

2010

Genome-Wide Localization And Novel Deposition Pathways Of Histone Variant H3.3 In Embryonic Stem And Neuronal Precursor Cells

Aaron David Goldberg

Follow this and additional works at: http://digitalcommons.rockefeller.edu/student_theses_and_dissertations

 Part of the [Life Sciences Commons](#)

Recommended Citation

Goldberg, Aaron David, "Genome-Wide Localization And Novel Deposition Pathways Of Histone Variant H3.3 In Embryonic Stem And Neuronal Precursor Cells" (2010). *Student Theses and Dissertations*. Paper 98.

This Thesis is brought to you for free and open access by Digital Commons @ RU. It has been accepted for inclusion in Student Theses and Dissertations by an authorized administrator of Digital Commons @ RU. For more information, please contact mcsweej@mail.rockefeller.edu.



**GENOME-WIDE LOCALIZATION AND NOVEL DEPOSITION
PATHWAYS OF HISTONE VARIANT H3.3 IN EMBRYONIC
STEM AND NEURONAL PRECURSOR CELLS**

A Thesis Presented to the Faculty of
The Rockefeller University
in Partial Fulfillment of the Requirements for
the degree of Doctor of Philosophy

by
Aaron David Goldberg
June 2010

GENOME-WIDE LOCALIZATION AND NOVEL DEPOSITION PATHWAYS OF HISTONE VARIANT H3.3 IN EMBRYONIC STEM AND NEURONAL PRECURSOR CELLS

Aaron David Goldberg, Ph.D.

The Rockefeller University 2010

The eukaryotic genome is composed of chromatin, a complex polymer of genomic DNA, RNA, and closely associated proteins. Histone proteins form the core of the nucleosome, the fundamental repeating unit of chromatin. Variant histone proteins play critical roles in the epigenetic regulation of gene expression and in the development of multicellular organisms.

In this thesis, I describe the first genome-wide profiles of histone H3 variants in pluripotent mammalian embryonic stem (ES) cells, and I establish the dependence and independence of these patterns on the histone chaperone Hira. To distinguish H3 variants, I use designed zinc finger nucleases (ZFNs) to rapidly knock epitope tags into a single allele of the endogenous histone H3.3B gene in mouse ES cells.

Genome-wide analysis reveals that H3.3 is enriched in specific patterns at active and repressed genes with high CpG content promoters, including developmentally repressed bivalent H3K4me3 / H3K27me3 genes in ES cells, in addition to transcribed non-coding regions, telomeres, ribosomal DNA (rDNA), and genic and intergenic transcription factor binding sites (TFBS). Transcriptional termination sites of highly transcribed genes are marked by peaks of H3.3 and phosphorylated RNA polymerase II. Differentiation of ES cells into neural precursor cells (NPCs) leads to specific changes in H3.3 localization, demonstrating that the localization of H3.3 is dependent on cellular state. Targeted gene editing of H3.3B

to H3.2 or H3.1 using ZFNs demonstrates that these patterns are dependent on the amino acid sequence of endogenous H3.3.

Using wild-type and Hira $-/-$ ES cells, I show that the H3.3 chaperone Hira is required for H3.3 enrichment at active and repressed genes. Strikingly, Hira is not essential for deposition of H3.3 at rDNA, telomeres, and specific TFBS. Immunoaffinity purification and mass spectrometry reveal that the proteins Atrx and Daxx associate with H3.3 in a Hira-independent manner. Using Atrx^{flox} and Atrx^{null} mouse ES cells, I find that Atrx is specifically required for Hira-independent enrichment of H3.3 at telomeres and rDNA, and for repression of telomeric and ribosomal RNA. Overall, the data in this thesis demonstrate that multiple and distinct pathways are responsible for H3.3 deposition at specific genomic locations in mammalian cells.

For my family

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Dr. David Allis, for his guidance, mentorship, and unwavering support. That Dave is a brilliant and creative scientist is obvious to anyone familiar with the paradigm-shifting discoveries that have been made in our laboratory. However, even more importantly, Dave is a mentor who cares deeply about the members of his lab. My parents taught me from early on that “it’s all about character,” and Dave has that in abundance. In Dave, I found a mentor who matches me in his enthusiasm for science, and who is a truly a “*mensch*” when it comes to all of his interactions. Dave believed in me when I was a young graduate student with a crazy idea of using a new and weird technique to tag an endogenous histone gene and track its genome-wide localization in stem cells. He has supported me through all of the tough times in graduate school, not only tough times in the lab, but even in life, and we have celebrated the good times together. Despite his phenomenal success as a scientist, he remains humble, and he maintains a deep sense of perspective on what matters in life. His generosity, creativity, and his character provide a model for how to be a scientist.

I thank my faculty advisory committee, Dr. Elaine Fuchs, Dr. Shahin Rafii, and Dr. Robert Roeder for their time, their expertise, and their intellectual input on my work over the last several years. It has been an honor for me to have such great mentors.

I would like to thank Dr. Elaine Fuchs for her incredible expertise and insights regarding all aspects of stem cell biology. Her insightful, critical questions during my committee meetings were extremely valuable, and significantly helped to advance the course of my project.

In particular, I would like to thank Dr. Bob Roeder for his longstanding support and mentorship. I am very grateful for the opportunity to have spent time in Dr. Roeder's lab during my first rotation as a MD/PhD student. Bob's own work as a graduate student provided the foundation for our modern understanding of gene regulation, and his continued contributions have made him an absolute giant in our field. I spent many hours discussing the details of gene regulation with him during my rotation, and his insights over the course of my thesis project have been extremely valuable. I would also like to thank him for serving as the chair of my committee.

I am very grateful to Dr. Shahin Rafii for his service on my committee, his collaboration, and his mentorship. Over the years, I have spent enough time in Shahin's laboratory that he is almost a co-mentor to me. Shahin is a very creative and generous scientist, and he has been incredibly supportive, through thick and thin. Shahin also shares the enthusiasm that Dave and I have for science, as is clearly evidenced by some late Sunday night phone conversations about how much salt to use in a specific immunoprecipitation. I also thank him for our collaborative work on the stem cell grants that made this project possible. As my work has progressed, Shahin has become the most important collaborator for all of the *in vivo* aspects of this work, and it is because of his support that we can look forward to many future studies of H3.3 in mammalian biology. As a highly successful physician-scientist with a full-time laboratory of his own, Shahin provides a role model for me as I look forward to my future career.

I would also like to thank Dr. Eric Nestler for serving as my outside examiner. It is clear that there are many more discoveries to be made at the intersection of epigenetics and neuroscience, and I appreciate having the opportunity to present

my work to Dr. Nestler. Like Shahin, Eric is an extraordinarily successful physician-scientist, and also provides me with another role model.

I owe a special thanks to Dr. Olaf Andersen, for his fantastic leadership of our MD/PhD program, and for the excellent advice and support that he has given me over the years. I also thank Dr. Andersen for his willingness to step in and serve as a backup committee member.

There are many past and present members of the Allis laboratory to whom I owe an enormous debt of gratitude.

First, I must recognize my friends and labmates who have become the Allis lab “H3.3 team”: Dr. Laura Banaszynski, Dr. Kyung-Min Noh, Dr. Peter Lewis, and my fellow H3.3 graduate student Simon Elsaesser. I have learned so much from each of them, and I cannot thank them enough. Particularly in the last year, this work would have been absolutely impossible without their enormous efforts. It has been a privilege to share this project with them, and to spend countless hours working as part of a team. I can always count on them. I am especially grateful to the postdoctoral fellows, Laura, Min, and Peter for their guidance, and for taking time out of their own postdoctoral projects to work with an enthusiastic, but overwhelmed graduate student. I am honored that my labmates will continue our work on H3.3 even after I have returned to medical school.

I would also like to thank all of my Allis lab colleagues for their friendship and intellectual input. I would like to particularly acknowledge Dr. Beth Duncan, who helped pioneer the embryonic stem cell model in the Allis lab with me. In addition to Dr. Duncan, Lindsey Baker, Dr. Greg Wang, and Dr. David Shechter have all

read through drafts of my work at various stages. I would also like to thank the past and present members of the original “bay,” in particular Amrita Basu, Dr. Holger Dormann, Dr. Monika Lachner, Fabio Casadio, and Dr. Ronen Sadeh. I thank Dr. Christina Hughes for her gift of the 12CA5 monoclonal antibody, and Sarah Whitcomb and Dr. Tom Milne for their intellectual input and advice on chromatin immunoprecipitation (ChIP).

I would like to thank Dr. Sandra Hake for serving as my original Allis “lab coach” during my initial rotation in the lab, and for helping to inspire my work on H3 variants. I would also like to thank Dr. Joanna Wysocka and Dr. Tomek Swigut for their support and encouragement during my early years in the Allis lab. I would like to thank Dr. Emily Bernstein for her intellectual input over the years, and for the opportunity to work closely together with her and Dave on our essay for *Cell*.

I would like to thank Jamie Winshell for her extraordinary efforts on behalf of all of us in the Allis lab. In addition, I would like to thank and recognize Nichole Diaz for her recent work in culturing embryonic stem cell lines. I would also like to thank Chandra McKern and Jamie for their assistance in scheduling various events through my time in the lab, and for all of their help in setting everything up for my thesis defense.

During the course of my graduate work, in addition to my labmates, I have been blessed to work with a great number of extraordinary collaborators. This work would not have been possible without them.

Foremost, I must absolutely recognize the invaluable contributions of Dr. Deyou Zheng. Without Deyou, this thesis would be significantly thinner. For years, I

have dreamed of establishing the genome-wide localization of histone H3.3 in mammalian stem cells and differentiated cells. Since we began working together in July 2008, Deyou's partnership has enabled this dream to become a reality. We have spent countless hours talking about this project on the phone, in person, and over email. Deyou is simply brilliant. Though I recognize that the processing and analysis of genome-wide sequencing data takes extraordinary expertise, time, and effort, I have always thought of Deyou as something of a bioinformatics genie. He makes it look that easy. I owe Deyou an enormous debt of thanks for his intellectual contributions, his patience, and his generosity. We make a fantastic team, and I look forward to continuing our collaboration.

I also owe an incredible debt of thanks to Dr. Ileana Cristea. Ileana and I clicked from the moment we started to work together, and I have learned so much from her. It is because of Ileana's extraordinary expertise in protein complex purification, proteomics, and mass spectrometry that we were able to identify Atrx and Daxx as Hira-independent players in H3.3 deposition, and this advance has changed the whole direction of our project, and opened up a whole new perspective on our field. I want to thank Ileana for her generosity, her expertise, and her friendship. I would also like to thank Kelly Molloy, who worked with Ileana to perform some of the mass spectrometric analyses shown in this thesis.

I also owe a special debt of thanks to Dr. Fyodor Urnov and his colleagues at Sangamo Biosciences. Without Fyodor, Sangamo, and their extraordinary gene targeting technology, this project would have been impossible in its present form. I thank them for their collaboration and advice throughout this project. Fyodor shares my enthusiasm for science, and he has been an enthusiastic supporter of this project from day one. I would also like to thank Fyodor, Dr. Michael Holmes,

and Dr. Philip Gregory, and all of their colleagues at Sangamo for their generosity in hosting me and taking the time to talk about our work when I visited them in California.

I would like to thank Dr. Chingwen Yang and her team at the Rockefeller Gene Targeting Facility for their outstanding generosity over the years. Chingwen welcomed me into her lab and taught me the basic principles of mouse embryonic stem cell gene targeting years ago, and without her help, this project would never have gotten off the ground. Chingwen and her colleagues have been full collaborators in the generation of many of the mouse ES lines that are described in this thesis. I would especially like to recognize Xuan Li for her time and efforts.

I would like to thank Scott Dewell and his team at the Rockefeller Genomics Resource Center. Scott is brilliant, and he is a fantastic asset to the Rockefeller University. He has been an exceptional collaborator from the beginning of our genome-wide experiments. I would also like to thank Dr. Connie Zhao for her leadership of the Rockefeller Genomics Resource Center, her support, and for giving me the opportunity to present my work to the Rockefeller community.

I owe a special thanks to Dr. John Greally. John has been a fantastic resource of knowledge with regard to genomic level approaches and analyses. I am very grateful for his support, and I look forward to some very exciting collaborative work on the relationship between histone H3 variant deposition and DNA methylation.

I would like to thank Dr. Duancheng Wen in Shahin Rafii's laboratory for the countless hours that he has spent on the *in vivo* approaches to H3 variant biology. Because of his tireless efforts, we now have the opportunity to explore the localization and

function of H3.3 in mouse development. This part of the project would absolutely not have been possible without his skills, expertise, advice, and extraordinary work. I am extremely excited to see what the next few years will bring.

I would also like to thank other members of the Rafii laboratory for their collaboration on various projects, including some that will not be described in this thesis. In particular, I would like to thank Dr. Marco Seandel for his support and mentorship, as well as Dr. Daylon James, Dr. Ilaria Falciatori, and Ying Liu for their advice and collaboration. Jiyeon Kim also deserves special thanks for her assistance with ChIP and teratoma formation experiments.

I would like to thank Dr. Peter Scambler for his time, intellectual input, and for generously sharing wild-type and Hira ^{-/-} ES cells, as well as for sharing unpublished data from his laboratory. In addition, I would like to thank Dr. Ariane Chappier, a member of the Scambler laboratory, for her input and efforts.

I would like to thank Dr. Doug Higgs, Dr. Richard Gibbons, Dr. David Garrick, and Dr. Martin Law for their time, for the close reading of our collaborative manuscript, and for generously sharing their Atrxflox and Atrxnull ES cells, as well for sharing unpublished data from their laboratory. They have been amazingly generous, and all of our work on Atrx would not have been possible without their ideas, expertise, and reagents.

I would also like to thank Dr. Xingyi Guo in Dr. Deyou Zheng's group for his collaborative analysis of H3.3 enriched regions, and Erica Moehle for her graphical assistance with some of the zinc finger nuclease figures. I would also like to thank Geulah Livshits from Dr. Elaine Fuchs' laboratory for her intellectual input, and for

sharing lentiviral shRNA constructs for H3.3A and H3.3B.

Rockefeller University has some amazing core facilities run by exceptional scientists. In addition to the Gene Targeting Facility and the Genomics Resource Center mentioned above, I would also like to thank Dr. Alison North of the Bio-Imaging Resource Center, and her staff for all of her support over the years. I would also particularly like to thank Dr. Svetlana Mazel and her fantastic team at the Flow Cytometry Resource Center for the many hours that they have spent working with me on this project.

Many scientists have been extremely generous with their reagents and protocols. I would particularly like to thank Dr. Peter Adams for sharing his antibodies to Hira, Dr. Rudolf Jaenisch for sharing F1 hybrid ES cells, Dr. Genevieve Almouzni for sharing H3.1 and H3.3-FLAG-HA HeLa cells, Dr. Alex Ruthenburg for sharing WDR5 plasmid, Dr. Daniela Rhodes for sharing ATRX ADD domain protein, and Dr. Keji Zhao for sharing his native ChIP-seq protocol.

In addition, I would like to thank Dr. Titia de Lange and her group, particularly Peng Wu and Dr. Agnel Sfeir, for their time, reagents, and their generous intellectual input with regard to telomere biology.

In addition to Dr. Olaf Andersen mentioned above, I would also like to thank everyone in the MD/PhD program office for their incredible support and advice over the years, and particularly Ruth Gotian.

I would like to thank the MD/PhD program and the National Institutes of Health Medical Scientist Training Program for financial support, as well Rockefeller

University and my advisor for their financial support during the last year of my graduate work.

While this may seem unusual, I would also like to take a moment to thank everyone for their extraordinary support during an important experience in my life that took place during my graduate career. In February 2006, as my friends and family know very well, I had a potentially devastating skiing accident. I broke my back, my ribs, and damaged the tip of my spinal cord. I cannot thank my family, friends, loved ones, my advisor, my committee, labmates, the Rockefeller Graduate School, and the MD/PhD program enough for stepping up during this difficult time in my life. I would also like to thank the neurosurgeons, rehabilitation physicians, doctors, nurses, and physical therapists who gave me a new lease on life, particularly Dr. Martin Zonenshayn, Dr. Michael O'Dell, and their amazing teams. Because of their efforts, and because of the love and support I received from everyone around me, particularly my parents, my family, and my girlfriend Grace, I recovered quickly. I regained almost full neurological function within a month, and I was out of my brace and back in the lab by June of that same year. There are no words to express my gratitude.

My friends have all been extraordinary. I would particularly like to thank Dr. Jotin Marango, my best friend and college roommate, for his support and friendship over the last decade, and for making the trip cross country to attend my defense.

Finally, I would like to thank my family and loved ones. To my parents, Steven and Carmen Goldberg, and my sister Miriam, thank you for your boundless love throughout my life. Mom and Dad, despite what you have likely told all of your friends, I don't think that this work will win the Nobel Prize.

I also wish to thank my girlfriend, Dr. Grace Lau, for her love and encouragement. Grace has supported me throughout this long journey, and we have shared every experience during the course of graduate school. She has my everlasting love and thanks.

TABLE OF CONTENTS

Dedication.....	iii
Acknowledgments.....	iv
Table of Contents.....	xiv
List of Figures.....	xviii
List of Tables.....	xxiii
List of Abbreviations.....	xxiv
Chapter 1: General Introduction.....	1
One (mostly) identical genome, many cell types	1
A brief history of gene regulation.....	5
Epigenetics.....	13
Chromatin biology and epigenetic mechanisms.....	14
Chromatin – the physiological form of the eukaryotic genome.....	16
Non-coding RNA.....	18
DNA methylation.....	20
Covalent histone modifications and chromatin remodeling.....	21
Chromatin immunoprecipitation, high-throughput sequencing, and genome-wide approaches to chromatin state.....	26
Genome-wide localization of nucleosomes.....	31
Genome-wide patterns and functions of histone modifications at genes and regulatory elements.....	33
H3K4 methylation.....	34
H3K36 methylation.....	38
Histone lysine acetylation.....	39
Cell type specific differences in chromatin state.....	41
H3K4me3 and H3K27me3 bivalent domains mark specific repressed genes.....	43

Histone variants.....	47
H3.3 and the Histone H3 Variant Family.....	49
Evolution of H3 variants.....	49
H3.3 differs from H3.2 and H3.1 at a small number of amino acids.....	50
H3.3 nucleosomes are less stable than H3.1 nucleosomes.....	50
Histone H3 genetics in mammals.....	54
Genomic localization of H3.3 at genes and regulatory elements.....	55
Roles for H3.3 in large-scale nucleosome replacement.....	57
H3.3 serine 31 phosphorylation is a mitosis specific mark.....	60
Telomeres and telomeric chromatin.....	60
Developmentally regulated telomeric and pericentromeric localization of H3.3 and H3.3 serine 31 phosphorylation.....	63
Functional requirements for H3.3 in transcriptional regulation and development.....	64
H3.3 in neurobiology.....	65
Requirements for H3.3 in sexual reproduction.....	66
Histone H3 deposition machinery.....	70
Deposition of H3-H4 as a heterodimer.....	72
HIRA – histone regulatory protein A.....	75
Functional roles for HIRA and CHD1 in histone deposition, cellular differentiation, and development.....	77
Localization and function of H2A.Z in undifferentiated and differentiated cells.....	80
Concluding remarks – setting the stage for the study of H3.3 in mammalian embryonic stem cells.....	81
Chapter 2: Genome-wide patterns of histone H3.3 in mammalian cells are dependent on endogenous amino acid sequence and cellular state.....	83
Summary.....	83
Designed zinc finger nucleases enable highly efficient gene targeting of histone H3.3B.....	85

Specific enrichment of H3.3 in gene-rich euchromatic chromosome arms and telomeres is dependent on amino acid sequence.....	94
ChIP-seq provides genome-wide maps of histone H3 variants.....	102
Genome-wide patterns of H3 variants at repetitive elements.....	107
H3.3 is enriched around transcription start sites of both active and repressed genes, but is enriched in the bodies of only active protein-coding and non-coding RNA genes.....	108
Transcription end sites of highly expressed genes are marked by peaks of H3.3 and phosphorylated RNA polymerase II.....	117
The profile of H3.3 at cell-type specific genes changes with cell differentiation.....	123
Genome-wide enrichment of H3.3 at transcription factor binding sites is dependent on amino acid sequence and cellular differentiation state.....	128
Specific patterns of H3.3 at genes, regulatory elements, and repeats.....	136
Chapter Three: Hira and Atrx are required for H3.3 deposition at distinct genomic regions.....	138
Summary.....	138
Hira is required for enrichment of H3.3 at active and repressed genes.....	139
Hira ^{-/-} ES cells maintain global patterns of ES cell gene expression.....	144
Hira-independent enrichment of H3.3 at transcription factor binding sites.....	148
Hira-independent association of Atrx and Daxx with histone H3.3.....	150
Atrx is required to maintain H3.3 deposition at telomeric DNA and for repression of telomeric (TERRA) in ES cells.....	159
Atrx is required to maintain H3.3 deposition at ribosomal DNA and for repression of ribosomal RNA in ES cells.....	164
ATRX ADD domain interacts with the histone H3 tail.....	165
A new perspective on H3.3 deposition pathways.....	167

Chapter Four: General Discussion, Work in Progress, and Future Directions.....	169
Genomic localization of H3.3 is specific and requires H3.3 amino acid sequence.....	172
Constitutive incorporation of H3.3 at active and repressed genes with CpG rich promoters.....	173
Gene body deposition and transcription end site peaks of H3.3 in mammalian cells are proportional to transcriptional activity.....	174
H3.3 localization changes at specific loci following embryonic stem cell differentiation.....	177
Comparison of H3.3 and H2A.Z localization in embryonic stem cells and neural precursor cells.....	178
Hira-dependent enrichment of H3.3 at active and repressed genes.....	180
Hira-mediated deposition of H3.3 and embryonic stem cell differentiation.....	181
CHD1 in mammalian H3.3 deposition.....	184
Hira-independent H3.3 deposition at regulatory elements.....	185
ATRX and Daxx: new histone variant deposition machinery?.....	186
rDNA and telomeres: specialized, transcribed repetitive regions of heterochromatin.....	188
Daxx as a potential H3.3 binding partner.....	191
Potential implications for ATR-X syndrome pathogenesis.....	192
Localization and function of H3.3B in mammalian development – recent progress towards a H3.3B-HA-IRES-EYFP mouse line.....	197
Ongoing experiments to probe the function of H3.3 through H3.3B knockout, regulated protein expression, and H3.3A shRNA.....	202
Perspective, predictions, future questions, and challenges.....	204
Chapter Five: Materials and Methods.....	209
Appendix: Chromatin Immunoprecipitation and Detailed Protocols for ChIP-seq.....	228
References.....	238

LIST OF FIGURES

Figure 1.1: Mammalian development generates diverse cell types with identical genomes.....	2
Figure 1.2: Cell differentiation is reversible.....	4
Figure 1.3: Contemporary view of gene activation.....	9
Figure 1.4: Cellular reprogramming generates induced pluripotent stem cells from differentiated cells.....	11
Figure 1.5: Early and contemporary views of epigenetics.....	15
Figure 1.6: Chromatin structure and organization.....	17
Figure 1.7: Mechanisms for introducing variation into chromatin.....	19
Figure 1.8: Chromatin immunoprecipitation and methods of analysis.....	28
Figure 1.9: Nucleosomes are phased around active transcription start sites.....	33
Figure 1.10: Patterns of specific histone H3 methylation marks at active genes.....	35
Figure 1.11: Patterns of specific histone H3 methylation marks at repressed genes.....	36
Figure 1.12: Different genomic patterns of H3K27me3.....	41
Figure 1.13: Embryonic stem cells and differentiated cells display distinct chromatin states.....	42
Figure 1.14: Transcriptionally repressed H3K4me3 / H3K27me3 bivalent domains resolve to various chromatin states.....	44
Figure 1.15: Large-scale resolution of bivalent domains at the HoxA locus.....	45
Figure 1.16: Sequence of H3 variants: most distinguishing H3 variant amino acids are not accessible in the context of the nucleosome.....	51
Figure 1.18: Genomic organization of mouse H3 variant genes.....	54
Figure 1.19: Developmentally regulated localization of H3.3S31 phosphorylation.....	61
Figure 1.20: Telomeric chromatin has characteristics of heterochromatin...	62

