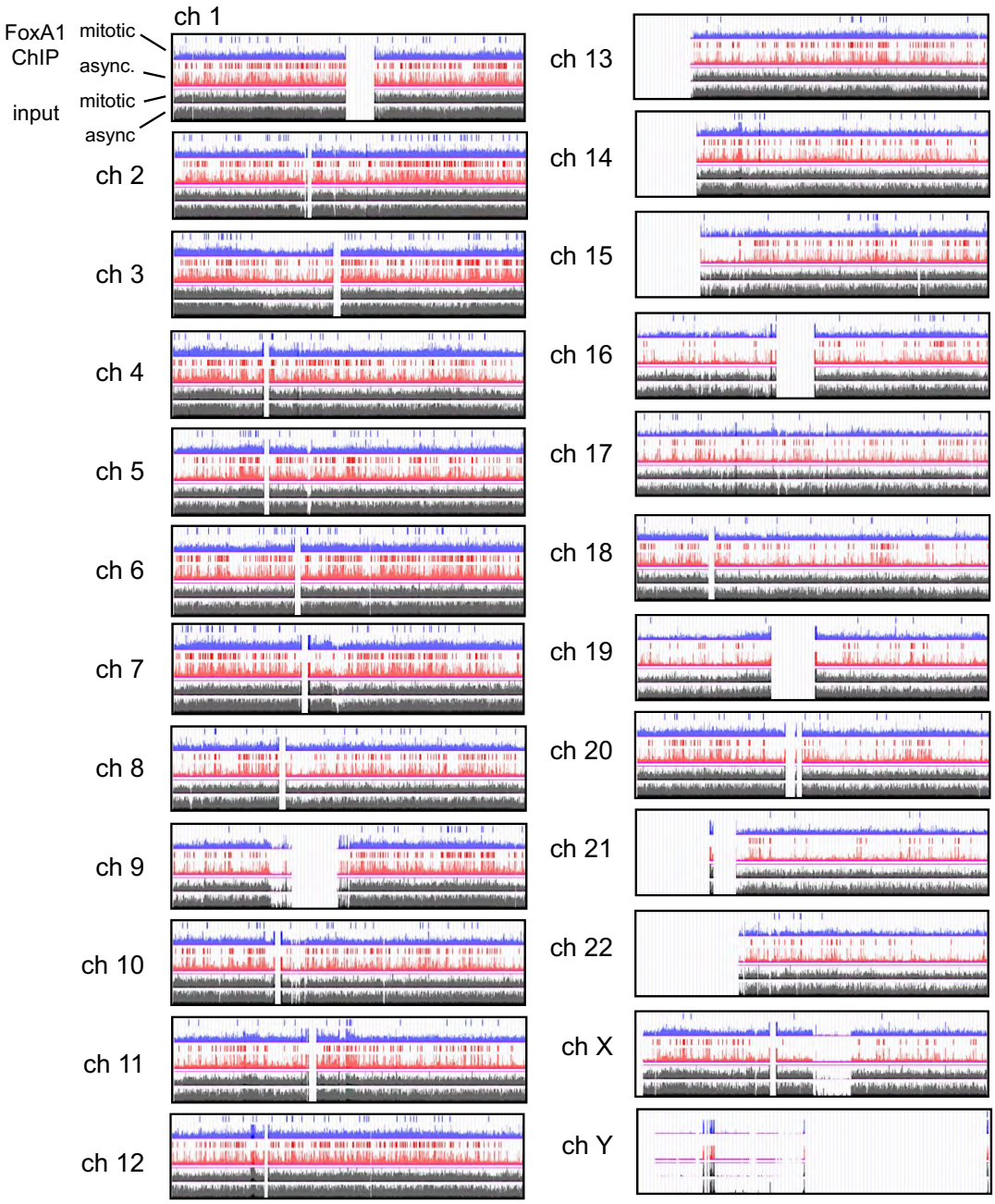




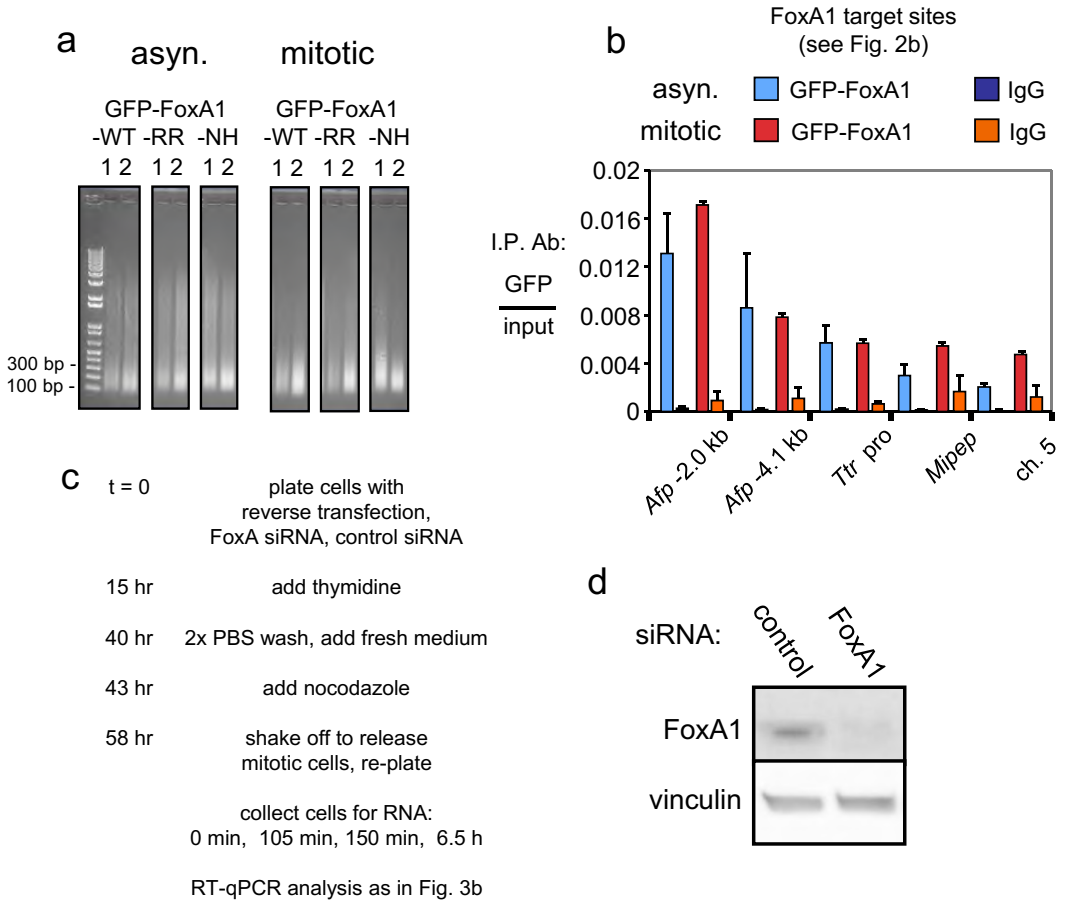
Caravaca / Zaret Supplementary Fig. 3



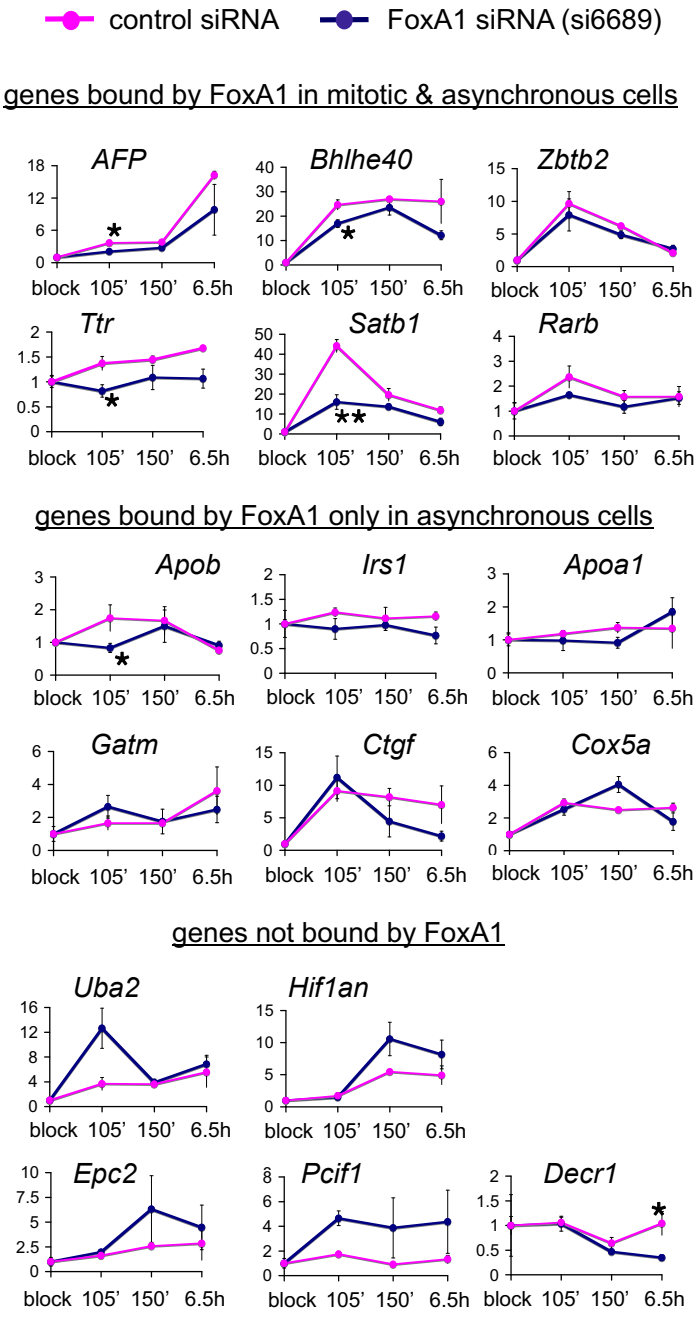
Caravaca / Zaret Supplementary Fig. 4



Caravaca / Zaret Supplementary Fig. 5



Caravaca / Zaret Supplementary Fig. 6



Caravaca / Zaret Supplementary Fig. 7