Figure 1.21: Patterns of H3.3A and H3.3B expression in the adult mouse brain.....	67
Figure 1.22: Current view of H3 variant deposition pathways.....	73
Figure 2.1: Expression of H3.3A and H3.3B in embryonic stem (ES) cells and neural precursor cells (NPCs).....	85
Figure 2.2: Zinc finger nuclease targeting scheme.....	87
Figure 2.3: Zinc finger nucleases increase the efficiency of H3.3B gene targeting in mouse embryonic stem cells by over 1000 fold.....	89
Figure 2.4: Gene conversion of endogenous H3.3B to epitope-tagged H3.2 or H3.1.....	90
Figure 2.5: Efficiency of H3.3B gene targeting is dependent on donor sequence homology.....	91
Figure 2.6: Generation and validation of H3.3B targeted ES lines.....	92
Figure 2.7: Misintegrations and mutations in some ZFN targeted lines.....	95
Figure 2.8: Genomic localization of histone H3.3 is dependent on amino acid sequence.....	96
Figure 2.9: Complete chromosome spreads of H3.3-EYFP, H3.2-EYFP, H3.1-EYFP, and H3.3-HA ES cells.....	97
Figure 2.10: Foci of H3.3 serine 31 phosphorylation represent a small subset of overall H3.3 on metaphase chromosomes.....	99
Figure 2.11: H3.3 significantly colocalizes with H3K4me3, and does not colocalize with H3K9me3 pericentric heterochromatin on metaphase chromosomes.....	100
Figure 2.12: Chromosome-scale map of ChIP-seq data reveals genic enrichment of H3.3.....	103
Figure 2.13: Genomic locations of H3 variant enriched regions.....	104
Figure 2.14: Genome-wide patterns of histone modifications are not affected by heterozygous conversion of H3.3B to H3.3-HA or H3.2-HA.....	105
Figure 2.15: Genome-wide correlation of ChIP-seq datasets.....	106
Figure 2.16: Enrichment of H3 variants in different classes of repeats.....	107
Figure 2.17: H3.3 is enriched around transcription start sites of both active and inactive genes with high CpG content promoters, and into the body of active genes.....	109

Figure 2.18: Mononucleosome resolution profiles of H3 variants and H3K4me3 at TSS.....	110
Figure 2.19: Correlation of H3.3 with gene expression.....	113
Figure 2.20: Specific examples of H3.3 incorporation at the transcription start sites of active and repressed coding genes, and actively transcribed non-coding RNA.....	115
Figure 2.21: Genome-wide profiles of H3.3 at transcription start sites change with cellular differentiation and resolution of bivalent domains.....	116
Figure 2.22: H3.3 and RNA polymerase Ser5p are enriched after the transcriptional end site of highly expressed genes.....	117
Figure 2.23: Profiles of H3 variants, H3 modifications, and RNA polymerase around the highly expressed housekeeping gene beta-actin...	118
Figure 2.24: TSS and TES enrichment of H3 variants, histone modifications, and RNAPII.....	119
Figure 2.25: Analysis of H3.3 post-TES at HCP genes without repeats in their TES.....	122
Figure 2.26: H3.3 localization changes at cell-type specific genes with cell differentiation and resolution of bivalent domains.....	124
Figure 2.27: Cell-type specific enrichment of H3.3 at pluripotency genes and transcription factor binding sites.....	125
Figure 2.28: H3.3 is enriched at genes involved in cellular and metabolic processes in both ES cells and NPC cells.....	126
Figure 2.29: Expression of pluripotency and differentiation-specific genes in H3.3-HA and H3.2-HA ES cells and NPCs.....	127
Figure 2.30: Cell-type specific enrichment of H3.3 at transcription factor binding sites.....	129
Figure 2.31: Genome-wide patterns of H3.3 at transcription factor binding sites are dependent on H3 amino acid sequence and cellular state.....	130
Figure 2.32: Genome-wide changes in intergenic peaks of H3.3 and H3K4me1.....	133
Figure 2.33: Cell-type specific enrichment of H3.3 at multiple transcription factor binding	135
Figure 3.1: Expression of H3.3 and pluripotency genes in wild-type and Hira ^{-/-} ES cells.....	140

Figure 3.2: Chromosomal localization of H3.3 in wild-type and Hira ^{-/-} ES cells.....	141
Figure 3.3: H3.3 serine 31 phosphorylation is Hira-independent.....	141
Figure 3.4: Enrichment of H3.3 at active and repressed genes is Hira-dependent.....	143
Figure 3.5: Microarray analysis of gene expression in wild-type and Hira ^{-/-} ES cells.....	145
Figure 3.6: H3K4me1 and H3K36me3 are maintained at genes in the presence and absence of Hira.....	147
Figure 3.7: Hira is not essential for H3.3 enrichment at transcription factor binding sites.....	149
Figure 3.8 H3.3 enrichment at ES TFBS is partially perturbed in Hira ^{-/-} ES cells, but is not significantly affected in Atrxnull ES cells.....	151
Figure 3.9: Real-time PCR confirmation of ChIP-seq results.....	153
Figure 3.10: Atrx and Daxx association with H3.3 is Hira-independent.....	154
Figure 3.11: Probing the specificity of Atrx association with H3.3 using hypothesis-driven MS/MS analysis.....	155
Figure 3.12: H3.3 specific association with Atrx and Daxx is conserved in differentiated human cells.....	159
Figure 3.13: Expression of H3.3-EYFP in Atrxflox, Atrxnull, and Hira ^{-/-} ES cells.....	160
Figure 3.14: Similar genome-wide profiles of H3.3 in Atrflox and Atrxnull ES cells at genes and TFBS.....	161
Figure 3.15: Atrx is required for Hira-independent enrichment of H3.3 at telomeres.....	162
Figure 3.16: Atrx is required for repression of telomeric RNA.....	163
Figure 3.17: Atrx is required for Hira-independent enrichment of H3.3 at ribosomal DNA, and for repression of ribosomal RNA.....	164
Figure 3.18: ATRX ADD domain interacts with H3 1-20, and does not show a preference for H3.3 specific sequences.....	166
Figure 4.1: Hira and Atrx are required for deposition of H3.3 at distinct genomic regions.....	170

Figure 4.2: Genome-wide localization of H3.3 requires H3.3-specific amino acid sequence.....	171
Figure 4.3: H3.3 localization at cell-type specific genes changes with cell differentiation.....	175
Figure 4.4: H3.3 localization at cell-type regulatory elements changes with cell differentiation.....	176
Figure 4.5: Predicted genic locations of H3.3 / H2A.Z nucleosomes in ES cells and NPCs based on current study and reported literature.....	178
Figure 4.6: Hira ^{-/-} ES cells are defective in teratoma formation.....	180
Figure 4.7: Defects in Hira ^{-/-} ES cells upon cell differentiation may be due to impaired replication-independent histone deposition.....	182
Figure 4.8: Hira is required for deposition of H3.3 at active genes, repressed genes, and specific transcription factor binding sites.....	185
Figure 4.9: Atrx is required for H3.3 deposition at ribosomal DNA and telomeres, and for the repression of ribosomal RNA and telomeric repeat-containing RNA.....	187
Figure 4.10: Hypothetical model for Atrx/Daxx-mediated deposition of H3.3 at telomeres and ribosomal DNA.....	189
Figure 4.11: Clinical and molecular features of alpha-thalassemia and X-linked mental retardation syndrome.....	193
Figure 4.12: Speculative model for spreading of telomeric heterochromatin in ATR-X syndrome.....	194
Figure 4.13: Pluripotent embryonic stem cells give rise to chimeras and are competent for tetraploid complementation.....	196
Figure 4.14: Generation of H3.3B-HA-IRES-EYFP mice.....	198
Figure 4.15: Strategy for homozygous targeting of H3.3B.....	201
Figure 4.16: Experiments to assess the function of H3.3 in mammalian development and cellular differentiation.....	202
Figure 4.17: Speculative model: the primary non-reproductive functions of H3.3 are nucleosome replacement and chromatin repair at regions of chromatin disruption.....	205
Figure A1: Gel image of crosslinking ChIP-seq DNA pre and post-amplification for Solexa sequencing.....	236

LIST OF TABLES

Table 3.1: Global fraction of ES TFBS with differential H3.3 deposition between wild-type and Hira ^{-/-} ES cells.....	153
Table 3.2: List of H3 variant associated proteins identified by mass spectrometry.....	156
Table 5.1: Real-time PCR primers and validation of ChIP-seq.....	218
Table 5.2: Summary of ChIP-seq experiments.....	220

LIST OF ABBREVIATIONS

aa	amino acid
Ab	antibody
ADD	ATRX-DNMT3-DNMT3L
ASF1	anti-silencing factor 1
ATP	adenosine triphosphate
ATPase	adenosine triphosphatase
Atrx	alpha thalassemia and X-linked mental retardation protein (mouse)
ATRX	alpha thalassemia and X-linked mental retardation protein (human)
ATR-X	alpha thalassemia and X-linked mental retardation syndrome
Atrx ^{flox}	Atrx conditional allele flanked by loxP sites
Atrx ^{null}	Atrx null allele
bp	base pairs
BSA	bovine serum albumin
CAF-1	chromatin assembly factor 1
C6	inbred wild-type mouse embryonic stem cell background
CHD1	chromodomain helicase DNA binding protein 1
ChIP	chromatin immunoprecipitation
ChIP-chip	ChIP followed by microarray
ChIP-seq	ChIP followed by massively parallel sequencing
CpG	region of DNA where a cytosine nucleotide occurs next to guanine
DAPI	4',6-diamidino-2-phenylindole
DAXX	death domain associated protein
DAVID	Database for Annotation, Visualization and Integrated Discovery
DD	destabilizing domain
DMEM	Dulbecco's Modified Eagle's Medium

DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DNMT	DNA methyltransferase
DSB	double-strand break
EDTA	ethylenediaminetetraacetic acid
EB	embryoid body
ES	embryonic stem
Egfr	epidermal growth factor receptor
Essrb	estrogen related receptor beta
EYFP	enhanced yellow fluorescent protein
FACS	Fluorescence activated cell sorting
GFP	green fluorescent protein
GO	gene ontology
H2A	histone H2A
H2B	histone H2B
H3	histone H3
H3.1	histone H3.1
H3.1S31	histone H3.1 with a serine retained at position 31
H3.2	histone H3.2
H3.3	histone H3.3
H3.3S31p	histone H3.3 phosphorylated at serine 31
H3K27me3	histone H3 tri-methylated at lysine 27
H3K36me3	histone H3 tri-methylated at lysine 36
H3K4me1	histone H3 mono-methylated at lysine 4
H3K4me2	histone H3 di-methylated at lysine 4
H3K4me3	histone H3 tri-methylated at lysine 4
H3K9me3	histone H3 tri-methylated at lysine 9

H3K9ac	histone H3 acetylated at lysine 9
H3K14ac	histone H3 acetylated at lysine 14
H4	histone H4
H4K5ac	histone H4 acetylated at lysine 5
H4K12ac	histone H4 acetylated at lysine 12
H4K20me3	histone H4 tri-methylated at lysine 20
HA	hemagglutinin
HAT	histone acetyltransferase
HCP	high CpG content promoter
HDAC	histone deacetylase
HR	homologous recombination
ICM	inner cell mass
IF	immunofluorescence
Hira	histone regulatory protein A (mouse)
HIRA	histone regulatory protein A (human)
Hira -/-	homozygous Hira knock-out mouse
Hox	homeobox
HEK293	human embryonic kidney 293
HeLa	Henrietta Lacks
IP	immunoprecipitation
iPS	induced pluripotent stem
IRES	internal ribosomal entry site
K kb	lysine kilobases
kDa	kilo dalton
KO	knock out
LCP	low CpG content promoter
Ldha	lactate dehydrogenase A

LIF	leukemia inhibitory factor
m	meter
lincRNA	large intervening non-coding RNA
M	molar
MEF	mouse embryonic fibroblast
MNase	micrococcal nuclease
mRNA	messenger RNA
MS	mass spectrometry
MTL	multiple transcription factor binding locus
NPC	neural precursor cell
Oct4	Octamer-4 (also known as Pou5f1)
ORF	open reading frame
PcG	polycomb group
PCR	polymerase chain reaction
PEV	position effect variegation
PHD	plant homeodomain
PGC	primordial germ cell
Pou5f1	POU domain, class 5, transcription factor 1 (also known as Oct4)
RA	retinoic acid
RC	replication-coupled
RD	replication-dependent
rDNA	ribosomal DNA
RI	replication-independent
RNA	ribonucleic acid
RNAPI, II, III	RNA polymerase I, II, III
RNase	ribonuclease
RNAi	RNA interference

RNA-seq	whole transcriptome sequencing
rRNA	ribosomal RNA
RT	room temperature
RT-PCR	reverse transcriptase PCR
shRNA	short hairpin RNA
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SNF2	sucrose non-fermentable 2
SSC	spermatogonial stem cell
SWI/SNF	SWitch / sucrose non-fermentable
TERRA	telomeric repeat-containing RNA
TES	transcription end site
TF	transcription factor
TFBS	transcription factor binding site
TPE	telomere position effect
TrxG	trithorax group
TS	tissue-specific
TSS	transcription start site
UTR	untranslated region
W9.5 WT	wild-type ES cell line background wild-type
YFP	yellow fluorescent protein
ZFN	zinc finger nuclease

CHAPTER 1

GENERAL INTRODUCTION

“The major problem, I think, is chromatin...you can inherit something beyond the DNA sequence. That’s where the real excitement of genetics is now.”

---James Watson (Watson, 2003).

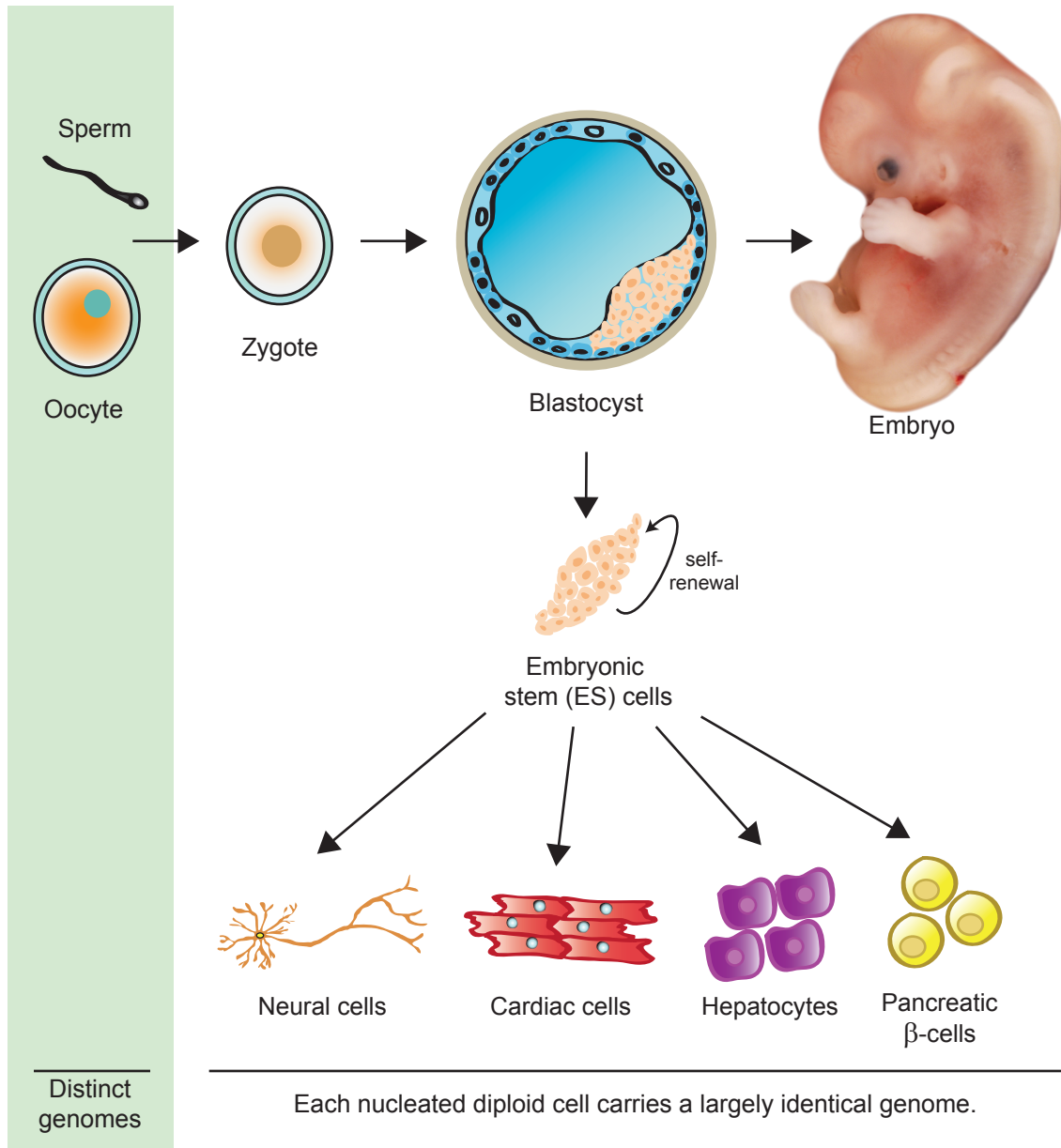
One (mostly) identical genome, many cell types

One of the singular wonders of biology is the capacity of a single cell to give rise to a multicellular organism. Every human begins life as an individual cell, the product of fusion of sperm and oocyte, formed with a unique combination of paternal and maternal genetic material. Despite the fact that a fully grown human being contains an estimated ten trillion different cells (Dobzhansky, 1970; Savage, 1977), organized into multiple organs and over four hundred spatially and functionally distinct cell types (Vickaryous and Hall, 2006), the vast majority of all of these cells contain identical genetic information (Gurdon and Melton, 2008) (**Figure 1.1**).

During the 1940s, scientists led by Avery, MacLeod, and McCarty here at the Rockefeller University demonstrated that genetic information is carried by DNA (Avery et al., 1979). However, it was not immediately clear that nearly every cell in a multicellular organism is genetically identical. One possibility was that organismal development might permanently alter the genomes of different cell types, thereby rendering them genetically distinct from other tissues of the body (Gurdon and Byrne, 2003; Gurdon and Melton, 2008). Beginning in the mid twentieth century, developmental biologists including Robert Briggs, Thomas King, and John Gurdon conducted a series of pioneering experiments involving the transplantation of differentiated frog cell nuclei from various embryonic and

Figure 1.1: Mammalian development generates diverse cell types with identical genomes.

Sperm and oocyte contain distinct haploid genomes. Following fertilization, the paternal and maternal pronuclei fuse to form the new genome of the diploid **zygote**, the unicellular embryo. From this point on, most new and old cells of the developing mammal carry an identical diploid genome throughout the lifetime of the organism, despite a vast diversity of differentiated cellular phenotypes. After repeated mitotic divisions of the zygote, the mammalian embryo becomes a morula (not shown), a ball of cells which begins at the 12-16 cell stage. About 4 days after fertilization in humans, a fluid filled cavity forms in the morula, and the embryo is now known as a **blastocyst**. Cells differentiate into an outer extraembryonic layer known as the trophoblast (blue cells), which attaches to the endometrial epithelium and later gives rise to the placenta, and an inner layer known as the inner cell mass (ICM) (pink cells), which gives rise to the **embryo** itself. If the blastocyst stage embryo is harvested prior to implantation, pluripotent **embryonic stem (ES) cells** can be derived from the ICM. ES cells have the capacity to self-renew and maintain the pluripotent state in culture for extended periods of time. In addition, ES cells can differentiate into cells of all germ lineages (ectoderm, endoderm, mesoderm), including neural cells, cardiac cells, and pancreatic β -cells depicted below. When injected back into blastocyst mouse embryos, ES cells contribute to the embryo to form chimeras, and can even give rise to the germline for passage to the next generation. See (Meissner et al., 2009; Moore et al., 2008).



Human embryo image adapted from Ed Uthman.
Cell icons adapted from Yamanaka, *Cell* 2009.

Figure 1.1

adult tissues (including lung, heart, liver, and intestine) into enucleated frog eggs (Gurdon and Byrne, 2003). Remarkably, Gurdon and colleagues found that some of the frog embryos generated from nuclei of fully differentiated somatic cells were able to develop into tadpoles and even into adult frogs (Gurdon and Byrne, 2003; Laskey and Gurdon, 1970). These experiments were later successfully conducted in mammals with the infamous cloning of Dolly the sheep from the nucleus of a cultured adult mammary gland cell by Wilmut and colleagues (Campbell et al., 1996; Wilmut et al., 1997) (**Figure 1.2**). Although this process of somatic cell nuclear

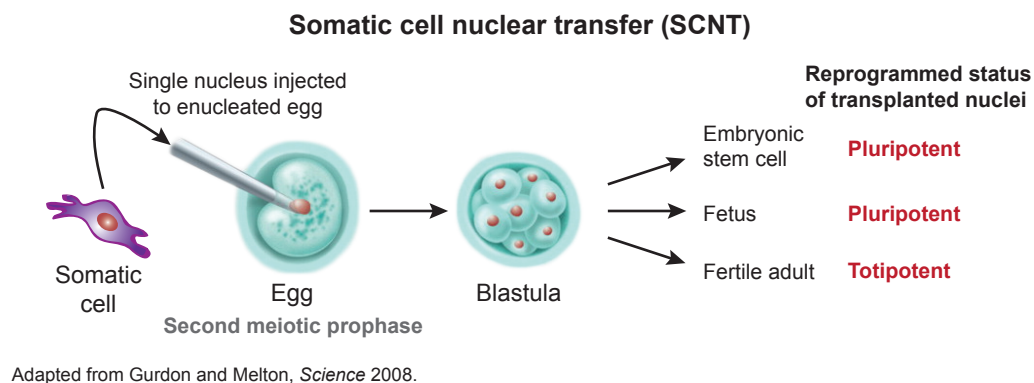


Figure 1.2: Cell differentiation is reversible

Nuclear transfer experiments by John Gurdon, Ian Wilmut and colleagues demonstrate that cell differentiation is reversible (Gurdon and Byrne, 2003; Rossant, 2009). The nucleus of a differentiated somatic cell can be transferred to an unfertilized, enucleated egg. Some of the new embryos that are generated from nuclear transfer experiments can give rise to pluripotent embryonic stem cells, to a new fetus, or a fertile adult.

transfer is not very efficient, particularly with nuclei from adult cells, the fact that even 1-2% of differentiated cell nuclei could develop into a fully formed vertebrate demonstrated that cell differentiation is reversible, and does not necessarily alter the genetic content of a cell (Gurdon and Byrne, 2003; Gurdon and Melton, 2008; Laskey and Gurdon, 1970). Gurdon and colleagues therefore concluded that the genetic blueprint for normal cellular development is preserved in all somatic cells of an organism, even in terminally differentiated cells that would normally have no

chance of passing on their genetic material.

There are of course a few exceptions to this general principle of the conservation of genetic identity during cellular differentiation. Certain immune cells contain genes that are specifically mutated to increase the body's ability to respond to the diverse array of pathogens present in our environment (Teng and Papavasiliou, 2007). Some cells such as red blood cells are so specialized that they no longer contain any stable genetic information at all, having ejected their DNA during terminal differentiation so as to be optimized for carrying oxygen from the lungs to various tissues of the body (Chasis and Mohandas, 2008). However, other than a few important exceptions, the vast majority of the trillions of cells that make up the human body are genetically identical – unique and subtly different from the cells of another human being, but identical to one another. Yet despite this genetic uniformity, the cells of our body are organized into tissues and cell types that are enormously different from one another in form and function (Ross et al., 2009; Vickaryous and Hall, 2006) (**Figure 1.1**). If their genetic content is identical, what makes one cell different from another?

A brief history of gene regulation

Some clues to the dizzying yet highly organized array of cellular diversity lie in biochemical studies. Despite generally identical DNA content, the RNA and protein content of different tissues can differ greatly (Alaiya et al., 2005; Cloonan et al., 2008; Wang et al., 2009a). While some proteins and RNA are common to many cell types, even the levels of these so-called 'housekeeping' genes and proteins can be highly variable (de Jonge et al., 2007). For example, two-dimensional electrophoresis studies of different tissue types clearly demonstrate that many proteins, even those common to multiple tissues, have dramatically different levels in distinct cell types (Alberts, 2002; Minden et al., 2009). Moreover, patterns of

RNA and protein content are so characteristic of cell type that they can be used to identify cells of uncertain tissue origin (Alaiya et al., 2005; Bloom et al., 2004; Rosenfeld et al., 2008). If different cells of an individual organism contain identical DNA, perhaps what makes one cell phenotypically different from one another therefore lies largely in the RNA and protein content of the cell.

Classic studies from Jacob and Monod in bacteria provided a model for basic genetic control mechanisms by which the biochemical composition (RNA and protein content) of a cell was determined by regulated transcription of DNA into RNA, offering a potential solution to the problem of how an identical cellular genome can give rise to such phenotypic diversity (Monod and Jacob, 1961). The expression of specific genes can be regulated in a temporal and spatial manner that allows for cellular differentiation and organismal development. In 1961, Jacob and Monod wrote:

It may be in the interpretation and analysis of differentiation that the new concepts derived from the study of microorganisms will prove of the greatest value. One point at least already seems to be quite clear: namely that biochemical differentiation (reversible or not) of cells carrying an identical genome, does not constitute a “paradox,” as it appeared to do so for many years, to both embryologists and geneticists....The control mechanisms discovered in microorganisms govern the *expression* of genetic potentialities (Monod and Jacob, 1961).

Thus, the concept of regulated cellular gene expression in the development of higher eukaryotes was born in the field of bacterial genetics. However, the mechanisms underlying the regulation of eukaryotic gene expression remained initially obscure.

In 1969, a key component of this puzzle for eukaryotes was solved with the landmark discovery of RNA polymerases by Robert Roeder (Roeder and Rutter, 1969). As a graduate student in William Rutter’s lab, Roeder purified three DNA-dependent RNA polymerases, now known as RNA polymerase (RNAP) I, II, and III, and demonstrated that RNAPI was localized to the nucleolus (the site

of ribosomal DNA transcription), while RNAPII and RNAPIII were restricted to the nucleoplasm (Roeder and Rutter, 1969, 1970). The discovery that three different enzymes catalyzed the transcription of DNA into RNA was quite surprising, particularly given that most contemporary scientists believed that animal cells contained a single enzyme (Roeder, 2003). In the following years, Roeder and colleagues demonstrated that each of the three RNAP enzymes was dedicated to the transcription of specific classes of genes (Abmayr and Workman, 2003; Roeder, 2003). Specifically, RNAPI catalyzed the synthesis of ribosomal RNA (rRNA), RNAPII catalyzed the synthesis of pre-messenger RNA (pre-mRNA), and RNAPIII catalyzed the synthesis of transfer RNA (tRNA) and 5S rRNA (Weinmann et al., 1974; Weinmann and Roeder, 1974).

With further biochemical studies, it soon became clear to Roeder and colleagues that the mechanisms underlying eukaryotic gene regulation must involve many more factors beyond the three different types of RNA polymerase. For example, each of the different types of polymerase was found to be composed of multiple subunits, most of which are unique to the specific type of polymerase (Sklar et al., 1975). In addition, transcription of mammalian gene sequences by the different RNA polymerases required multiple additional enzyme-specific recognition factors, and the gene regulation field spent many productive years defining the identity of these general transcription initiation factors, and the ordered assembly of functional transcription pre-initiation complexes (Matsui et al., 1980; Roeder, 2003; Segall et al., 1980; Sikorski and Buratowski, 2009). After the sequencing of the human genome, it was estimated that 200-300 genes code for proteins that make up the general transcriptional machinery (Lander et al., 2001; Venter et al., 2001). However, these general transcription factors are promiscuous in their capacity to transcribe numerous genes (Luse et al., 1981; Sikorski and Buratowski, 2009), and the identification of general transcriptional machinery therefore did not

solve the mystery of gene-specific and cell-type specific transcription that occurs during normal development and cellular differentiation.

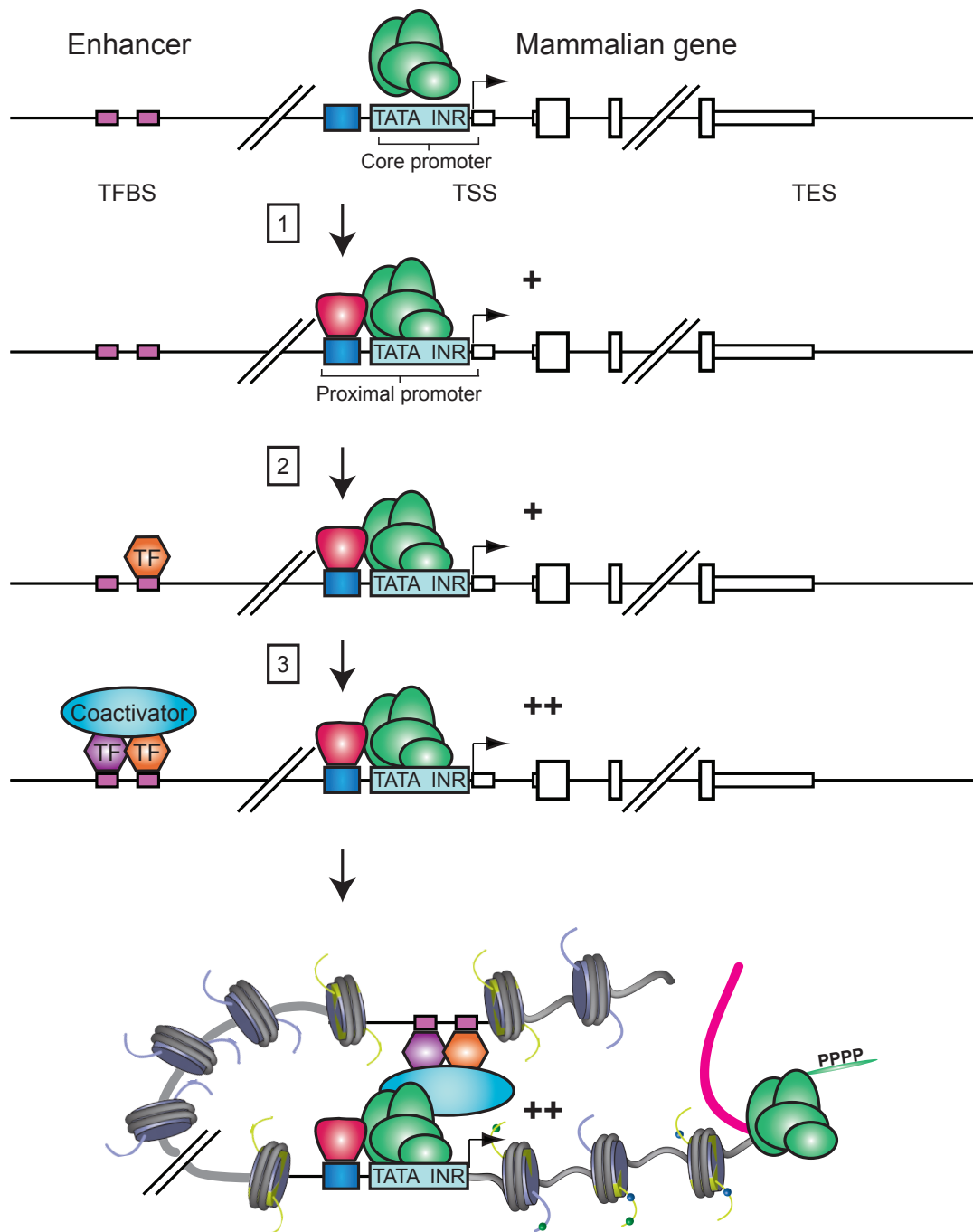
To activate or repress the transcription of particular genes, cells utilize factors that recognize specific sequences of DNA. Sequence-specific DNA binding proteins were initially identified in prokaryotes (Jacob and Monod, 1961). In 1980, the identification and purification of TFIIIA provided the first example of a sequence-specific DNA binding transcription factor in eukaryotes (Abmayr and Workman, 2003; Engelke et al., 1980; Roeder, 2003). Numerous sequence-specific DNA binding proteins have been subsequently identified, and ~1400 proteins are estimated to serve as transcription factors that bind to sequence-specific elements in humans (Farnham, 2009; Vaquerizas et al., 2009). More recently, multiple groups have identified large subunit complexes that mediate physical and functional links between sequence-specific transcription factors and the general transcriptional machinery, with certain subunits capable of interacting with specific TFs (Biddick and Young, 2005; Kornberg, 2005; Malik and Roeder, 2005; Meisterernst et al., 1991). Many of the activities of these complexes converge on the post-translational modification of histones in chromatin (Black et al., 2006; Venters and Pugh, 2009), a subject that will be addressed in a following section (**Figure 1.3**).

Tissue and lineage-specific transcription factors can play critical roles in regulating and even defining cellular identity (Hemberger et al., 2009; Niwa et al., 2000). Recently, these lineage-defining properties of transcription factors have been used to switch and reprogram cells from one somatic lineage into differentiated cells of a different lineage and even to pluripotent cells (Gurdon and Melton, 2008; Jaenisch and Young, 2008; Yamanaka, 2009). For example, the expression of the muscle specific transcription factor myoD is sufficient to switch a variety of nonmuscle cell types into muscle (Weintraub et al., 1989). Some developmentally important transcription factors often act in combination with other transcription

Figure 1.3: Contemporary view of gene activation

General transcription factors and RNA polymerase II (RNAPII) (green ovals) bind to thousands of core promoter regions through recognition of common elements such as TATA boxes and initiators (INR), but this binding alone provides very low transcriptional activity. Transcription can be increased by (1) binding of a site-specific DNA binding factor (red trapezoid) to cis elements in the proximal promoter region and stabilizing the recruitment of the transcriptional machinery. Promoter activity can be further stimulated (+) by site-specific factors (orange) binding to enhancers (2), but this step is not sufficient for high levels of transcription. Cell-type specific enhancer factors (pink) can cooperate with site-specific factors to stimulate transcription by recruiting a coactivator (3) such as a histone modifying enzyme (for example, a histone acetyltransferase, HAT) that can create a more favorable environment for transcription (++). The bottom illustration represents long-distance looping of a distal enhancer element to promote transcription in the context of chromatin. Nucleosomes are depicted, with the possible contribution of histone variants (light green) and specific histone tail modifications (small blue and pink spheres) to be discussed later. The elongating RNAPII is phosphorylated on its C-terminal domain, and the pink strand represents the growing RNA transcript. This figure has been adapted from (Farnham, 2009).

TF = transcription factor. TFBS = transcription factor binding site. TSS = transcription start site. TES = transcription end site.



Adapted from Farnham, *Nat Rev Genet* 2009.

Figure 1.3

factors (Chen et al., 2008b; Farnham, 2009), a realization that was utilized in the landmark development of cellular reprogramming (Takahashi and Yamanaka, 2006). In 2006, Yamanaka and colleagues demonstrated that the retroviral introduction of genes encoding 4 transcription factors enriched in ES cells (Oct4, Sox2, Klf4, and c-Myc) enabled differentiated fibroblasts to take on a cellular phenotype similar to pluripotent ES cells, subsequently known as induced pluripotent stem (iPS) cells (Takahashi and Yamanaka, 2006; Yamanaka, 2009) (**Figure 1.4**). Along with the previous experiments of Gurdon and colleagues described above, such cellular

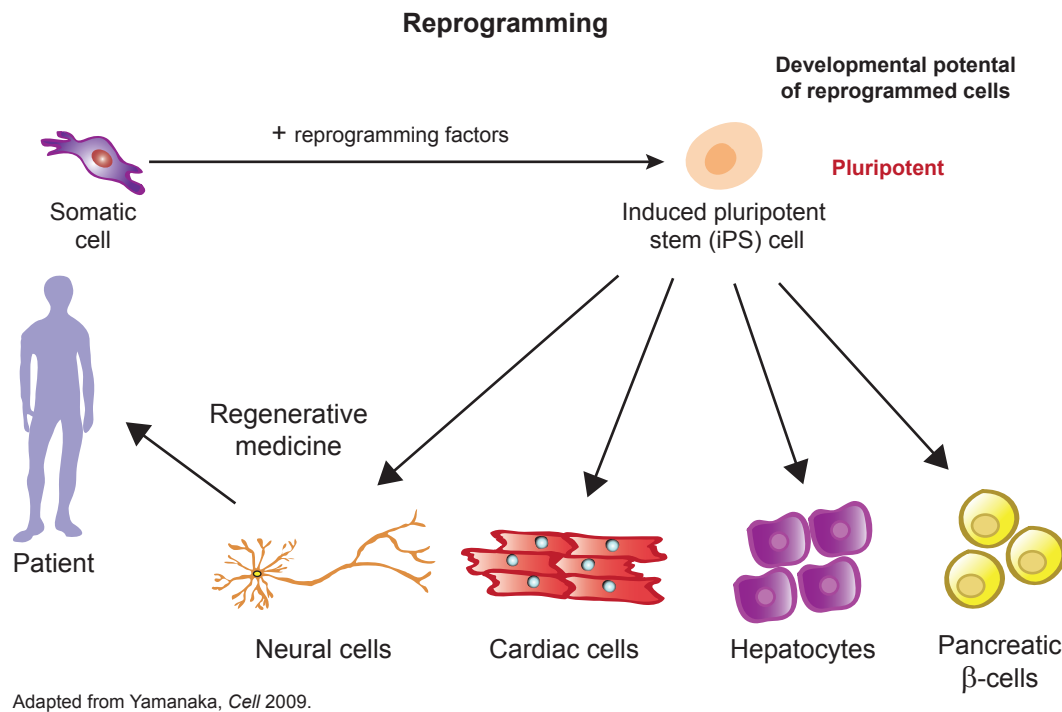


Figure 1.4: Cellular reprogramming generates induced pluripotent stem cells from differentiated cells

Cellular reprogramming further demonstrates the reversibility of cell differentiation. As originally developed by Yamanaka and colleagues in 2006, reprogramming factors (such as transcription factors Oct4, Sox2, Klf4, and c-myc) can be introduced in a variety of ways into a differentiated somatic cell. After a period of time, a number of these somatic cells will take on the morphology and cellular phenotype of pluripotent ES cells, and such reprogrammed cells are subsequently known as induced pluripotent stem (iPS) cells. Like ES cells, iPS cells have the capacity to differentiate into cells of all germ lineages. Patient-derived iPS cells can be used to produce various differentiated somatic cells with the same genotype as the patient, and these cells therefore have great potential for use in cell transplantation therapy. This figure and legend has been adapted from (Yamanaka, 2009).

reprogramming conclusively demonstrates that under specific conditions, cellular differentiation is reversible. Moreover, cellular reprogramming also paves the way for the development of patient specific iPS cells that could be used for therapeutic purposes in regenerative medicine (Rossant, 2009; Yamanaka, 2009).

Although the advent of reprogramming has revolutionized the concept of cellular identity, reprogramming remains a very inefficient process (Gurdon and Melton, 2008). In addition, cellular reprogramming is most efficient with embryonic cells, and becomes progressively less efficient as cells become more differentiated (Gurdon and Melton, 2008; Yang et al., 2007). As cells differentiate, they not only establish coordinated changes in patterns of cellular gene expression, but maintain these patterns long after they are established (Hemberger et al., 2009). Despite recent progress, the mechanisms underlying the establishment, maintenance, and propagation of these epigenetic states remain major challenges in biology. What accounts for this stability of cellular phenotype?

Nearly fifty years ago, Jacob and Monod already recognized that this relative stability of gene expression in higher eukaryotes marked a clear difference with the gene regulatory systems that they were uncovering in prokaryotes.

Most of the actual systems [involved in gene regulation in microorganisms]... are entirely reversible, in the sense that the effects of inhibitors, inducers, or repressors do not survive for any length of time after elimination of the active agent, and the cells soon return to their initial state. Differentiation, on the other hand, is stable, and persists once it has been induced. (Monod and Jacob, 1961)

Some of this phenotypic stability in the differentiation of metazoan cells can be attributed to autoregulatory loops of specific master regulatory transcription factors that reinforce cellular gene expression programs (Jaenisch and Young, 2008; Odom et al., 2006). However, transcription factors are not the only determinants of cellular identity. Even after somatic cell nuclei have undergone two rounds of transfer to the completely different cellular environment of the egg, these nuclei

will sometimes inappropriately 'remember' their original tissue of origin, expressing genes characteristic of their previous differentiated state (Ng and Gurdon, 2005, 2008). How do cell nuclei remember their tissue identity, even in the face of a completely different cellular environment, with a distinct repertoire of transcriptional regulatory proteins? What is the material actually being transferred during somatic cell nuclear transfer? Recently, genetic and biochemical evidence has converged to connect many of the mechanisms underlying cellular memory at the level of chromatin, the complex of DNA and intimately associated proteins at the center of a eukaryotic nucleus (Goldberg et al., 2007).

Epigenetics

As described above, the differentiation of cells in a multicellular organism involves changes in gene expression, but largely not in DNA content (Hemberger et al., 2009). Such a stable and heritable alteration in cellular phenotype in the absence of DNA changes can be referred to as an 'epigenetic' phenomenon (Goldberg et al., 2007).

Epigenetics has recently evolved from a collection of diverse phenomena into a defined and far-reaching field of study (Allis et al., 2007). Historically, the word 'epigenetics' was used to describe events that could not be explained by genetic principles. Conrad Waddington (1905-1975), who is given credit for coining the term, defined epigenetics as "the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being" (Waddington, 1942). Over the years, numerous biological phenomena, some considered bizarre and inexplicable, have been lumped into the category of epigenetics. These include seemingly unrelated processes, such as paramutation in maize (an interaction between two alleles in which one allele causes heritable changes in the other allele) (Cuzin et al., 2008); position effect variegation (PEV)

in the fruit fly *Drosophila* (in which the local chromatin environment of a gene determines its expression) (Girton and Johansen, 2008); and imprinting of specific paternal or maternal loci in mammals (Kacem and Feil, 2009; Weaver et al., 2009). Although mysteries abound, the field is now beginning to uncover common molecular mechanisms underlying epigenetic phenomena.

Epigenetics, in a broad sense, is a bridge between genotype and phenotype—a phenomenon that changes the final outcome of a locus or chromosome without changing the underlying DNA sequence. As described above, even though the vast majority of cells in a multicellular organism share an identical genotype, organismal development generates a diversity of cell types with disparate, yet stable profiles of gene expression and distinct cellular functions. Thus, cellular differentiation may be considered an epigenetic phenomenon, largely governed by changes in what Waddington described as the “epigenetic landscape” rather than alterations in genetic inheritance (Waddington, 1957); **Figure 1.5**). More specifically, epigenetics may be defined as the study of any potentially stable, and ideally, heritable change in gene expression or cellular phenotype that occurs without changes in Watson-Crick base-pairing of DNA (Allis et al., 2007; Goldberg et al., 2007). Despite the field’s recent progress, significant and fundamental questions remain to be answered, many of which center on the propagation of epigenetic information through cellular division and differentiation.

Chromatin biology and epigenetic mechanisms

Current epigenetic research has recently converged on the study of covalent and non-covalent modifications of DNA and histone proteins, and the mechanisms by which such modifications influence overall chromatin structure. Various overlapping mechanisms have been implicated in the epigenetic regulation of gene expression, including transcription factors (TFs), non-coding RNAs, covalent and non-covalent

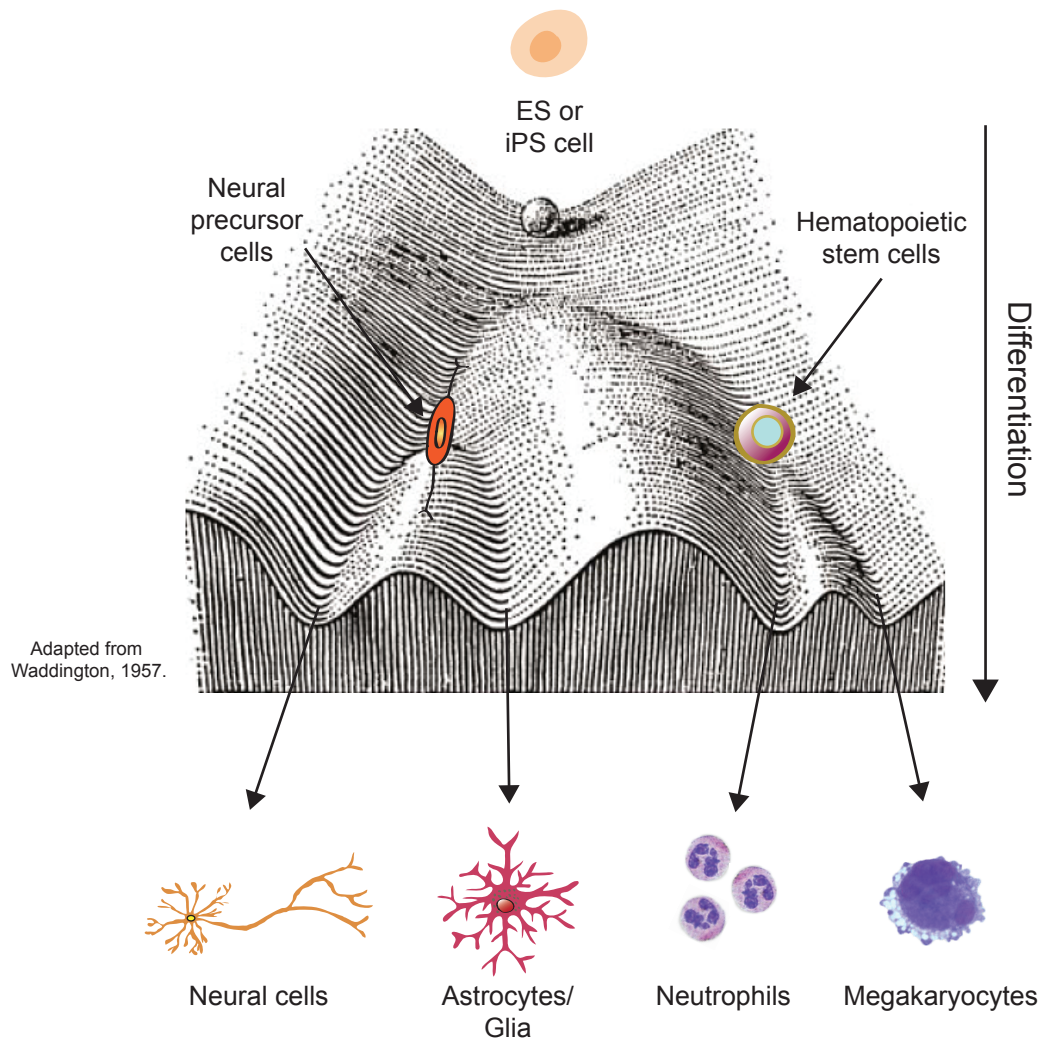


Figure 1.5: Early and contemporary views of epigenetics

In 1957, Conrad Waddington proposed the concept of an epigenetic landscape to represent the process of cellular decision-making during development. At various points in this dynamic visual metaphor, the cell (represented by a ball) can take specific permitted trajectories, leading to different outcomes or cell fates. Today, developmental biology would represent the ball at its highest point as an embryonic stem (ES) cell or induced pluripotent stem (iPS) cell. Further differentiation can give rise to precursor cells of a committed lineage, such as neural precursor cells or hematopoietic stem cells, each of which can also give rise to further precursors or terminally differentiated cells within that lineage. Figure reprinted and modified from Waddington, 1957.

modifications to histone proteins, and DNA methylation (Guttman et al., 2009; Henikoff, 2008; Klose and Bird, 2006; Rando and Chang, 2009; Takahashi and Yamanaka, 2006). Chromatin, the complex of protein and nucleic acids within the nucleus of eukaryotic cells, provides an attractive candidate for shaping the features of a cell's epigenetic landscape (Bernstein et al., 2007; Goldberg et al.,

2007).

Chromatin – the physiological form of the eukaryotic genome

The term 'chromatin' was originally coined by Walther Flemming around 1880 to refer to "that substance in the nucleus which is readily stained" (Flemming, 1882; Olins and Olins, 2003), and now generally refers to the complex of nuclear DNA, RNA, and closely associated proteins (Allis et al., 2007). Chromatin packages the eukaryotic genome into a periodic nucleoprotein structure, and serves as the physiological substrate of all DNA-dependent processes, including gene regulation (Wolffe, 1992). In each cell of higher eukaryotes, chromatin packs two meters worth of genomic DNA more than 10,000 fold in order to fit into a nuclear compartment only a few microns in diameter (Allis et al., 2007; Felsenfeld and Groudine, 2003; Jiang and Pugh, 2009) (**Figure 1.6A**). Highly basic histone proteins form the core of chromatin's repeating nucleosome core particle (Kornberg, 1974; Kornberg and Thomas, 1974; Olins and Olins, 1974; Oudet et al., 1975). In a single nucleosome, two superhelical turns of approximately 146 base pairs of DNA wrap around an octamer of histone proteins H2A, H2B, H3, and H4 (Luger et al., 1997) (**Figure 1.6B**). Critically, these repeating octamers of histone proteins and DNA do not form a homogenous, uniform chromatin structure inside the cell's nucleus. Rather, the eukaryotic nucleus is compartmentalized, and the location of a gene within the nucleus can significantly affect its expression (Lieberman-Aiden et al., 2009; Zhao et al., 2009). In general, the interphase eukaryotic nucleus can be divided into areas of highly condensed chromatin, known as heterochromatin, and regions with less compacted chromatin, known as euchromatin (Allis et al., 2007) (**Figure 1.6C**). Actively transcribed genes are generally found in euchromatin, which replicates early in S-phase, while the more densely packed heterochromatin is gene poor, enriched in repetitive DNA, and replicates late in S-phase (Carmo-Fonseca, 2002;

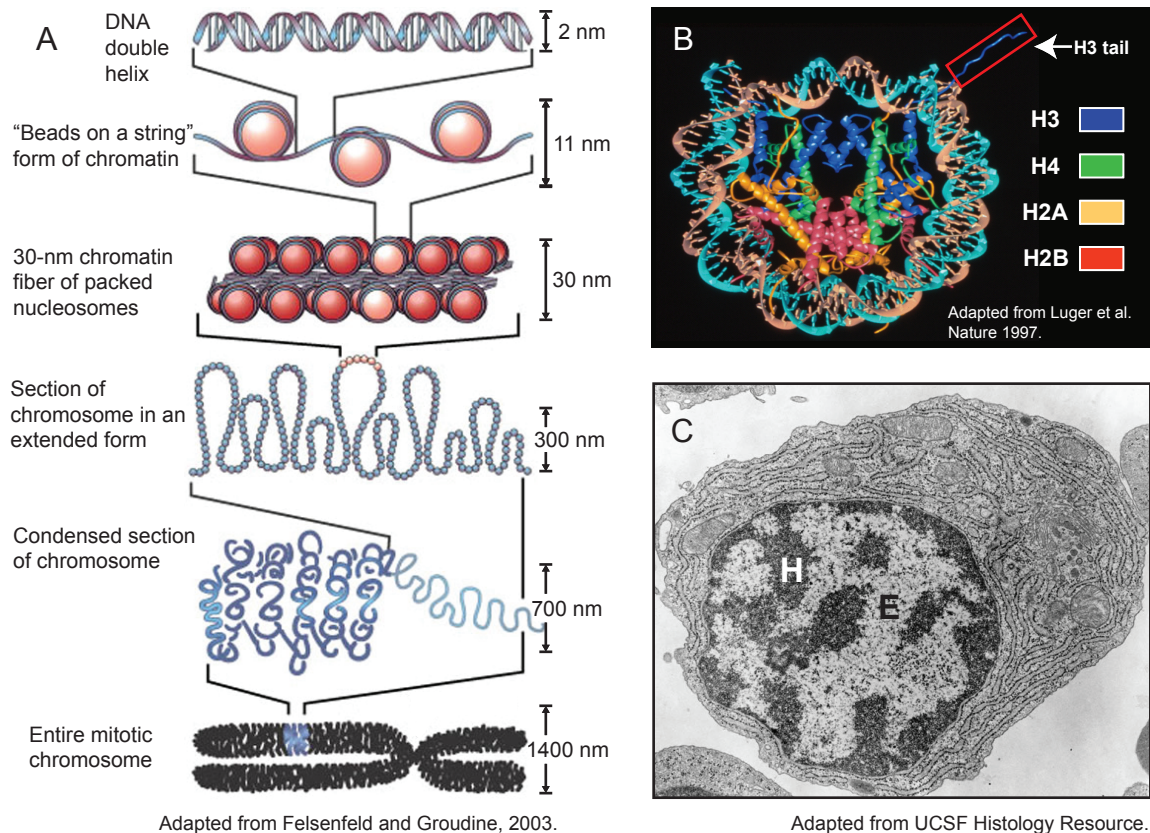


Figure 1.6: Chromatin structure and organization

A. Chromatin is the physiological form of the eukaryotic genome. The fundamental unit of chromatin organization is the nucleosome, in which 146 bp of DNA are wound around an octamer of histone proteins. Nucleosomes are connected to one another by short stretches of linker DNA. At the next level of organization, the string of nucleosomes is folded into a 30nm fiber, and these fibers are further folded into higher order structures. Figure and legend adapted from (Felsenfeld and Groudine, 2003).

B. The crystal structure of the nucleosome core particle was solved by Luger and colleagues (Luger et al., 1997). Note the DNA double helix wrapped around an octamer of histone proteins H3, H4, H2A, and H2B. One protruding histone H3 N-terminal tail is indicated in red, and such histone tails are sites of dense covalent modifications (Allis et al., 2007; Kouzarides, 2007).

C. Electron micrograph of a plasma cell from bone marrow. The nucleus is divided into electron-dense dark regions of heterochromatin (H) and lighter regions of euchromatin (E).

Martens et al., 2005).

Eukaryotic cells have evolved multiple overlapping mechanisms for introducing biologically-relevant variation into chromatin. Nucleosome positioning in eukaryotic genomes is non-random, with some aspects of nucleosome organization encoded by intrinsic nucleotide sequence (Segal et al., 2006; Segal

and Widom, 2009). In addition, the localization and stability of nucleosomes at specific genomic locations appears to be regulated by a combination of factors, including ATP-dependent chromatin remodeling enzymes, covalent modifications to histone proteins and DNA, the binding of chromatin-associated effector proteins, the localization and molecular interactions of specific non-coding RNAs, and the incorporation of primary sequence histone variants (Henikoff, 2008; Jiang and Pugh, 2009; Klose and Bird, 2006; Loyola and Almouzni, 2007; Rando and Chang, 2009; Schones et al., 2008; Wilusz et al., 2009). **(Figure 1.7)** It is likely that the macromolecular processes described below all significantly contribute to the physiologically relevant organization of most eukaryotic genomes. These entities, and possibly others yet unknown, should be considered collectively when exploring epigenetic mechanisms. Each of these mechanisms will be discussed briefly below, with a larger focus on histone variants.

Non-coding RNA

Recently, it has become evident that RNAs, particularly non-coding RNAs, play critical roles in multiple epigenetic phenomena (Bernstein and Allis, 2005; Rinn et al., 2007). Clear examples of RNA involvement range from dosage compensation mechanisms in *Drosophila* and mammals mediated by the *rox* and *XIST* RNAs, respectively, to the silencing of both genes and repetitive sequences by post-transcriptional (PTGS) and transcriptional (TGS) RNA interference (RNAi)-related pathways, respectively, in almost all eukaryotes (Bernstein and Allis, 2005). These RNAs often act in concert with various components of the cell's chromatin modifying and DNA methylation machinery to maintain active or inactive states of gene expression (Wilusz et al., 2009). Although we might not consider PTGS-inducing RNAs (e.g. microRNAs, siRNAs etc.) to be epigenetic in nature, TGS-evoking RNAs (e.g. repeat associated siRNAs, *Xist* RNA, and small RNAs in *S. pombe*) are

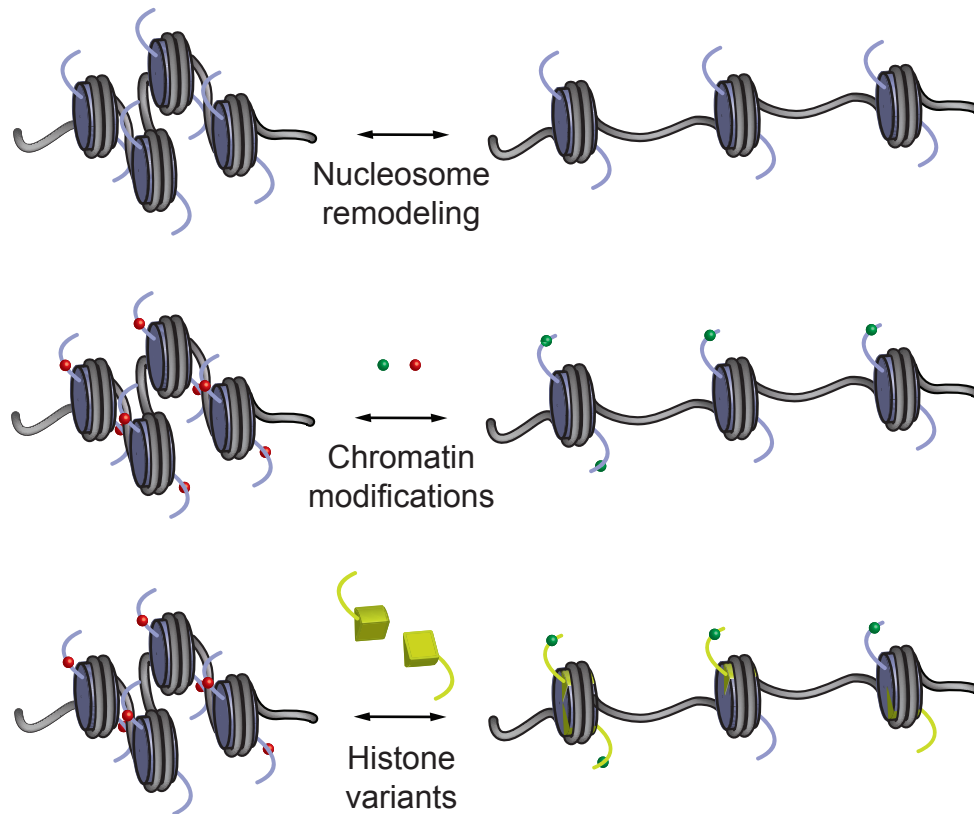


Figure 1.7: Mechanisms for introducing variation into chromatin

Eukaryotic cells utilize three primary mechanisms for introducing variation into the chromatin fiber. Nucleosomes can be physically mobilized and remodeled by ATP-dependent chromatin remodeling enzymes (top) (Clapier and Cairns, 2009). In addition, histones and DNA can be marked by various types of covalent modifications (red and blue spheres, middle). These modifications can result in distinct biological outcomes through direct physical effects, or through the binding of modification-specific effectors (Kouzarides, 2007). Finally, variant histone proteins with distinct amino acid compositions can be substituted into chromatin (bottom). These histone variants can bring with them a unique repertoire of covalent modifications, or they can serve to replace existing modifications (Loyola and Almouzni, 2007). In addition, histone variants can also have direct effects on nucleosome stability and mobility (Henikoff, 2008; Jin and Felsenfeld, 2007).

more clearly epigenetic, as they can induce long term silencing effects that can be inherited through cell division (Bernstein and Allis, 2005). In the last year, more than ~1600 large intervening non-coding RNA (lincRNA) transcripts were identified in mammalian cells, many of which were tissue-specific (Guttman et al., 2009). A number of these lincRNAs have been shown to associate with chromatin-related silencing proteins and affect gene expression (Khalil et al., 2009), and it seems likely that our appreciation for the role of RNA in epigenetic phenomena will grow

significantly in the near future.

DNA methylation

DNA methylation is perhaps the best-characterized and most clearly heritable chemical modification of chromatin. In mammals, nearly all DNA methylation occurs on cytosine residues of CpG dinucleotides (Ng and Bird, 1999). Regions of the genome that have a high density of CpGs are referred to as CpG islands, and DNA methylation of these islands correlates with transcriptional repression (Goll and Bestor, 2005). Genomic patterns of cytosine methylation in mammals, whether donated by *de novo* or maintenance DNA methyltransferases (DNMTs), play a critical role in gene regulation and chromatin organization during embryogenesis and gametogenesis (Goll and Bestor, 2005). Maintenance methyltransferases add methyl groups to hemi-methylated DNA during DNA replication, whereas *de novo* DNMTs act after replication (Groth et al., 2007b). Moreover, the formation of heterochromatin in many organisms is mediated in part by DNA methylation and its binding proteins, in combination with RNA and histone modifications characteristic of silent chromatin (Zaratiegui et al., 2007). Methylated DNA can be recognized by binding proteins such as MeCP2 (Jones et al., 1998), and mutations in MeCP2 have been shown to cause the human neurodevelopmental disorder Rett syndrome (Chahrour and Zoghbi, 2007). DNA methylation plays a role in many cellular processes including silencing of repetitive and centromeric sequences from fungi to mammals; X chromosome inactivation in female mammals; and mammalian imprinting, all of which can be stably maintained through cellular division (Yang and Kuroda, 2007). Taken together, DNA methylation provides a stable, heritable, and critical component of epigenetic regulation.

Covalent histone modifications and chromatin remodeling

For many years, chromatin was considered largely an inert scaffold and packing mechanism for nuclear DNA (Felsenfeld and Groudine, 2003). If anything, histone proteins were recognized as a simple impediment to gene transcription, as they were found to inhibit RNA synthesis (Allfrey et al., 1963), but they were not widely seen as playing any active or regulatory role. However, over the last four decades, significant genetic and biochemical evidence has accumulated to connect histone modifications with the regulation of gene expression.

Histone proteins were originally recognized to carry a high level of acetylation over 45 years ago (Phillips, 1963). In 1964, Vincent Allfrey and colleagues here at the Rockefeller University proposed a possible function of post-translational histone acetylation in the positive regulation of gene expression (Allfrey et al., 1964). Allfrey and colleagues noted that histones were acetylated even in the presence of inhibitors of protein synthesis, indicating that histone acetylation could take place following protein translation (Allfrey et al., 1964). Moreover, they found that unlike unacetylated histones, acetylated histones were relatively permissive for RNA synthesis in isolated nuclei (Allfrey et al., 1964). Intriguingly, the timing of histone acetylation was also correlated with gene transcription, with histone acetylation preceding the increase in RNA synthesis (Pogo et al., 1966). In a prescient suggestion, Allfrey and colleagues wrote:

The findings introduce the possibility that histone effects on nuclear RNA metabolism may involve more than a simple inhibition of RNA synthesis, and that more subtle mechanisms may exist which permit both inhibition and reactivation of RNA production at different loci along the chromosome (Allfrey et al., 1964).

The relationship between histone acetylation and gene transcription was further supported by later studies showing that hyperacetylated core histones preferentially associate with active areas of the genome that are more easily accessible to

nuclease digestion (Sealy and Chalkley, 1978; Vidali et al., 1978). In addition, genetic studies by Michael Grunstein and colleagues demonstrated that histone acetylation sites are required to alleviate nucleosomal gene repression in yeast (Durrin et al., 1991; Kurdistani and Grunstein, 2003). Biochemical experiments also suggested that histone tail acetylation could directly facilitate transcription factor access to nucleosomal DNA (Lee et al., 1993). However, despite this wealth of correlative evidence linking histone acetylation and transcriptional regulation, the specific molecular mechanisms responsible for histone covalent modifications remained obscure for many decades.

In the mid 1990s, two discoveries led to a major paradigm shift in the field of eukaryotic gene regulation, and forever upended the view that histones act solely as passive “scaffolding” for packaging genomic DNA. Using a novel in-gel activity assay, James Brownell, David Allis, and colleagues identified and purified the first histone acetyltransferase (HAT) enzyme from the transcriptionally active macronucleus of the ciliated protozoan *Tetrahymena thermophila* (Brownell and Allis, 1995; Brownell et al., 1996). Critically, this first HAT was highly homologous to the yeast protein Gcn5p, a transcriptional adaptor protein that had already been implicated in transcriptional activation (Brownell et al., 1996). This finding was soon followed by the identification of the first histone deacetylase (HDAC) enzyme by Stuart Schreiber and colleagues, and this first HDAC was found to be homologous to a yeast protein (Rpd3p) previously implicated in transcriptional repression (Taunton et al., 1996). Together, these landmark discoveries immediately suggested molecular mechanisms to employ “both inhibition and reactivation of RNA production at different loci along the chromosome,” as originally suggested by Allfrey, providing a critical and previously missing mechanistic link between chromatin structure and transcriptional output (Allfrey et al., 1964; Struhl, 1998).

It is now clear that histones can be marked with a wide variety of chemical

and peptide modifications in addition to acetylation, including methylation, phosphorylation, ADP-ribosylation, ubiquitination, and sumoylation (Kouzarides, 2007). These covalent modifications can also be dynamically removed by a variety of enzymes and mechanisms, including histone deacetylases, phosphatases, deubiquitinases, arginine demethyliminases, lysine demethylases, and direct histone proteolysis (Duncan et al., 2008; Keogh et al., 2006; Litt et al., 2009; Minucci and Pelicci, 2006; Nottke et al., 2009; Weake and Workman, 2008).

Genetic and biochemical evidence has recently converged to clearly connect covalent histone modifications with longstanding epigenetic phenomena. Genetic screens for suppressors of PEV (*Su(var)*) in *Drosophila*, for example, revealed over 100 genes that encode vital constituents of heterochromatin. Many of these genes are conserved from flies to humans, including heterochromatin protein 1 (HP1) and the histone H3K9 methyltransferase *Su(var)3-9* (Schotta et al., 2003). *Drosophila* genetics also provides another link between epigenetics and histone modifications, in the form of two evolutionarily conserved families of proteins that regulate homeotic genes antagonistically during development: the Polycomb Group (PcG) and the Trithorax Group (TrxG). Further molecular characterization of PcG genes reveals that they contribute to two distinct protein complexes that are responsible for 'writing' (Polycomb Repressive Complex 2) and 'reading' (PRC1) methylation of H3K27 and facilitating chromatin condensation (Whitcomb et al., 2007). Meanwhile, TrxG proteins mediate methylation of H3K4 and promote transcriptionally active chromatin (Ringrose and Paro, 2004; Ruthenburg et al., 2007a).

How might covalent histone modifications modify chromatin structure to influence patterns of gene expression (Jenuwein and Allis, 2001)? A longstanding literature suggests that charge-altering modifications such as acetylation and phosphorylation can directly alter the physical properties of the chromatin fiber,

leading to changes in higher-order structures (Hong et al., 1993; Struhl, 1998). More recently, effector-mediated functions have been well documented, whereby histone modifications recruit or stabilize the localization of specific binding partners to chromatin. In the first example of an emerging epigenetic paradigm, the catalytic enzyme *Su(var)3-9* serves as a 'writer' of histone H3K9 methylation (Rea et al., 2000), whereas HP1 serves as a 'reader' or 'effector' recognizing H3K9 methylation and directing biological processes (such as heterochromatin stabilization) to particular areas of the chromatin fiber (Lachner et al., 2001). In addition, effector-repulsive functions have been demonstrated in which histone modifications abrogate the binding of particular effectors (Fischle et al., 2005; Ooi et al., 2007).

Several investigators have noted that the combinatorial complexity of various covalent modifications could provide the physical structure of the genome with a powerful epigenetic marking system for conveying information outside of DNA sequence itself (Jenuwein and Allis, 2001; Strahl and Allis, 2000; Turner, 2000). This "histone code" of post-translational modifications could be read both by direct physical effects and by the binding of effector proteins, and could potentially contribute to the heritability of transcriptional states through cell division (Jenuwein and Allis, 2001; Strahl and Allis, 2000). Recent evidence suggests that multiple histone modifications may provide a multivalent platform for combinatorial interactions with several linked effector modules (Morinieri et al., 2009; Ruthenburg et al., 2007b). However, it remains to be seen whether histone modifications are truly 'epigenetic,' and much research is currently focused on determining if and how patterns of histone modifications are inherited through DNA replication (Groth et al., 2007b).

Non-covalent mechanisms such as chromatin remodeling provide the cell with additional tools for introducing variation into the chromatin template (**Figure 1.7**). ATP-dependent chromatin remodeling complexes are thought to modify

chromatin accessibility by altering histone-DNA interactions, perhaps by sliding or ejecting nucleosomes (Smith and Peterson, 2005). Specifically targeted remodeling complexes may open or close chromatin in a localized area of the genome, or globally remodel chromatin structure throughout the nucleus (Lusser and Kadonaga, 2003; Smith and Peterson, 2005).

Chromatin remodelers are critical for a variety of nuclear processes, including the proper spacing of nucleosomes following replication, access to DNA during DNA repair processes, regulating the exposure of regulatory elements, and the replacement and proper positioning of nucleosomes following transcription (Clapier and Cairns, 2009). For precise control of gene expression, such modulation of nucleosome location and stability may be particularly important in regulating access of transcription factors to regulatory elements, as well regulating access of the transcriptional machinery to promoters (Henikoff, 2008). Mutations in specific chromatin remodelers have been linked to several human developmental disorders, including alpha thalassemia and X-linked mental retardation (Gibbons et al., 2008), a subject that will be discussed in more detail in Chapters Three and Four.

Mechanistic links between covalent and non-covalent mechanisms have also been uncovered, as modification 'effectors' can include subunits of nucleosome remodeling complexes (Wysocka et al., 2006). It is becoming clear that significant 'crosstalk' exists between different epigenetic pathways. In *S. pombe*, for example, biochemical purification of the RNAi machinery revealed an interaction between small RNAs and chromatin regulators such as the H3K9me-reading chromodomain-containing protein Chp1 (Grewal and Jia, 2007). Silencing of the inactive X chromosome is another prime example, where *Xist* RNA, DNA methylation, histone modifications, their 'writers' and 'readers,' all play a role (Heard, 2005). Moreover, DNMTs act in part by interacting directly or indirectly with chromatin modifying enzymes such as histone deacetylases

(Dobosy and Selker, 2001). Recent studies reveal that DNMT3L, an essential accessory protein to the *de novo* DNMTs in the germline, interprets patterns of histone modifications, adding weight to an emerging theme that DNA methylation and histone modifications functionally interact (Freitag and Selker, 2005; Ooi et al., 2007). Covalent modification, nucleosome remodeling, and histone variants (discussed in detail below) work together to introduce meaningful variation into the chromatin fiber, and their collective contribution to epigenetics is only now being rigorously explored.

Chromatin immunoprecipitation, high-throughput sequencing, and genome-wide approaches to chromatin state

The development of chromatin immunoprecipitation (ChIP) has provided a powerful tool to assess the precise localization of specific proteins or covalent modifications in the genome (Kuo and Allis, 1999; O'Neill and Turner, 2003). For many years, histone acetylation was indirectly associated with transcriptionally active areas of the genome. For example, acetylated histones and active genes were both found enriched in more open, accessible regions of the genome that were preferentially released following nuclease digestion (Sealy and Chalkley, 1978; Vidali et al., 1978; Weintraub and Groudine, 1976). However, one possibility was that the use of nucleases released multiple populations of chromatin, one of which was enriched in acetylated histones, and one of which was enriched in active genes (Hebbes et al., 1988). How could one determine if the population of chromatin with acetylated histones actually carried the sequences of active genes? The direct association between histone modifications and active or repressed genes therefore required the development of a new technique. In 1988, Crane-Robinson and colleagues published the first example of ChIP, demonstrating that acetylated histones were directly associated with active genes, and establishing a robust

method that separates chromatin on the basis of the histone modification itself (Hebbes et al., 1988).

ChIP allows one to determine whether a specific histone modification or protein of interest is directly associated with a specific region of the genome. ChIP is generally performed with either formaldehyde-mediated crosslinking of proteins and nucleic acids within living cells followed by sonication (crosslinking, x-link ChIP), or by micrococcal nuclease (MNase) digestion of harvested cell nuclei and isolation of solubilized mono or short oligonucleosomes (native ChIP) (**Figure 1.8**). Crosslinking ChIP ‘freezes’ protein and nucleic acid interactions and enables the analysis of nucleosomal as well as non-nucleosomal proteins such as TFs and RNA polymerase, while native ChIP is more suitable for the analysis of core chromatin and nucleosomal components such as histones and histone modifications (Kuo and Allis, 1999; O’Neill and Turner, 2003). After sonication or nuclease digestion to separate chromatin into oligonucleosomal fragments, an antibody to a specific protein or modification of interest is used to immunoprecipitate nucleosomes or oligonucleosomal fragments that contain the specific protein or modification. Unbound areas of the genome that do not contain the specific protein or modification are washed away. Bound DNA can then be isolated and a variety of techniques can be used to probe for the presence of specific sequences, including blotting with radioactively labeled DNA probes, PCR, or semi-quantitative real-time PCR (**Figure 1.8**).

Recently, the development of microarrays and high-throughput sequencing has allowed ChIP to provide global, genome-wide views of chromatin state in various cell types (Barski et al., 2007; Mendenhall and Bernstein, 2008; Mikkelsen et al., 2007). The hybridization of ChIP DNA to microarrays that tile across a portion of the genome (ChIP-chip) was one early iteration of genome-scale location analysis, and this technology yielded important insights, particularly in model organisms such as

Figure 1.8: Chromatin immunoprecipitation and methods of analysis

For crosslinking chromatin immunoprecipitation (x-link ChIP), proteins and nucleic acids in living cells are first crosslinked with a crosslinker (x-linker) such as paraformaldehyde. For native ChIP, no crosslinker is used. Cells are lysed, and nuclei are sonicated (x-link ChIP) or digested with a nuclease (native ChIP) to break up chromatin into smaller oligonucleosome or mononucleosome fragments. Highly selective antibodies are used to immunoprecipitate specific proteins or modifications of interest. In this case, the antibodies are depicted as previously bound or conjugated to a magnetic bead (left), and the antibodies bind to a specific histone tail modification (small green spheres on histone tails). Chromatin that is not bound by the antibody (right) is washed away. Chromatin fragments containing the protein or modification of interest are kept, and DNA from this chromatin is isolated. This immunoprecipitated DNA can be analyzed in a variety of ways, including dot blots with radiolabeled probes for specific genomic regions (left), PCR with specific primers for genomic DNA (lower left), real-time PCR for relative quantitation (lower center), and massively parallel sequencing (ChIP-seq, right). ChIP-seq DNA reads can then be aligned and mapped to a reference genome of interest, enabling nucleosome level resolution (lower right). Each blue or red dot indicates a read that has been specifically aligned to the + or – strand.

For review of crosslinking and native ChIP, see (Kuo and Allis, 1999; O'Neill and Turner, 2003).

For review of ChIP-seq, see (Barski and Zhao, 2009; Mendenhall and Bernstein, 2008).

Please see Appendix for detailed discussion of ChIP and ChIP-seq, and for specific ChIP-seq protocols.

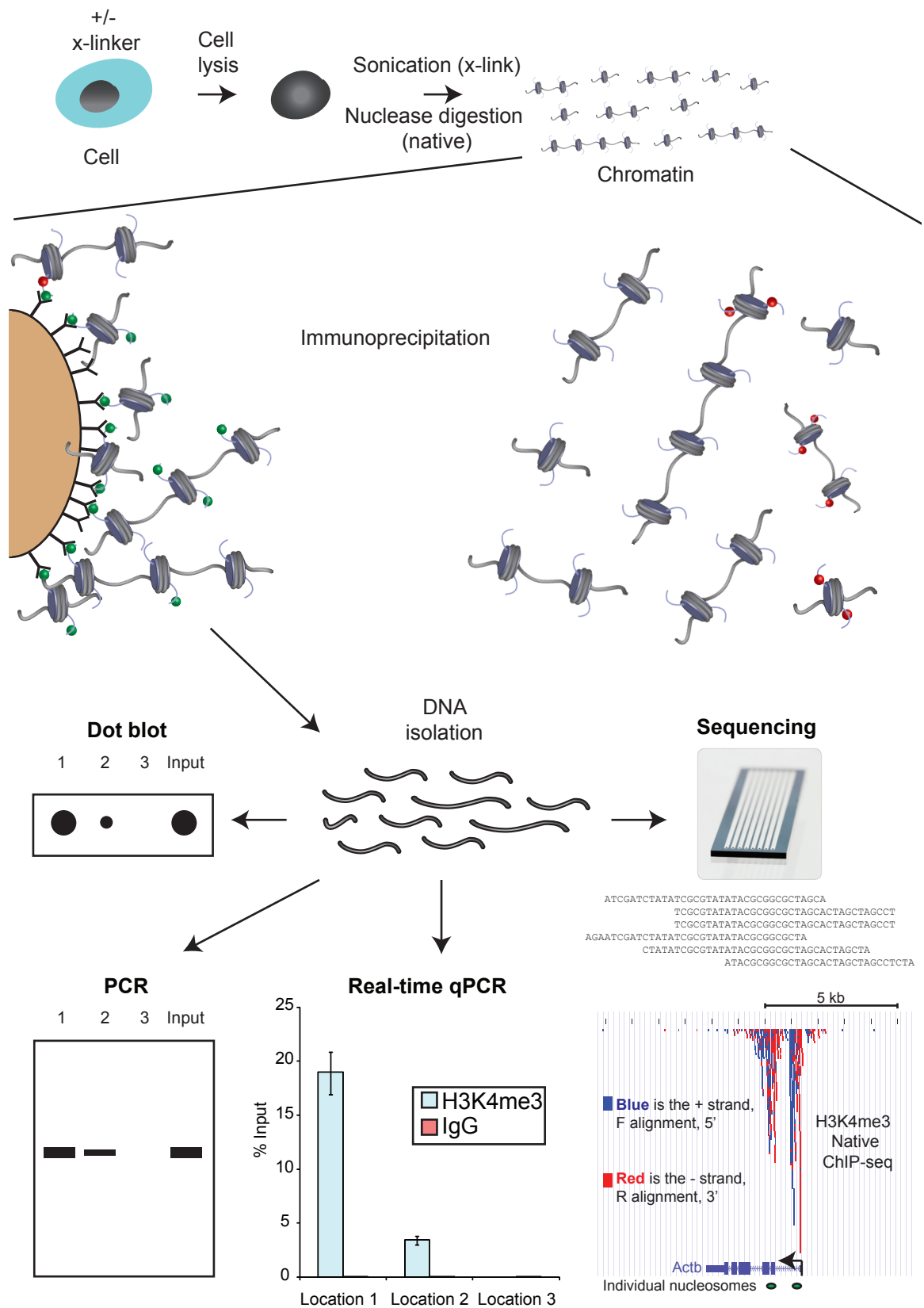


Figure 1.8

yeast where small genome size allowed for interrogation of the entire genome (Lee et al., 2002b; Pokholok et al., 2005; Ren et al., 2000; Yuan et al., 2005). However, for genome-wide analysis of larger, more complex eukaryotic genomes, ChIP-chip has largely been rendered obsolete by the advent of relatively low cost, ultra high-throughput sequencing. Following ChIP and DNA isolation, bound DNA can be directly sequenced using massively parallel sequencing technologies such as those developed by Illumina/Solexa, and this technique is known as ChIP-seq (Barski et al., 2007; Barski and Zhao, 2009; Mikkelsen et al., 2007). Cluster generation and single-molecule-based sequencing produces millions of short (usually 36 bp) sequence tags in a single experiment, and these short sequences can be aligned to a reference genome to generate a genome-wide map (**Figure 1.8**) (Barski and Zhao, 2009). There are multiple advantages of ChIP-seq, including a relatively unbiased approach to genomic location, higher resolution, the analysis of repetitive areas of the genome, and the potential use of sequence polymorphisms to analyze allele-specific incorporation (Barski et al., 2007; Barski and Zhao, 2009; Mikkelsen et al., 2007).

As with any technique, there are limitations to ChIP-seq, particularly in the inherent bias of both crosslinking and native ChIP for the representation of specific, more accessible regions of the genome. This bias makes it advisable to compare the genome-wide pattern of a specific modification or protein to the genome-wide pattern of input chromatin prior to IP, or to the genome-wide pattern of an IgG control (Barski and Zhao, 2009). In addition, crosslinking ChIP can preserve multimeric complexes across large looped regions of DNA, and ChIP-seq data from crosslinking ChIP therefore has the caveat that enrichment might be due to an artifact of looping and protein-protein interactions rather than direct DNA binding (Barski and Zhao, 2009). Finally, like any ChIP experiment from multiple cells, genome-wide patterns from ChIP-seq represent an aggregate of the

chromatin state of all cell types in the examined population. Cell type heterogeneity can therefore produce conflated results from multiple chromatin states, and even the population of a single cell type can represent cells with specific genes at various stages of transcription. For example, populations of undifferentiated mouse embryonic stem cells display heterogeneous expression of multiple genes such as *Dppa3*, *Nanog*, *Rex1*, *Pecam1*, *Brachyury*, and *Twist 2* (Tanaka, 2009). However, despite these limitations, ChIP-seq has emerged as a powerful high-resolution technique for examining genome-wide localization (Barski and Zhao, 2009; Mendenhall and Bernstein, 2008).

Genome-wide localization of nucleosomes

The majority of histone modifications are found in the context of nucleosomal histones that are incorporated into chromatin (Loyola et al., 2006). Various studies have examined the localization of nucleosomes in multiple organisms (Jiang and Pugh, 2009; Schones et al., 2008; Yuan et al., 2005). Most nucleosome positioning studies use MNase digestion followed by DNA hybridization to tiling microarrays or sequencing in order to profile nucleosome locations, or perform ChIP-chip or ChIP-seq with an antibody to a core histone (Schones et al., 2008; Yuan et al., 2005). Nucleosome positioning is non-random, and many aspects of nucleosome position are dictated by the underlying DNA sequence, while nucleosome positioning at specific loci can also be modulated by chromatin modifications, histone variants, and a variety of chromatin remodeling machinery (Field et al., 2009; Henikoff, 2008; Segal et al., 2006; Segal and Widom, 2009). Each nucleosome wraps approximately 146 bp of DNA, and positioned nucleosomes have a characteristic length of short linker DNA between them, which averages ~38bp in humans (Jiang and Pugh, 2009; Luger et al., 1997; Schones et al., 2008).

At most loci, nucleosome position can be represented by a continuous

distribution, while at other loci, highly phased nucleosomes are more precisely positioned (Jiang and Pugh, 2009). The first nucleosome (+1) after the transcription start site (TSS) of both active and inactive genes is well-positioned, although the exact location differs slightly between active and inactive genes (Schones et al., 2008). At active genes in human cells, the 5' end of the +1 nucleosome peaks at +40bp, while the +1 nucleosome peaks at +10bp in inactive genes, a phenomenon that can perhaps be attributed to positional constraints placed by binding of the transcriptional machinery at the TSS of active genes (Schones et al., 2008). Multiple studies have also established that a nucleosome free, or at least nucleosome-depleted region, exists at the TSS of both active and inactive genes, a phenomenon that causes a dip in general histone profiles at the TSS (Jiang and Pugh, 2009; Jin et al., 2009; Schones et al., 2008) (**Figure 1.9**). The level of nucleosome depletion at the TSS is dependent on gene expression, with increased depletion at the TSS of active genes (Schones et al., 2008). Nucleosome depletion at the TSS has recently been attributed to the instability of nucleosome core particles at this region, as well as the presence of specific nucleosome destabilizing DNA sequences at high CpG content promoters (Jin et al., 2009; Ramirez-Carrozzi et al., 2009). Unlike inactive genes, where only the +1 nucleosome is precisely positioned, multiple nucleosomes surrounding the TSS of actively transcribed genes are highly phased, with specific positioning and regular spacing of at least three nucleosomes upstream and five nucleosomes downstream of the TSS (Schones et al., 2008) (**Figure 1.9**). Recent studies have established that nucleosomes are also highly phased at TFBS with bound TFs, suggesting that TF binding may promote regulated nucleosome positioning by chromatin remodeling complexes, potentially modulating TF access to regulatory elements (Henikoff, 2008; Zheng et al., 2009).

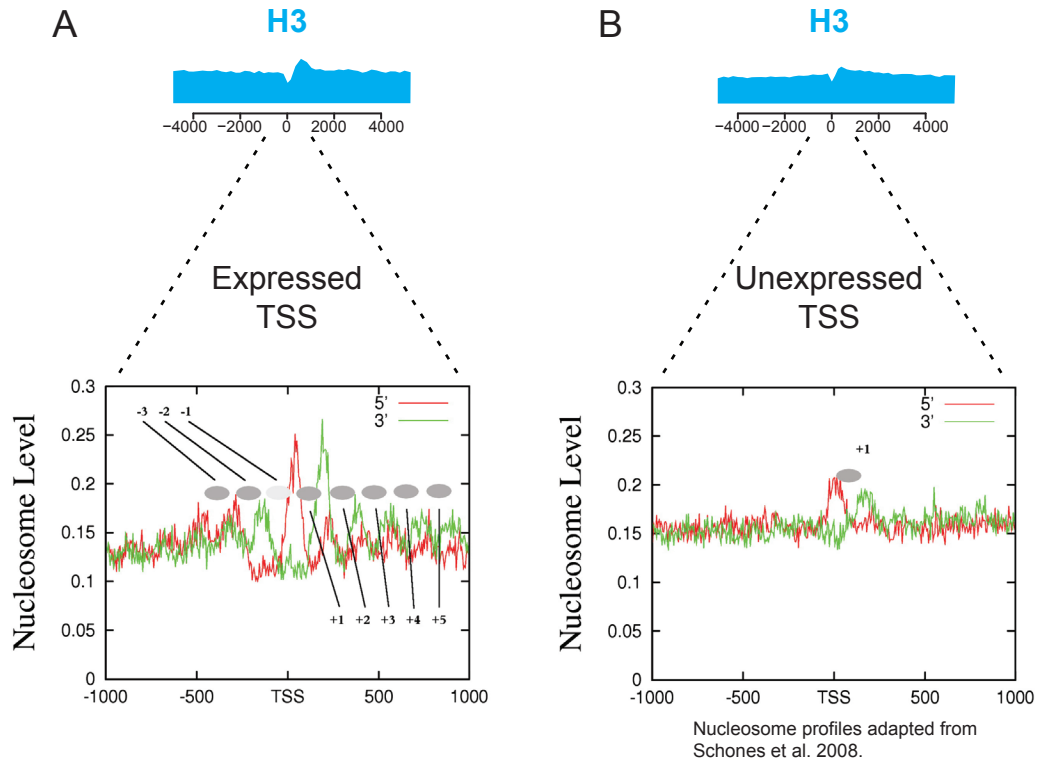


Figure 1.9: Nucleosomes are phased around active transcription start sites

A. Top profile represents the average pattern of general H3 around the TSS of transcriptionally active genes. Note the slight dip of H3 at the TSS itself, corresponding to a localized region of nucleosome depletion. The bottom data represent nucleosome profiles taken from genome-wide sequencing of micrococcal nuclease digested chromatin. Data in this bottom profile and its description are adapted from (Schones et al., 2008). Nucleosomes near the TSS of expressed genes are phased with respect to the TSS. The y axis shows the normalized number of sequence tags from the sense strand (red) and antisense strand (green) of DNA at each position. The inferred nucleosomes are shown by filled ovals that are numbered. The -1 nucleosome (light gray) is lower than the -2 or +1 nucleosome level in both active and inactive genes. This nucleosome depletion has been quantified by (Schones et al., 2008), though a recent study has suggested that this apparent depletion may stem from the fragility of H3.3/H2A.Z nucleosomes near the TSS (Jin et al., 2009).

B. Top profile represents the average pattern of general H3 around the TSS of unexpressed genes. Bottom data represents nucleosome profiles around unexpressed genes, with data represented in similar manner to **A**. The bottom data and legend is also adapted from (Schones et al., 2008). Only one well-positioned nucleosome exists near the TSS of unexpressed genes.

Genome-wide patterns and functions of histone modifications at genes and regulatory elements

The genome-wide localization of histone modifications is intimately related to their function. Using ChIP-chip and ChIP-seq, landmark studies have recently established genome-wide maps of covalent histone modifications in multiple cell

types, including pluripotent ES cells, differentiated NPCs, CD4⁺ T cells, and various immortalized cancer cell lines (Barski et al., 2007; Heintzman et al., 2009; Meissner et al., 2008; Mikkelsen et al., 2007; Wang et al., 2008). Focusing predominantly on covalent modifications of histone H3 and H4 to which highly selective antibodies are available, these genome-wide studies demonstrate that specific histone modifications are highly correlated with gene expression (Barski et al., 2007; Mikkelsen et al., 2007). Transcriptionally active and repressed genes maintain distinct modification-specific patterns across their transcription start site (TSS), gene body, and transcription end site (TES) (**Figure 1.10, 1.11**). Although ChIP-chip and ChIP-seq data typically depict the patterns for individual proteins or histone modifications, it is important to recall that individual histones, nucleosomes, and chromatin domains are highly modified *in vivo*, and the combinatorial nature of multiple weak multivalent interactions are likely to be critical to chromatin biology (Ruthenburg et al., 2007b; Wang et al., 2008). As described, histones are highly covalently modified (Kouzarides, 2007), but only a subset of these modifications have been rigorously studied at the genome-wide level. Therefore, the following section will focus only on the specific genome-wide patterns of histone methyl and acetyl marks that have been well characterized in mammalian cells.

H3K4 methylation

Histone H3 lysine 4 trimethylation (H3K4me₃) is strongly associated with transcription initiation, and represents one of the best studied histone modifications to date (Guenther et al., 2007; Ruthenburg et al., 2007a). H3K4me₃ is thought to serve as a genomic landmark for the transcription start site (TSS) in various biological processes (Guenther et al., 2007). In yeast, H3K4me₃ is found at the 5' end of active but not inactive genes (Santos-Rosa et al., 2002). However, the correlation of H3K4me₃ with gene expression in higher eukaryotes is more complex. High CpG

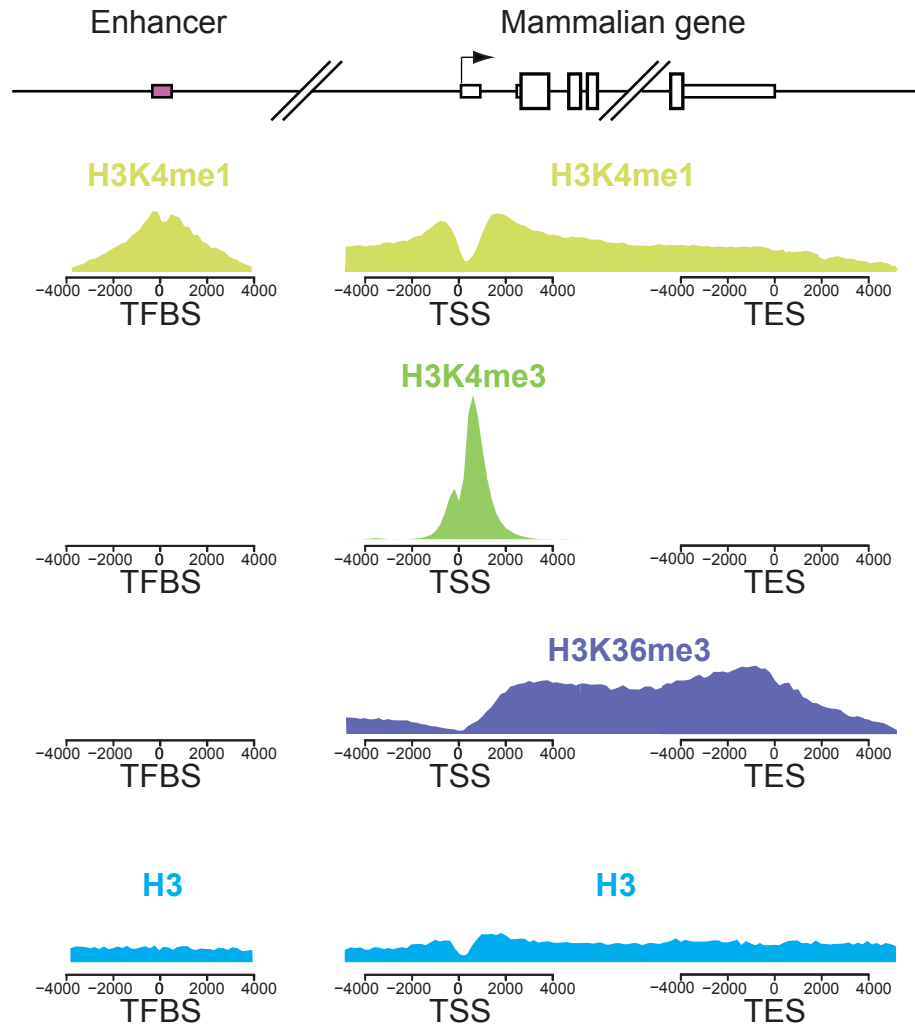


Figure 1.10: Patterns of specific histone H3 methylation marks at active genes

Representative patterns of histone methylation marks and general histone H3 are shown at active genes and regulatory elements. Note that H3K4me1 is uniquely enriched at enhancer elements as well as across active genes. H3K4me3 is more restricted to a few nucleosomes surrounding the TSS, and is relatively unenriched at enhancer elements. In contrast, H3K36me3 marks a broad domain across an actively transcribed gene, which in mammalian cells can span many kilobases. Nucleosomes containing histone H3 are widespread throughout the genome, with specific patterns around TSS, and nucleosome depletion at the TSS itself.

content promoter (HCP) genes in both ES cells and differentiated cells are marked by a peak of histone H3K4me3 and the presence of RNAPII, regardless of whether the gene is active or repressed (Barski et al., 2007; Guenther et al., 2007; Mikkelsen et al., 2007). Notably, the magnitude of the H3K4me3 signal at active and repressed HCP genes still correlates with transcriptional activity, with greater

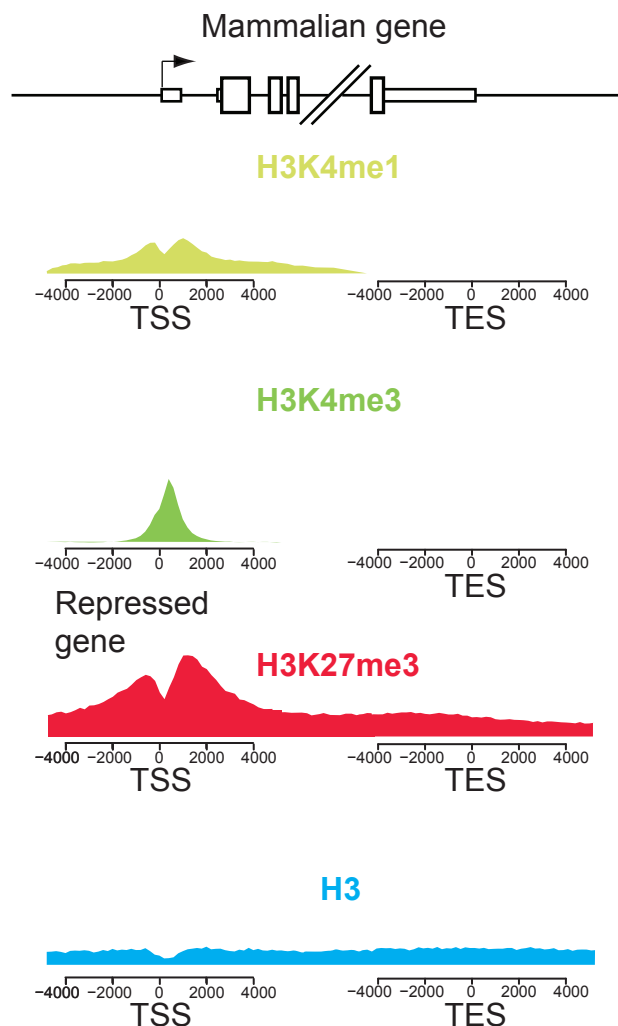


Figure 1.11: Patterns of specific histone H3 methylation marks at repressed genes

Representative patterns of histone methylation marks and general histone H3 are shown at repressed genes. The TSS of repressed genes can be marked with lower levels of H3K4me1 and H3K4me3, and maintenance of H3K4me3 at the TSS may help to poise repressed genes for future activation (Guenther et al., 2007; Wang et al., 2009). Some repressed genes have no detectable H3K4 or H3K27 methylation (not shown) (Mikkelsen et al., 2007), and these genes may be less amenable to future activation. Repressed genes in mammalian cells are also frequently marked with H3K27me3, and this modification is deposited and read by polycomb group proteins (Whitcomb et al., 2007). Note also that repressed genes do not show any enrichment of H3K36me3 depicted in the previous figure. Nucleosomes containing histone H3 are widely distributed throughout repressed genes, with a slight depletion at the TSS.

H3K3me3 signal at active TSS, and lower, but still distinctive peaks of H3K4me3 at repressed HCP gene TSS (Barski et al., 2007; Mikkelsen et al., 2007) (**Figure 1.11**). Although the pattern of H3K4me3 is generally represented by a tight peak at a few nucleosomes following a gene TSS (with specific nucleosome peaks at +50,

+210, and +360 bp), certain regions of the genome such as the Hox loci can show a broader pattern of H3K4me3 in specific developmental contexts, a pattern that may be related to large-scale transcription of intergenic non-coding RNA (Barski et al., 2007; Bernstein et al., 2005; Guenther et al., 2005; Guenther et al., 2007; Guttman et al., 2009; Rinn et al., 2007; Shilatifard, 2008).

H3K4me3 can be recognized by various specific effector proteins to mediate multiple different biological outputs (Guenther et al., 2007; Ruthenburg et al., 2007a; Shilatifard, 2008). For example, in the facilitation of transcription initiation, H3K4me3 can be recognized by PHD finger domains of a component of the NURF chromatin remodeling complex, as well as a PHD domain of the well studied protein TFIID, a component of the basal transcription machinery (Shilatifard, 2008; Vermeulen et al., 2007; Wysocka et al., 2006). In contrast, recognition of H3K4me3 by ING proteins is also important for the repression of proliferation genes following DNA damage, while recognition of H3K4me3 by RAG2 is required for efficient V(D)J recombination (Matthews et al., 2007; Shi et al., 2006). Intriguingly, the H3K4 methyltransferase Set1 is required for ribosomal DNA silencing and telomere length maintenance in yeast, as well as for normal rates of cell growth, but not for yeast viability (Briggs et al., 2001; Dehe and Geli, 2006). In higher eukaryotes, proteins responsible for 'writing' and 'reading' H3K4 methylation are encoded by trithorax group proteins, and H3K4 methylation may play a more important role in fine-tuning and maintaining the positive regulation of gene expression in metazoans (Schuettengruber et al., 2007). Recognition of H3K4me3 by NURF has been functionally linked to the maintenance of Hox gene expression during vertebrate development (Wysocka et al., 2006). In addition, a recent genetic study demonstrated that deletion of the enzymatic SET domain of the H3K4 methyltransferase mixed lineage leukemia (MLL) protein recapitulates the homeotic phenotype in mice, further suggesting a role for H3K4 methylation in

the regulation of epigenetic memory in mammals, though other non-histone targets of MLL could also be functionally important (Terranova et al., 2006).

Different methylation states on the same histone amino acid can show distinct genome-wide patterns of localization and distinct functions. For example, in contrast to H3K4me3, H3K4me1 shows a distinct pattern, with enrichment surrounding the TSS of active genes, a dip at the peak of H3K4me3 nucleosomes, and a broader distribution across the gene body of active genes (**Figure 1.10**) (Barski et al., 2007; Meissner et al., 2008). Unlike H3K4me3, H3K4me1 has recently been shown to have a strong genome-wide co-localization with TFBS, enhancers, and regulatory elements (Heintzman et al., 2007; Robertson et al., 2008; Wang et al., 2008) (**Figure 1.10**). H3K4me2 also shows a pattern that is similar to H3K4me1, with enrichment surrounding the TSS, depletion at the peak of H3K4me3, and broader enrichment into the gene body of active genes, as well as enrichment at regulatory elements (Barski et al., 2007). Specific functions for H3K4me1 and H3K4me2 at regulatory elements have not been determined, although H3K4me2 was recently shown to play a role in the recruitment of HDACs and the suppression of nucleosome acetylation at the 5' end of transcribed genes in yeast, just downstream of the high levels of nucleosome acetylation at H3K4me3 near the TSS (Kim and Buratowski, 2009).

H3K36 methylation

H3K36me3 is a robust marker of actively transcribed gene bodies (Barski et al., 2007; Mikkelsen et al., 2007). In contrast to the tight peaks of a few nucleosomes marked by H3K4me3, H3K36me3 is often found over wide transcribed domains, particularly in larger eukaryotic genomes where smaller exons can be separated by long introns of many kilobases (Barski et al., 2007; Mikkelsen et al., 2007) (**Figure 1.10**). H3K36me3 is a highly conserved histone modification, and it has

been associated with transcriptional elongation in a variety of eukaryotes, from yeast to mice and humans (Bannister et al., 2005; Barski et al., 2007; Mikkelsen et al., 2007). Intriguingly, despite its association with actively transcribed genes, H3K36me3 has been functionally linked to transcriptional repression, and has specifically been shown to function in the repression of spurious intragenic transcripts (Carrozza et al., 2005; Keogh et al., 2005). Set2-mediated H3K36me3 is required for the recruitment of the HDAC Rpd3S complex to gene bodies in yeast, leading to the model that H3K36me3 prevents aberrant transcriptional initiation at sites other than the appropriate TSS, particularly at long or infrequently transcribed genes (Carrozza et al., 2005; Li et al., 2007a; Li et al., 2007c). The presence of unannotated H3K4me3 peaks followed by regions of H3K36me3 have been recently used to identify novel transcripts and alternative TSS, as well as thousands of large intergenic regions that give rise to non-coding RNA transcripts (Guttman et al., 2009; Mikkelsen et al., 2007).

Histone lysine acetylation

For decades, histone acetylation has been associated with gene activation, while histone deacetylation has been associated with gene repression (Struhl, 1998). Histone acetylation can have both direct and indirect effects on chromatin (Shahbazian and Grunstein, 2007). Histone acetylation can be recognized by bromodomains within specific effector proteins and chromatin regulatory complexes, but can also directly regulate the degree of chromatin compaction, and is thought to help overcome the inhibitory effects of nucleosomes on transcription (Dhalluin et al., 1999; Shahbazian and Grunstein, 2007). Like histone methyl marks, different histone acetyl marks are localized to distinct regions of active genes (Wang et al., 2008). Histone acetyl marks such as H3K9ac and H3K18ac are enriched in regions surrounding the TSS of active genes, while acetyl marks such as H4K12ac

and H4K16ac are enriched in promoter and transcribed regions of active genes (Wang et al., 2009b; Wang et al., 2008). Certain histone acetyl marks, such as H3K18ac but not H3K14ac, are also enriched in a significant fraction of enhancers (Wang et al., 2008). Genome-wide maps of histone acetylation, HATs, and HDACs have recently been established, and these data reveal an unexpected association of both HATs and HDACs with active genes (Wang et al., 2009b). Surprisingly, the majority of HDACs are associated with active and poised genes, although some HDACs are also associated with intergenic enhancers, and no HDACs were found enriched at silent genes (Wang et al., 2009b). Current models suggest that most HDACs function to remove acetylation from active genes, 'resetting chromatin' from acetyl marks laid down by HATs associated with elongating polymerase in order to maintain a level and localization of acetylation ideal for accurate transcription (Wang et al., 2002; Wang et al., 2009b).

Histone H3K27 methylation marks repressed genes

In contrast to active genes, repressed genes are often marked by H3K27me3 (Schuettengruber et al., 2007). Genome-scale studies have established that H3K27me3 can be found at repressed genes either in broad domains, such as at key developmental loci including the Hox genes, or in a more narrow domain surrounding repressed TSS (Mikkelsen et al., 2007) (**Figure 1.12**). As mentioned above, the 'writing' and 'reading' of H3K27 methylation is mediated by components of the evolutionarily conserved polycomb group (PcG) proteins that are associated with Hox gene repression (Whitcomb et al., 2007). Mammalian homologs of the polycomb protein that recognizes H3K27 methylation in *Drosophila* are encoded by the Cbx proteins, and these proteins have been found to play diverse roles in the regulation of cell cycle, homeotic, and developmental genes (Bernstein et al., 2006b; Whitcomb et al., 2007). H3K27me3 and polycomb proteins are enriched

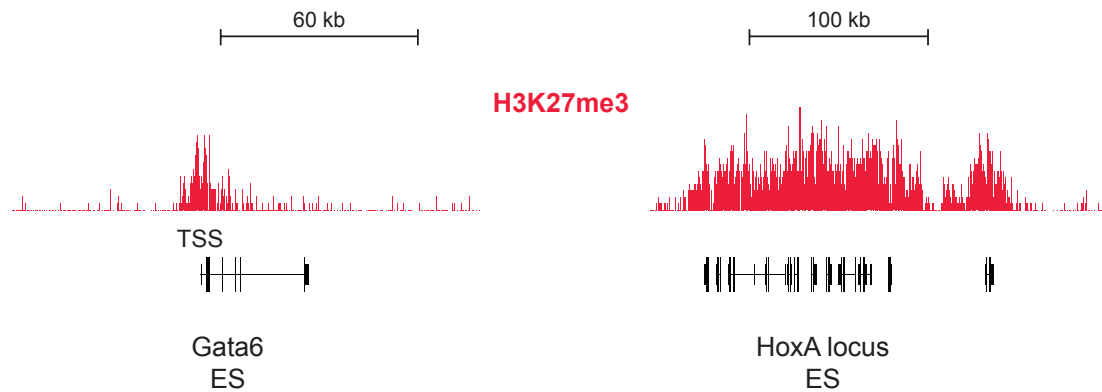


Figure 1.12: Different genomic patterns of H3K27me3

Two general patterns of H3K27me3 are found at repressed genes in mammalian cells. H3K27me3 can be found in relatively tight regions surrounding the TSS, as in the case of the repressed gene *Gata6* in ES cells (left). In contrast, larger regions such as the repressed *Hox* loci in ES cells can be marked by broader domains of H3K27me3 (right).

at key developmental genes in ES cells, and polycomb proteins play critical roles in gene repression and the maintenance of cellular state (Boyer et al., 2006; Lee et al., 2006a). The PcG protein and H3K27 methyltransferase Ezh2 is required for mammalian development, and Ezh2 and its homolog Ezh1 play complementary roles in mediating H3K27me3 and transcriptional repression at key developmental genes (Ho and Crabtree, 2008; O'Carroll et al., 2001; Shen et al., 2008). The PcG protein Eed is required for global H3K27me2 and H3K27me3, and loss of Eed leads to depression of multiple differentiation-specific PcG target genes in undifferentiated ES cells (Boyer et al., 2006; Montgomery et al., 2005). More recently, Ezh2 has been shown to play an important role in tissue-specific stem cells, repressing specific proliferative and structural genes to fine-tune epidermal differentiation (Ezhkova et al., 2009).

Cell type specific differences in chromatin state

Chromatin state has been shown to vary with cellular differentiation state. As described previously (**Figure 1.1**), pluripotent embryonic stem (ES) cells have the

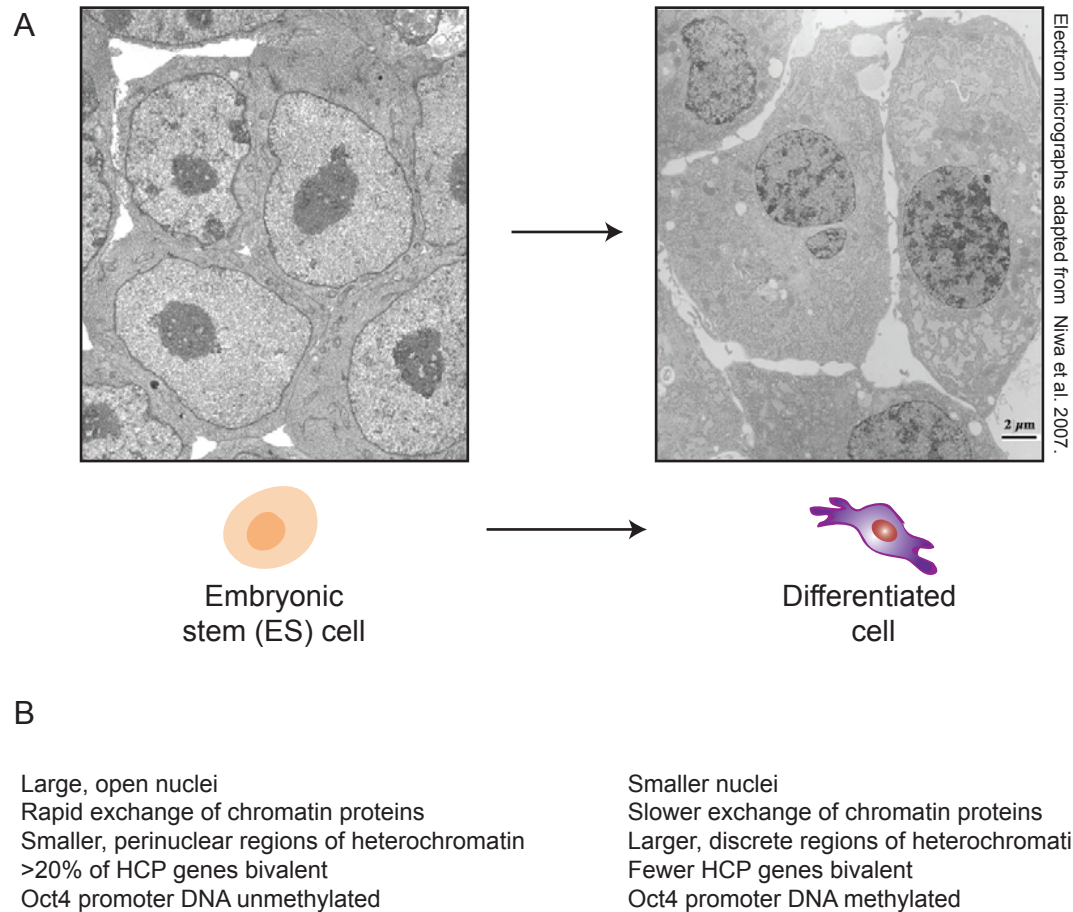


Figure 1.13: Embryonic stem cells and differentiated cells display distinct chromatin states

A. Electron micrographs of undifferentiated ES cells (left) and differentiated ES cells (right). Note the large proportion of open euchromatin in the ES nucleus. Following differentiation into primitive endoderm, the nucleus shrinks, and the distribution of electron-dense heterochromatin changes dramatically.

B. Characteristic features of ES and differentiated cell chromatin. A variety of evidence suggests that undifferentiated ES and iPS cells maintain a uniquely open and plastic epigenome (Bernstein et al., 2007; Efroni et al., 2008; Gaspar-Maia et al., 2009; Meshorer et al., 2006; Niwa, 2007).

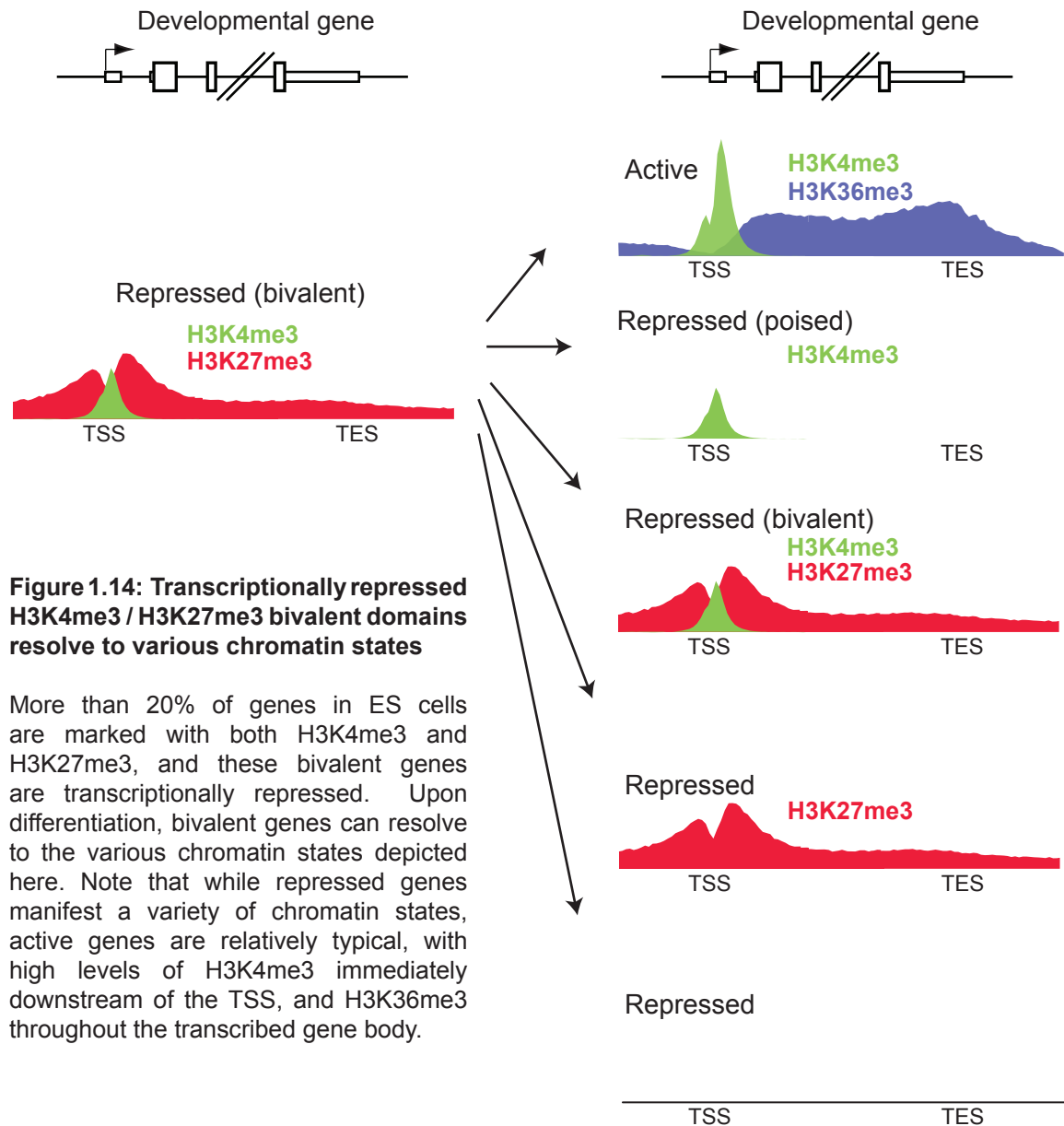
capacity to give rise to all differentiated cell types of the vertebrate embryo, and they manifest a unique genome-wide chromatin state (Bernstein et al., 2007; Niwa, 2007). In particular, pluripotent ES cells have large and open nuclei with smaller regions of heterochromatin, display hyperdynamic binding of structural chromatin proteins, and exhibit global transcriptional hyperactivity (Efroni et al., 2008; Gaspar-Maia et al., 2009; Meshorer et al., 2006) (**Figure 1.13**). ES cells also utilize unique

chromatin remodeling complexes to maintain the pluripotent state. For example, the esBAF chromatin remodeling complex is required for self-renewal and pluripotency of mouse ES cells, but not for proliferation of differentiated fibroblasts (Ho et al., 2009a; Ho et al., 2009b). In addition, the chromatin remodeling factor Chd1 is required to maintain open, hyperdynamic chromatin and ES cell pluripotency (Gaspar-Maia et al., 2009).

H3K4me3 and H3K27me3 bivalent domains mark specific repressed genes

As they have not yet committed to a particular lineage, ES cells must repress key developmental lineage-specific genes in order to maintain the pluripotent state (Chi and Bernstein, 2009; Jaenisch and Young, 2008). Some of the repression of key developmental genes and upregulation of pluripotency genes can be attributed to reinforcing gene expression programs driven by specific TFs such as Oct4, Sox2, and Nanog, as well as the expression of particular microRNA genes (Chen et al., 2008b; Jaenisch and Young, 2008; Marson et al., 2008; Niwa, 2007). In addition, recent studies have established that ES cells maintain unique global patterns of specific covalent histone modifications at key lineage-specific genes (Barski et al., 2007; Bernstein et al., 2006a; Meissner et al., 2008; Mikkelsen et al., 2007). More than one-fifth of HCP genes in ES cells carry both H3K4me3 and H3K27me3 in their promoters (Bernstein et al., 2006a; Mikkelsen et al., 2007) (**Figure 1.14**). These 'bivalent' genes are transcriptionally repressed in ES cells, and are enriched in genes encoding key developmental TFs and cell surface proteins, leading to the model that they represent repressed genes that are poised for activation following cell differentiation (Bernstein et al., 2006a; Mikkelsen et al., 2007).

Upon cell differentiation, most H3K4me3 / H3K27me3 bivalent genes resolve either to H3K4me3, H3K27me3, or neither mark, though a minority of these genes remain bivalent (**Figure 1.14**) (Bernstein et al., 2006a; Mikkelsen et al., 2007).



Although most bivalent domains are located in punctuate regions around specific TSS, certain regions of the genome such as the Hox loci are enriched in large-scale bivalent domains (**Figure 1.15**) (Mikkelsen et al., 2007). As bivalent gene promoters are transcriptionally repressed, but enriched in both H3K4me3 and H3K27me3, these data suggest that repressive PcG activity is dominant over activating TrxG activity (Bernstein et al., 2006a; Mikkelsen et al., 2007). Bivalent domains can be divided into two classes based on their occupancy by PcG proteins (Ku et al.,

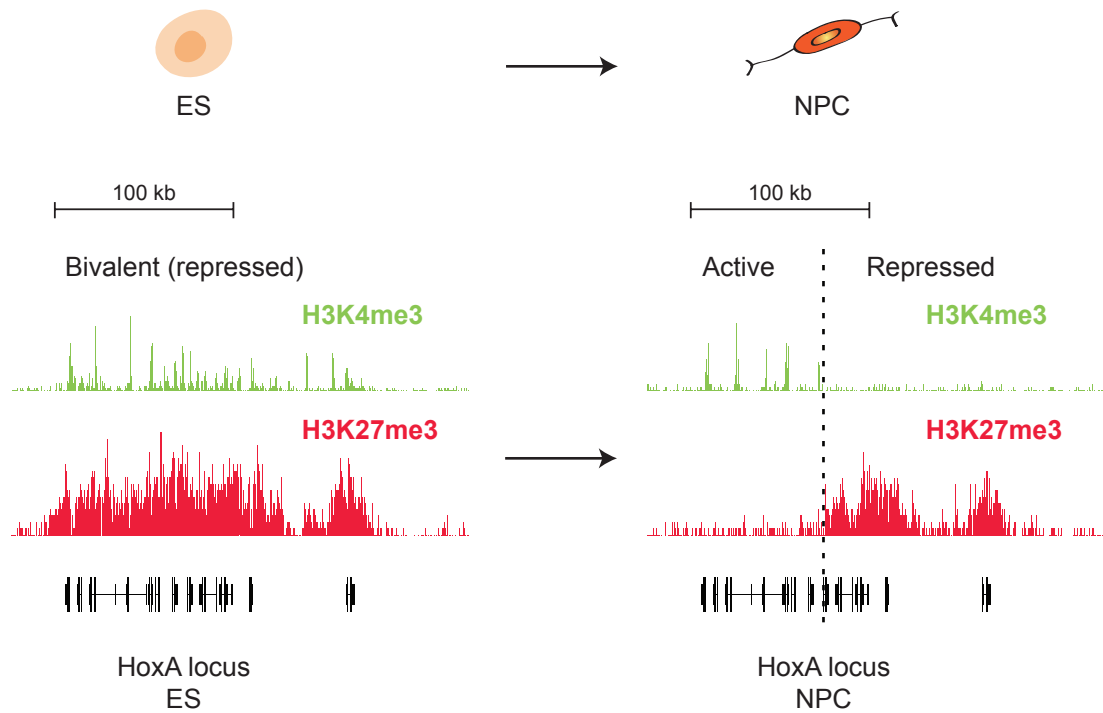


Figure 1.15: Large-scale resolution of bivalent domains at the HoxA locus

In undifferentiated ES cells (left), the HoxA locus is bivalent and marked by widespread H3K4me3 and H3K27me3. Following differentiation into NPCs (right), the large bivalent domains at the HoxA locus resolve into a region of active HoxA genes marked by H3K4me3 (left), and a region of repressed HoxA genes marked by H3K27me3 (right). H3K4me3 and H3K27me3 ChIP-seq profiles in ES cells were generated by Bernstein and colleagues (Mikkelsen et al., 2007), and are publicly available from <http://www.ncbi.nlm.nih.gov/geo/>.

2008). Bivalent domains occupied by the polycomb repressive complex 1 (PRC1) protein Ring1b are particularly enriched in developmental regulator proteins, and these PRC1 positive bivalent domains are more likely to maintain H3K27me3 upon differentiation (Ku et al., 2008). Although bivalent domains are not exclusive to pluripotent cells and have been observed at lower frequency in differentiated cells, ES cells are significantly enriched in bivalent domains relative to all other cell types examined (Barski et al., 2007; Mikkelsen et al., 2007). Moreover, generation of iPS cells also re-establishes bivalent domains at key developmental genes, further suggesting that bivalent domains can reflect developmental potential (Mikkelsen et al., 2008).

Although differentiated cells have a small number of repressed genes

marked by both H3K4me3 and H3K27me3, a subset of genes in differentiated cells marked only by H3K4me3 also remain surprisingly repressed (Guenther et al., 2007; Wang et al., 2009b). These poised genes or 'primed' genes are transcriptionally repressed, but still marked by the presence of H3K4me3, H2A.Z, and low levels of histone acetylation near their TSS (Wang et al., 2009b). Poised genes rapidly gain histone acetylation and increased RNAPII binding following treatment of cells with HDAC inhibitors, suggesting that transient binding and cycling of HAT and HDAC activity helps to keep these genes repressed but primed for future activation (Wang et al., 2009b). The emerging picture of gene expression in mammalian cells is therefore one in which there is a characteristic pattern of histone modifications at active genes, and various patterns at poised, bivalent, or silent genes (**Figure 1.14**). Evidence suggests that multiple genes in undifferentiated and differentiated cells are repressed, but poised for expression (Guenther et al., 2007; Mikkelsen et al., 2007; Wang et al., 2009b). In ES cells, bivalent domains may enable large numbers of genes to be poised with trithorax H3K4me3 activity and H2A.Z, but the use of polycomb and H3K27me3 ensures an additional level of transcriptional repression prior to differentiation.

Patterns of histone modifications at regulatory elements can also be highly predictive of cellular state, and often vary more significantly than histone modifications at promoters. A recent study has established that although histone modifications at many promoters are largely invariant across cell types with different patterns of gene expression, patterns of histone modifications at enhancers are highly cell-type specific, and correlate strongly with global cell-type specific gene expression programs (Heintzman et al., 2009). Taken together, recent genome-wide studies demonstrate that different cell types use TFs, chromatin remodeling machinery, DNA methylation, and histone modifications to reinforce cellular gene expression programs, patterns of specific histone modifications are highly

correlated with gene expression at promoters and enhancers, and genome-wide maps of histone modifications and DNA methylation reflect lineage commitment and cell state (Barski et al., 2007; Bernstein et al., 2006a; Boyer et al., 2006; Heintzman et al., 2009; Meissner et al., 2008; Mikkelsen et al., 2007). However, chromatin modifications and remodeling processes are not the only methods used by cells to establish variation in the chromatin fiber.

Histone variants

In addition to chromatin modifications and nucleosome remodeling, eukaryotic cells also generate variation in chromatin by the introduction of variant histone proteins (**Figure 1.7**). Primary sequence variants of histones have been known for some time (Allis et al., 1980; Allis et al., 1982; Bustin and Cole, 1968; Franklin and Zweidler, 1977; Kinkade and Cole, 1966). Excluding histone H4, all core histone proteins in mammals (H1, H2A, H2B, and H3) have several primary sequence variants (Hake and Allis, 2006; Pusarla and Bhargava, 2005). These variants can be classified into three general subtypes: replication-dependent, replication-independent, and tissue-specific, based on their expression during the cell cycle, tissue-specific expression, and their genomic organization (Albig and Doenecke, 1997; Frank et al., 2003; Hake and Allis, 2006).

During S-phase of the cell cycle, a eukaryotic cell must rapidly synthesize histones in stoichiometric amounts ($\sim 10^8$ molecules of each histone protein in a mammalian cell) for the formation of chromatin with newly replicated DNA (Albig and Doenecke, 1997; Marzluff et al., 2002). These cell-cycle regulated histones are known as replication-dependent (RD) histones, and they account for the bulk of histone synthesis in dividing cells (Albig and Doenecke, 1997). RD histone genes are found in large, multigene, intronless clusters in the genomes of all metazoans, and their transcripts lack a poly-A tail, ending instead in a highly conserved stem-

loop sequence (Heintz et al., 1981; Marzluff et al., 2002).

In contrast, replication-independent (RI) or so-called replacement histones are synthesized throughout the cell cycle, independently from the replication of DNA, and they are also synthesized in terminally differentiated or quiescent cells (Frank et al., 2003). RI histone genes are solitary, typically intron-containing genes located outside of the major histone gene clusters (Albig and Doenecke, 1997; Carozzi et al., 1984; Heintz et al., 1981; Marzluff et al., 2002). RI histone gene transcripts generally have long 5' and 3' UTRs, and are frequently polyadenylated (Frank et al., 2003). In addition to RD and RI histones, metazoans also express a number of tissue-specific (TS) histones. Some of the most highly studied examples of TS histones are expressed during specific stages of gametogenesis, such as the oocyte specific H1 variant H1oo (Tanaka et al., 2001), and the testis specific H1 variants H1t and HILS1 (Drabent et al., 1996; Iguchi et al., 2003; Yan et al., 2003).

Specific histone variants have been associated with particular biological processes. For example, histone H2A.Z was first found specifically enriched in transcriptionally active chromatin in *Tetrahymena* (Allis et al., 1986). More recently, H2A.Z has been observed to be enriched at the transcriptional start sites of both active and repressed genes and regulatory regions in specific organismal and developmental contexts, and has been shown to play a variety of diverse roles including the silencing of developmental genes in mouse ES cells, the suppression of antisense RNA transcripts in *S. pombe*, and the protection of euchromatin from the ectopic spread of heterochromatin in *S. cerevisiae* (Creyghton et al., 2008; Jin et al., 2009; Meneghini et al., 2003; Raisner et al., 2005; Zofall et al., 2009). Histone macroH2A has recently been shown to co-localize with Polycomb complex members, and to facilitate the fine-tuning of Hox gene expression during embryonic development (Buschbeck et al., 2009). In contrast, histone H2A.X is

specifically phosphorylated in response to DNA damage, and this “guardian of the genome” helps to facilitate the assembly of DNA repair complexes and maintain genomic integrity (Celeste et al., 2002; Fernandez-Capetillo et al., 2004; Rogakou et al., 1998). Variants of histone H1 are required for DNA methylation at specific imprinted genes, and for the maintenance of global chromatin structure (Fan et al., 2003; Fan et al., 2005; Happel and Doenecke, 2009).

H3.3 and the Histone H3 Variant Family

Recently, attention has focused on variants of histone H3, a highly evolutionarily conserved family of histones whose members appear to be intimately associated with a variety of critical biological processes, ranging from gene regulation and centromeric function, to gametogenesis and early development (Govin et al., 2004; Hake and Allis, 2006; Hodl and Basler, 2009; Howman et al., 2000; Loppin et al., 2005).

Evolution of H3 variants

Mammals express at least five different types of H3 variants: a centromere-specific H3 variant CENP-A, replication-dependent H3.1 and H3.2, replication-independent H3.3, and a testis-specific variant H3.1t (Hake and Allis, 2006). The evolutionary conservation of non-centromeric H3 variants, and particularly of H3.3, is remarkable (Malik and Henikoff, 2003). Outside of a centromeric H3 variant, the budding yeast *S. cerevisiae* possesses only one H3 protein, and this H3 variant is most closely related (>91% identity) to H3.3 in mammals (Malik and Henikoff, 2003). In yeast, the so-called H3.3 variant therefore serves all functions of histone H3. Indeed, there is no known eukaryote which expresses H3.2 or H3.1 without also expressing H3.3 (Malik and Henikoff, 2003). Over evolutionary time, various branches of the eukaryotic tree have independently evolved distinct replication-

independent H3.3 and replication-dependent H3 variants on multiple occasions, with replication-dependent H3 variants likely expanding in number to package larger eukaryotic genomes, and H3.3 perhaps serving a more specialized function (Malik and Henikoff, 2003; Thatcher et al., 1994; Waterborg and Robertson, 1996).

H3.3 differs from H3.2 and H3.1 at a small number of amino acids

Histone H3.3 differs from H3.2 and H3.1 at only 4 or 5 amino acids (**Figure 1.16A**). Intriguingly, the location of these differences suggests that only one of these amino acids (H3.3 serine 31) is accessible in the context of nucleosomal H3.3 (Davey et al., 2002; Henikoff and Ahmad, 2005; Luger et al., 1997) (**Figure 1.16B**). The H3 variant used in published structures of the nucleosome core particle was H3.2 (Davey et al., 2002; Luger et al., 1997). Examination of the nucleosome structure indicates that the remainder of the differences between H3 variants (at positions 87, 89, 90, and 96) are buried within the H3 α -helix 2 beneath the DNA double helices, and therefore inaccessible in an intact nucleosome (Henikoff and Ahmad, 2005). However, prior to nucleosome formation, pre-deposition machinery would still be able to access and distinguish these amino acids in free H3.3 (Henikoff and Ahmad, 2005).

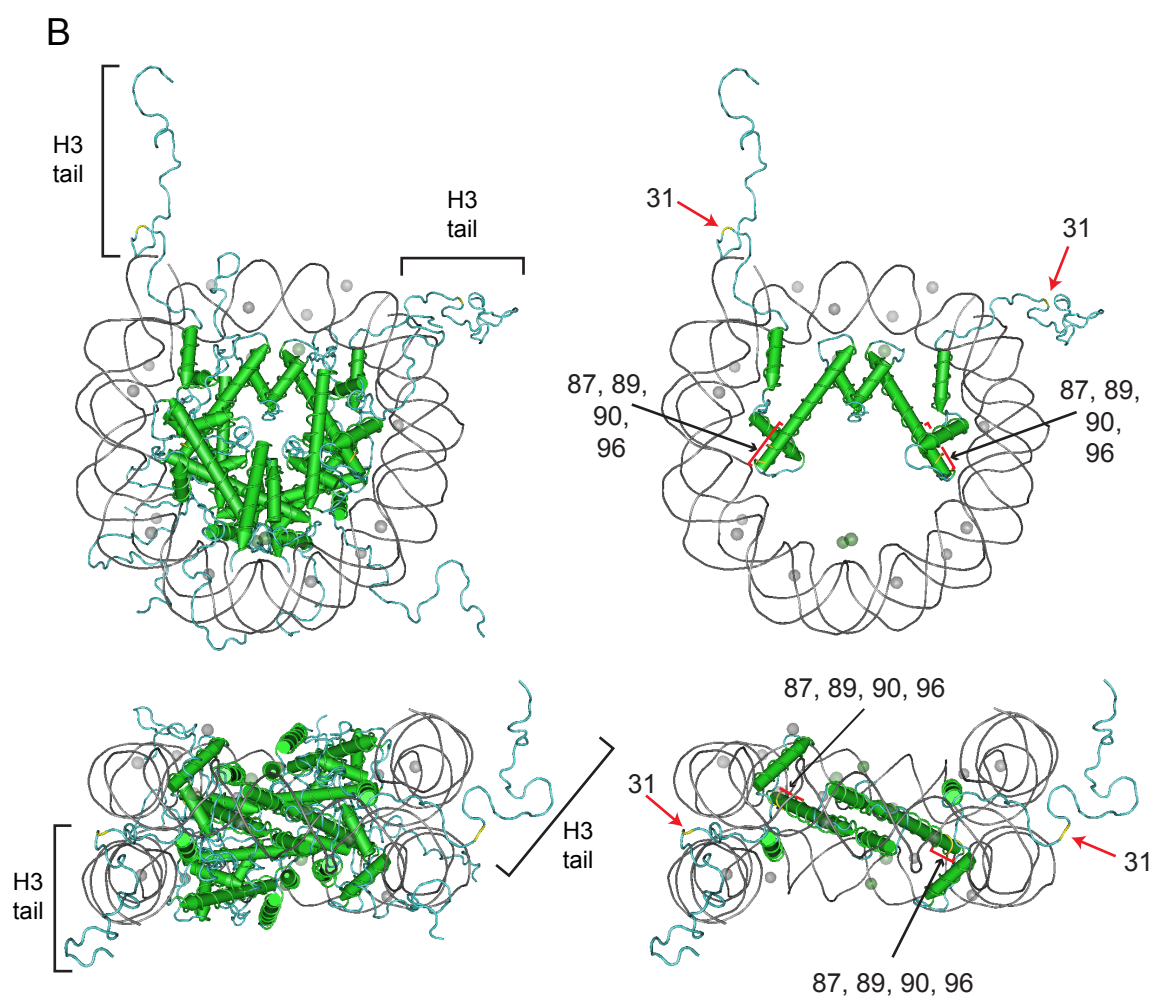
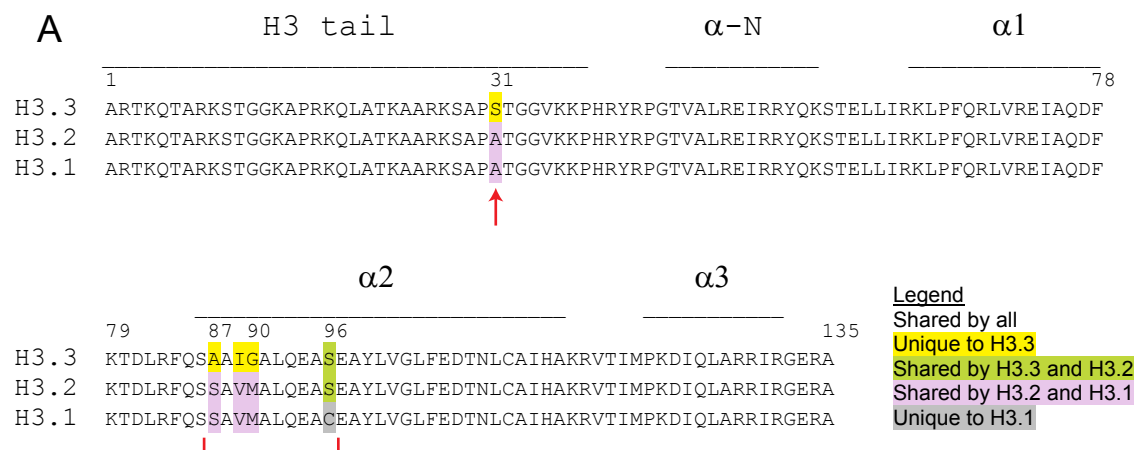
H3.3 nucleosomes are less stable than H3.1 nucleosomes

Recent studies by Jin and Felsenfeld indicate that H3.3 nucleosomes may be intrinsically less stable than nucleosomes containing H3.2 (Jin and Felsenfeld, 2007) (**Figure 1.17**). Nucleosomes containing H3.3 are unusually sensitive to salt concentration (Jin and Felsenfeld, 2007). Following exposure to moderate salt concentrations of 80mM, H3.3 nucleosomes isolated from chicken cells preferentially lost H2A/H2B and H2A.Z/H2B, and H2A.Z in particular was not even detectable. However, following nucleosome isolation under conditions of lower salt

Figure 1.16: Sequence of H3 variants: most distinguishing H3 variant amino acids are not accessible in the context of the nucleosome

A. Sequence alignment of mammalian H3.3, H3.2, and H3.1 histone variants. Histone H3 structural elements such as the H3 tail and alpha helices are indicated above the sequences (Luger et al., 1997). H3.3 differs from the replication-dependent (RD) variant H3.1 at two locations: position 31 in the H3 tail (red arrow), and at positions 87, 89, 90, and 96, a region towards the beginning of the second alpha helix (red bracket). The RD variants H3.2 and H3.1 are identical, with the exception of position 96.

B. Location of H3 variant amino acid differences in the context of the nucleosome. The nucleosome core particle as crystallized by Luger and colleagues is shown here (Davey et al., 2002; Luger et al., 1997). Core histone helices are shaded green. Histone protein backbones, including histone tails, are shaded light blue. The nucleotide backbone of the surrounding DNA double helix is shaded gray. The two views on the left represent the complete nucleosome core particle. Accessible histone H3 tails are indicated by black brackets (left). In the two views on the right, all other histones except H3 have been removed for purposes of visualization. The location of positions 31 (red arrows), as well as positions 87, 89, 90, and 96 (red brackets) are indicated. Note that only position 31 in the H3 tail is accessible in the context of the nucleosome. The remainder of the distinguishing H3 variant amino acids are buried and inaccessible in an intact nucleosome, though they may impact nucleosome stability (Jin and Felsenfeld, 2007).



Structure adapted from Davey et al. *J Mol Biol* 2002.

Figure 1.16

(10mM), some H2A.Z was detectable with H3.3 nucleosomes, and H2A was not preferentially depleted relative to H3.1 nucleosomes (Jin and Felsenfeld, 2007). Relative quantitation of these stabilities revealed that H3.3/H2A.Z nucleosomes were even more unstable than H3.3/H2A nucleosomes (Jin and Felsenfeld, 2007) (**Figure 1.17**). This instability of H3.3 does not appear to depend on histone

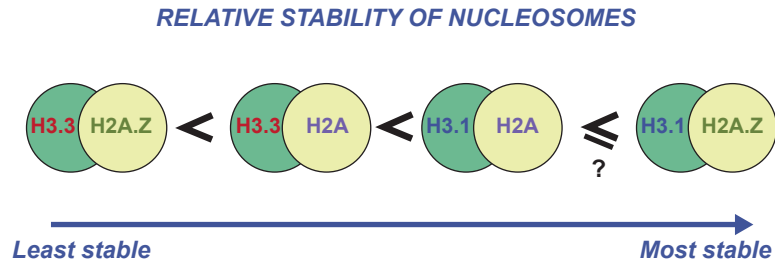


Figure 1.17: Relative stability of H3 variant nucleosomes

Schematic representation of relative stability of nucleosomes containing different histone variants, as assessed by Felsenfeld and colleagues (Jin and Felsenfeld, 2007).

acetylation, as identical results were observed in the presence and absence of the HDAC inhibitor sodium butyrate (Jin and Felsenfeld, 2007). Intriguingly, H3.3/H2A.Z nucleosomes exist in vivo, and are enriched in differentiated vertebrate cells in specific areas of active chromatin, including active promoters and enhancer elements (Jin and Felsenfeld, 2006, 2007; Jin et al., 2009). The relative instability of H3.3 nucleosomes, and particularly of H3.3/H2A.Z nucleosomes, has been suggested to reduce the energy required for nucleosome mobilization at dynamic regions of the genome (Jin and Felsenfeld, 2007). Notably, this hypothesis has yet to be rigorously demonstrated biochemically, as studies of nucleosome stability have not yet been carried out with recombinant histone variants and remodeling machinery. Overall, however, the results of Jin and Felsenfeld suggest that H3.3 may not simply function as an identical replacement variant, but may play a unique role in active chromatin.

Histone H3 genetics in mammals

In mice, the histone H3 variants H3.2 and H3.1 are encoded by 12 intron-less genes located in two clusters (Marzluff et al., 2002). These multiple H3.2 and H3.1 genes are expressed in all cycling cells, and their transcripts end in the highly conserved stem-loop sequence characteristic of replication-dependent histone genes (Marzluff et al., 2002; Plumb et al., 1983) (**Figure 1.18**). In contrast, H3.3 is encoded by two different genes (H3.3A and H3.3B) that code for the same

2 genes code for H3.3



12 genes in 2 clusters code for H3.2 and H3.1

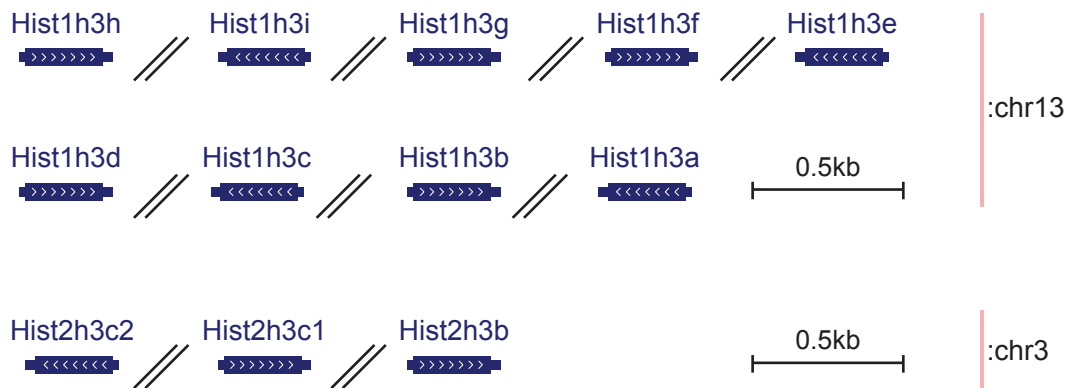


Figure 1.18: Genomic organization of mouse H3 variant genes

Genomic organization of mouse replication-independent H3.3 genes (top) and replication-dependent (RD) H3 genes (bottom) is depicted. Gene names are indicated in blue above the gene, and chromosomal locations are shown on the right. Coding exons are depicted by thick blue rectangles, non-coding exon sequence by thin blue rectangles. Note the presence of introns (thin hashed lines) in H3.3A and H3.3B. The RD histone genes do not contain introns. RD histone H3 genes are found in two clusters on mouse chromosomes 13 and 3, along with other RD histone H2B, H2A, and H4 genes (not shown). For more information on histone genomic organization in mammals, see (Marzluff et al., 2002)

amino acid sequence, but differ in their 5' and 3' UTRs, and also in their patterns of developmental expression (**Figure 1.18**) (Albig et al., 1995; Bramlage et al., 1997; Lopez-Fernandez et al., 1997; Wells et al., 1987). Unlike H3.2 and H3.1, the expression of H3.3 genes is replication-independent, and H3.3 has long been established as the only H3 variant synthesized in at high levels in quiescent, G1, and G2 cells (Wu et al., 1982).

Genomic localization of H3.3 at genes and regulatory elements

Despite the very small number of amino acid differences between non-centromeric H3 variants, groundbreaking work demonstrates that at least three of these amino acids (A87, I89, G90 in H3.3) are critically important for regulating the genomic localization of H3 variants in non-mammalian systems (Ahmad and Henikoff, 2002; Mito et al., 2005, 2007; Schwartz and Ahmad, 2005). Multiple lines of experimental evidence suggest that H3.3 plays a role in transcriptionally active chromatin in multicellular organisms. Using GFP-tagged histones H3.2 and H3.3 in *Drosophila*, Kami Ahmad and Steve Henikoff demonstrated in 2002 that a pulse of H3.3 and not H3.2 is deposited at transcriptionally active ribosomal DNA arrays in a replication-independent manner. In addition, they showed that three of the four amino acid changes that distinguish H3.3 from H3.2 (A87, I89, G90 in H3.3, see **Figure 1.16A**) allow H3.3-GFP to engage in replication-independent as well as replication-dependent assembly into chromatin (Ahmad and Henikoff, 2002). GFP-tagged H3.3 was also shown by David Spector's group to be deposited upon transcriptional induction of a transgene array in real-time analysis of single mammalian cells (Janicki et al., 2004). In complementary studies, Schwartz and Ahmad demonstrated that H3.3 deposition is temporally coupled to transcription on *Drosophila* heat-shock genes (Schwartz and Ahmad, 2005). Strikingly, they found H3.3 to be deposited and rapidly turned over at continually active genes,

while incorporated H3.3 became stable at genes after transcription has ceased, leaving behind an enduring marker of past transcriptional activity (Schwartz and Ahmad, 2005, 2006). Mass spectrometry studies from our laboratory and others have also revealed that specific post-translational histone modifications on histone H3 associated with gene activation (H3K4 methylation, H3K36 methylation) are significantly enriched on H3.3 in both *Drosophila* and mammalian cells (Hake et al., 2006; Loyola et al., 2006; McKittrick et al., 2004).

Genome-scale profiling of tagged H3.3 provides strong evidence of H3.3 enrichment at transcriptionally active genes and cis-regulatory elements in *Drosophila* S2 cells (Mito et al., 2005, 2007). Using chromatin immunoprecipitation (ChIP) and tiling microarrays, Henikoff and colleagues found that H3.3 replacement was significantly co-localized with sites of RNAPII and H3K4me2. They observed that H3.3 was particularly prominent in the gene bodies of highly active genes, which they suggest represent sites of ongoing transcription-coupled nucleosome replacement (Mito et al., 2005). In 2007, Henikoff and colleagues found that H3.3 was also enriched in *Drosophila* at cis-regulatory domain boundaries and sites of nuclease hypersensitivity, even outside of active or repressed genes (Mito et al., 2007). This enrichment of H3.3 at *Drosophila* regulatory elements represented an important demonstration that that specifically targeted H3.3 deposition was not limited to sites of transcriptional activation, and showed that regulatory elements are also sites of dynamic histone replacement (Mito et al., 2007).

Although genome-scale data regarding H3.3 localization in *Drosophila* has been consistent, more limited experiments suggested alternative and sometimes conflicting possibilities in vertebrates (Chow et al., 2005; Daury et al., 2006; Jin and Felsenfeld, 2006; Sutcliffe et al., 2009; Tamura et al., 2009; van der Heijden et al., 2007; Wirbelauer et al., 2005). One study found that H3.3 incorporation was confined primarily to active promoters, while another reported that H3.3 is enriched

throughout transcribed regions (Chow et al., 2005; Wirbelauer et al., 2005). Felsenfeld and colleagues found no clear correlation between H3.3 incorporation and levels of active transcription or histone modifications in chicken erythroid cells, with enrichment of H3.3 found in upstream regulatory regions, coding regions, or neither (Jin and Felsenfeld, 2006). H3.3 was shown to be constitutively unenriched around the transcription start sites (TSSs) of two inducible genes in T-cells, but enriched in the coding region upon gene induction (Sutcliffe et al., 2009). A recent genome-wide study of H3.3 in HeLa cells found that H3.3 was enriched at the TSSs of active genes, and also within the body of active genes in proportion to transcriptional activity, but significant levels of H3.3 were not detected at the TSSs of inactive genes in this differentiated human cancer cell line (Jin et al., 2009). However, H3.3 was also recently shown to be constitutively enriched in the promoter region of repressed interferon-inducible genes in fibroblasts, with levels of enrichment increasing in the coding regions upon gene induction (Tamura et al., 2009), and the precise pattern of H3.3 at active and inactive genes in vertebrates therefore has yet to be clearly established.

Roles for H3.3 in large-scale nucleosome replacement

H3.3 has not been exclusively linked with transcriptional regulation at specific loci, but has also been implicated in a variety of developmental processes and larger-scale chromatin remodeling events that are likely to involve nucleosome replacement. H3.3 appears to play particularly important roles in multiple phenomena related to reproduction, including meiotic sex chromosome inactivation, sperm chromatin remodeling, and paternal pronuclear decondensation (Orsi et al., 2009; Santenard and Torres-Padilla, 2009).

H3.3 has been associated with chromosome-wide formation of heterochromatin at the X and Y chromosomes during mouse male meiosis (van

der Heijden et al., 2007). During male meiotic prophase I, mammalian sex chromosomes form a separate repressive chromatin domain known as the XY body (Handel, 2004). The XY body is transcriptionally silenced, associated with a variety of heterochromatic proteins and histone modifications, and restricted from homologous recombination (Handel, 2004). Some of the heterochromatic proteins and modifications enriched in the XY body include the histone variants macroH2A and gammaH2A.X, H3K9 methylation, CBX1 (HP1beta), CBX3 (HP1gamma), as well as the small ubiquitin modifier protein SUMO and the SUMO substrates PML and Daxx (Handel, 2004; Metzler-Guillemain et al., 2003; Rogers et al., 2004).

Patterns of H3.3 gene expression coincide with the reported incorporation of H3.3 protein at the XY body during meiotic sex chromosome inactivation (Bramlage et al., 1997; van der Heijden et al., 2007). During mammalian spermatogenesis, H3.3A mRNA can be detected throughout pre- and post-meiotic cells, while expression of H3.3B appears to be more restricted but highly expressed during meiotic prophase (Bramlage et al., 1997). Using an antibody specific to H3.1/H3.2, van der Heijden and colleagues demonstrated that H3.1 and H3.2 are gradually removed from the XY body during male meiotic prophase I (van der Heijden et al., 2007). Moreover, using transgenic mice with tagged H3.3, they demonstrated that H3.3 is simultaneously incorporated throughout the transcriptionally silenced X and Y chromosomes during pachytene stages of spermatogenesis, coincident with loss of most histone post-translational modifications and meiotic sex chromosome inactivation (van der Heijden et al., 2007). This report marks the first demonstration that H3.3 can also be associated with the large scale formation of repressive heterochromatin.

H3.3 also plays a role in paternal pronuclear decondensation in a variety of species (Orsi et al., 2009). The early stages of post-fertilization mammalian development provide a remarkable example of genome-wide epigenetic

reprogramming, as gametic epigenetic marks are erased and replaced with embryonic marks important for early embryonic development and toti/pluripotency (Feil, 2009; Morgan et al., 2005). Soon after fertilization in mammals, paternal genomic DNA is actively demethylated, the sperm pronucleus is decondensed, and its protamines replaced by histones (Morgan et al., 2005). In *Drosophila*, nucleosomes containing H3.3 and not H3.2 were shown to be specifically assembled in the sperm pronucleus following fertilization, and the H3.3 chaperone HIRA and the chromatin remodeling protein CHD1 are also required for male pronuclear H3.3 deposition (Bonney et al., 2007; Konev et al., 2007; Loppin et al., 2005). There is some evidence that H3.3 plays a similar role in mammals. The male pronucleus in the pre-S-phase mouse zygote has been shown to be virtually devoid of H3.1/H3.2, while H3.1/H3.2 is enriched in the female pronucleus (van der Heijden et al., 2005). In addition, an antibody ostensibly specific to histone H3.3 has also been used to demonstrate that H3.3 is present in the mouse oocyte as a maternal factor, and incorporated preferentially into the male pronucleus prior to genome activation, although this antibody is no longer available (Torres-Padilla et al., 2006). Intriguingly, although most DNA in human sperm is packaged in protamines, approximately 15% remains associated with histones, and both H3.3 and H3.1 have been shown to be present in core histones isolated from human sperm (Gatewood et al., 1990). Therefore, it is also possible that incorporation of H3.3 might play a role in the recently demonstrated epigenetic marking that packages key developmental loci in human sperm (Hammoud et al., 2009). However, despite the association between H3.3 and various developmental processes, the precise role of endogenous H3.3 in mammalian reproduction remains difficult to study due to the lack of available antibody or genetic tools.

H3.3 serine 31 phosphorylation is a mitosis specific mark

As mentioned above, the only residue unique to H3.3 that is accessible in the context of the nucleosome is H3.3 serine 31 (H3.3S31) (**Figure 1.16**). H3.3S31 was recently established as a target for phosphorylation (H3.3S31p), and the genomic localization of this phosphorylation appears to be developmentally regulated (Hake et al., 2005; Wong et al., 2009). In human cervical cancer cells (HeLa) and immortalized human embryonic kidney cells (HEK293), H3.3S31p is specific to mitosis, and a timecourse of immunofluorescence (IF) studies revealed that H3.3S31p is restricted to late prometaphase and metaphase (Hake et al., 2005). Chromosome spreads in HeLa and HEK293 show that H3.3S31p is localized to regions surrounding centromeres on metaphase chromosomes (Hake et al., 2005). Unlike in HeLa and HEK293 cells, H3.3S31p is localized to telomeres in mouse ES cells (Wong et al., 2009) (**Figure 1.19**).

Telomeres and telomeric chromatin

Telomeres are specialized nucleoprotein structures that protect the ends of eukaryotic chromosomes from DNA degradation and inappropriate DNA repair (Palm and de Lange, 2008). Telomeres consist of tandem arrays of short DNA repeats, and these sequences are maintained by the reverse transcriptase enzyme telomerase (Blackburn et al., 2006; Luke and Lingner, 2009; Meyne et al., 1989; Palm and de Lange, 2008). Telomere length and telomerase expression play important roles in cellular lifespan and replicative capacity, and telomerase protects cells from the cellular senescence and genomic instability that would otherwise result from telomere shortening after excessive cell division (Allsopp et al., 1992; Bodnar et al., 1998; Harley et al., 1990; Levy et al., 1992; Luke and Lingner, 2009). The ends of linear eukaryotic chromosomes have the potential to be recognized as sites of DNA damage, and a multi-protein complex known as shelterin therefore protects

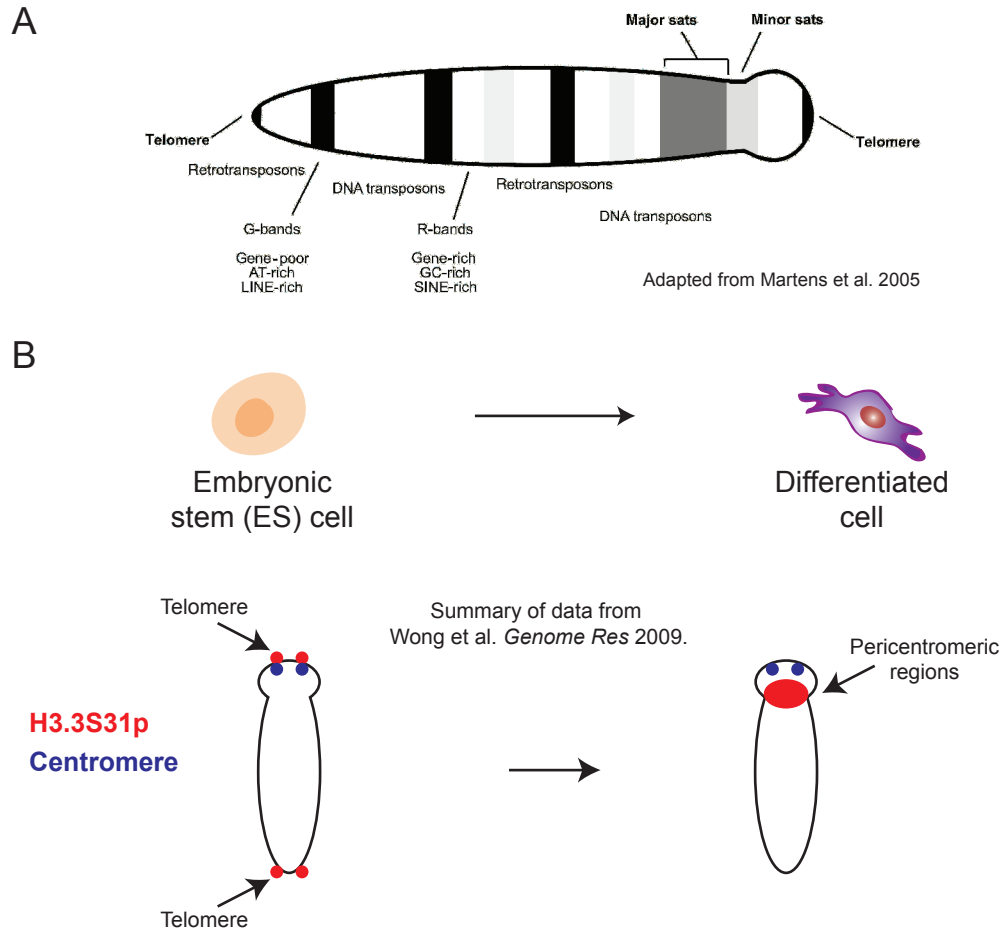


Figure 1.19: Developmentally regulated localization of H3.3S31 phosphorylation

A. Organization of mitotic mouse chromosomes. Unlike metacentric human chromosomes, with their centromeres in the center of the chromosome, mouse chromosomes are acrocentric, with their centromeres located extremely near one end of the chromosome. The distribution of major (pericentric) and minor (centromeric) satellite repeats and of the various interspersed repetitive elements are indicated.

B. The chromosomal localization of H3.3S31p is developmentally regulated (Wong et al., 2009). In undifferentiated ES cells, H3.3S31p is localized to telomeres. However, upon retinoic acid differentiation, H3.3S31p is relocalized to pericentromeric regions.

telomeres from recognition and repair by the non-homologous end joining and homology directed DNA repair pathways (Palm and de Lange, 2008). Telomeres are bordered by subtelomeric regions that are enriched in repetitive DNA and contain a low density of genes (Blasco, 2007).

Like the rest of the eukaryotic genome, vertebrate telomeres are not made up simply of naked DNA, but are packaged into nucleosomes containing core histones

(Lejnine et al., 1995; Makarov et al., 1993; Tommerup et al., 1994). However, telomeric chromatin appears to have unique features, including a short nucleosomal repeat length, hypersensitivity to MNase, and higher intrinsic nucleosome mobility (Pisano et al., 2007; Tommerup et al., 1994). Moreover, telomeric chromatin has features that are consistent with repressive heterochromatin, including H3K9me3, H4K20me3, hypoacetylation of H3 and H4, and DNA methylation of subtelomeric

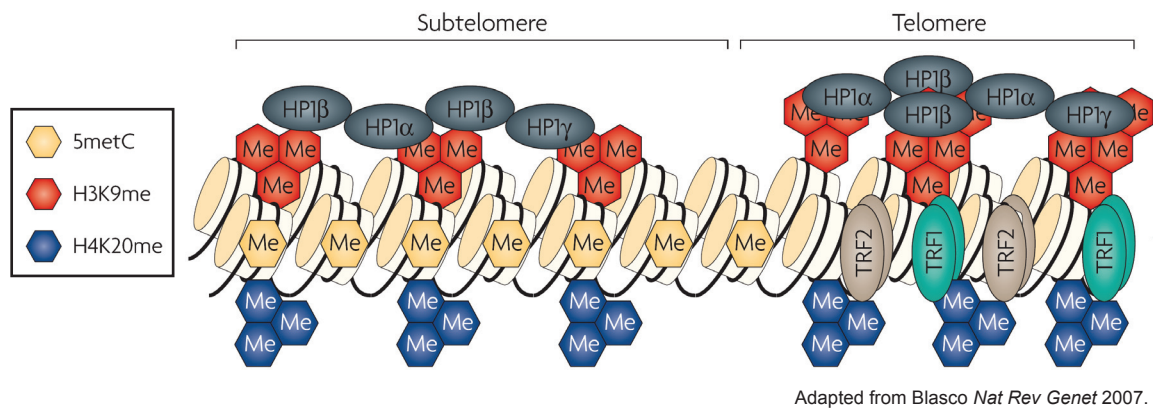


Figure 1.20: Telomeric chromatin has characteristics of heterochromatin

Telomeric and subtelomeric chromatin is heavily enriched in H3K9me3, H4K20me3, and HP1. Subtelomeric DNA is heavily methylated by the DNA methyltransferases DNMT1, DNMT3A, and DNMT3B. DNA methylation has been shown to inhibit telomere recombination. H3K9me3 and DNA methylation act independently as negative regulators of telomere length. For review, see: (Blasco, 2007).

regions (Blasco, 2007) (**Figure 1.20**). In other organisms such as yeast and *Drosophila*, telomeric heterochromatin has been shown to silence neighboring genes in a phenomenon known as telomeric position effect (TPE), though clear evidence for TPE in mammalian cells is still lacking (Ottaviani et al., 2008). Although telomeres were long thought to be transcriptionally silent, recent studies demonstrate that telomeres are transcribed into telomeric repeat-containing RNA (TERRA), a large non-coding RNA that forms an integral component of telomeric heterochromatin (Azzalin et al., 2007; Luke and Lingner, 2009; Luke et al., 2008; Schoeftner and Blasco, 2008, 2009). The function of TERRA remains unclear, but

it has been implicated in telomerase regulation, and has been suggested to play a role in the stabilization and maintenance of telomeric heterochromatin state (Luke and Lingner, 2009; Schoeftner and Blasco, 2009). Recent studies also indicate that telomere chromatin state is developmentally regulated (Marion et al., 2009; Schoeftner and Blasco, 2008). ES and iPS cell telomeres have unique features, including telomerase activation during reprogramming, increased transcription of telomeric RNA, and reduced levels of heterochromatic histone modifications (Marion et al., 2009; Schoeftner and Blasco, 2008).

Developmentally regulated telomeric and pericentromeric localization of H3.3 and H3.3 serine 31 phosphorylation

A study published in the last year suggests that H3.3 and H3.3S31p may play a role in ES cell telomeric chromatin (Wong et al., 2009). Although H3.3S31p is localized to telomeres of metaphase chromosomes in ES cells, upon ES cell differentiation with retinoic acid H3.3S31p is re-localized to pericentromeric DNA of metaphase chromosomes, similar to its localization in differentiated human cancer cells (Hake et al., 2005; Wong et al., 2009) (**Figure 1.19**). H3.3S31p was also found at the telomeres of mouse embryonic germ cells, further suggesting that telomeric localization of H3.3S31p may be related to pluripotency (Wong et al., 2009). Transfection of myc-tagged H3.3 and H3.2 revealed that H3.3 itself was also specifically associated with telomeres in both metaphase chromosomes and interphase nuclei, as demonstrated by colocalization with the telomeric protein TERF1 (Wong et al., 2009). The telomeric localization of H3.3 increases during early to late S/G2, suggesting that H3.3 telomeric deposition is coupled to telomeric DNA replication, although the timing of H3.3 telomeric deposition has not been conclusively established (Wong et al., 2009). Knockdown of H3.3A and H3.3B in ES cells induced a slight increase in telomere dysfunctional foci

(TIFs), as assessed by the presence of DNA damage factors on telomeres such as phosphorylation of gamma H2A.X (Wong et al., 2009). The authors of this study also found that the putative H3.3 chaperone HIRA partially co-localized to telomeres, and based on this co-localization they suggest that HIRA may mediate H3.3 telomeric deposition (Wong et al., 2009). Notably, the precise function of H3.3 at telomeres remains unclear. Moreover, despite the co-localization of HIRA with telomeres, the evidence described above also suggests that H3.3 is deposited at telomeres during S-phase, while HIRA-mediated nucleosome assembly has been demonstrated to occur independent of replication (Ray-Gallet et al., 2002; Tagami et al., 2004; Wong et al., 2009). The authors do not remove HIRA and assess telomeric incorporation, and therefore it is not clear whether HIRA is required for H3.3 telomeric deposition in ES cells (Wong et al., 2009).

Functional requirements for H3.3 in transcriptional regulation and development

Levels of H3 variants have been shown to be important for regulating gene expression in somatic cells and for organismal fertility (Couldrey et al., 1999; Hodl and Basler, 2009; Sakai et al., 2009; Tamura et al., 2009). The expression of specific genes in chicken erythroid cells was altered by expression of ectopic H3.2 and H3.3, suggesting an active role for H3 variants, and raising concerns regarding other experiments that use overexpressed tagged proteins to track H3 variants (Jin and Felsenfeld, 2006). In a mouse fibroblast cell line, knockdown of H3.3 was also shown to inhibit the expression of interferon induced genes, again implicating H3.3 in gene activation (Tamura et al., 2009). H3.3 may also play a role in epigenetic memory (Ng and Gurdon, 2008), although later experiments demonstrate that H3.3 is not absolutely required for development (Hodl and Basler, 2009; Sakai et al., 2009). Following nuclear transfer experiments, a number of

embryos sometimes 'remember' their tissue of origin by inappropriately expressing genes characteristic of their pre-transplantation lineage (Ng and Gurdon, 2005). Intriguingly, overexpression of H3.3 has been shown to enhance this expression of pre-transplantation lineage-specific genes following nuclear transplantation (Ng and Gurdon, 2008). Although the role of H3.2 in this process was not directly compared with H3.3, these experiments suggest that H3.3 levels (or at least H3 variant levels) may play a role in modulating epigenetic memory (Ng and Gurdon, 2008).

H3.3 in neurobiology

Chromatin regulation plays important roles in neuronal physiology and development (Borrelli et al., 2008). Histone modifications and chromatin remodeling have been linked with synaptic plasticity as well as with behavioral memory (Borrelli et al., 2008; Fagiolini et al., 2009; Levenson and Sweatt, 2005). Moreover, multiple neurological disorders involve mutations in genes that encode chromatin-binding or chromatin-modifying enzymes. Intriguing examples include mutations in a chromatin remodeling protein that causes alpha-thalassaemia and X-linked mental retardation (ATRX) syndrome, in a histone H3K4me3 demethylase that causes epilepsy and X-linked mental retardation, in a DNA methylation binding protein that causes Rett syndrome, and in a HAT that causes Rubinstein-Taybi syndrome (Chahrour and Zoghbi, 2007; Gibbons et al., 2008; Murata et al., 2001; Tahiliani et al., 2007). A common theme of these disorders is that mutations in epigenetic regulators can alter chromatin structure and induce a broad spectrum of neurodevelopmental defects.

Multiple lines of evidence suggest that H3.3 is likely to play an important role in neuronal biology. Most neurons in the mammalian central nervous system are post-mitotic, and H3.3 is the only H3 variant expressed at significant levels

outside of S phase (Wu et al., 1982). H3.3 RNA and protein levels were observed to increase during models of cell differentiation in non-neuronal systems (Krimer et al., 1993; Lord et al., 1990; Wunsch and Lough, 1987). High levels of H3.3 protein have been found in differentiating rat neurons, and H3.3 has been shown to become the dominant histone H3 variant protein in rat brain cortical neurons (Pina and Suau, 1987; Scaturro et al., 1995). H3.3A and H3.3B mRNA are highly expressed in distinctive, yet partially overlapping patterns of expression in the mouse brain (Lein et al., 2007). Indeed, RNA in situ hybridization studies indicate that H3.3A expression is relatively ubiquitous throughout the mouse brain, while H3.3B is mostly expressed in the cerebellum and the hippocampus, particularly in the dentate gyrus (Lein et al., 2007) (**Figure 1.21**). Gene trap mice with reduced expression of H3.3A also showed neuromuscular defects, with impaired performance on tasks involving grip strength and muscle tone, further suggesting a function for H3.3 in neuronal biology (Couldrey et al., 1999). However, due to the hypomorphic nature of available studies and the lack of an available H3.3 antibody, the specific role of H3.3 in mammalian neuronal biology has yet to be elucidated.

Requirements for H3.3 in sexual reproduction

H3.3 plays a critical role in sexual reproduction in multiple organisms. Given the association of H3.3 with specific processes in the reproduction of a variety of species, as described above (Cui et al., 2006; Loppin et al., 2005; Ooi et al., 2006; Torres-Padilla et al., 2006), it is not surprising that recent studies have shown that H3.3 is required for fertility in flies, ciliates, and mammals. Although the functional requirements for H3.3 in transcriptional regulation are not absolute, the requirements for H3.3 in fertility are clear (Couldrey et al., 1999; Hodl and Basler, 2009; Sakai et al., 2009).

In 1999, Couldrey and colleagues reported a retroviral gene trap insertion

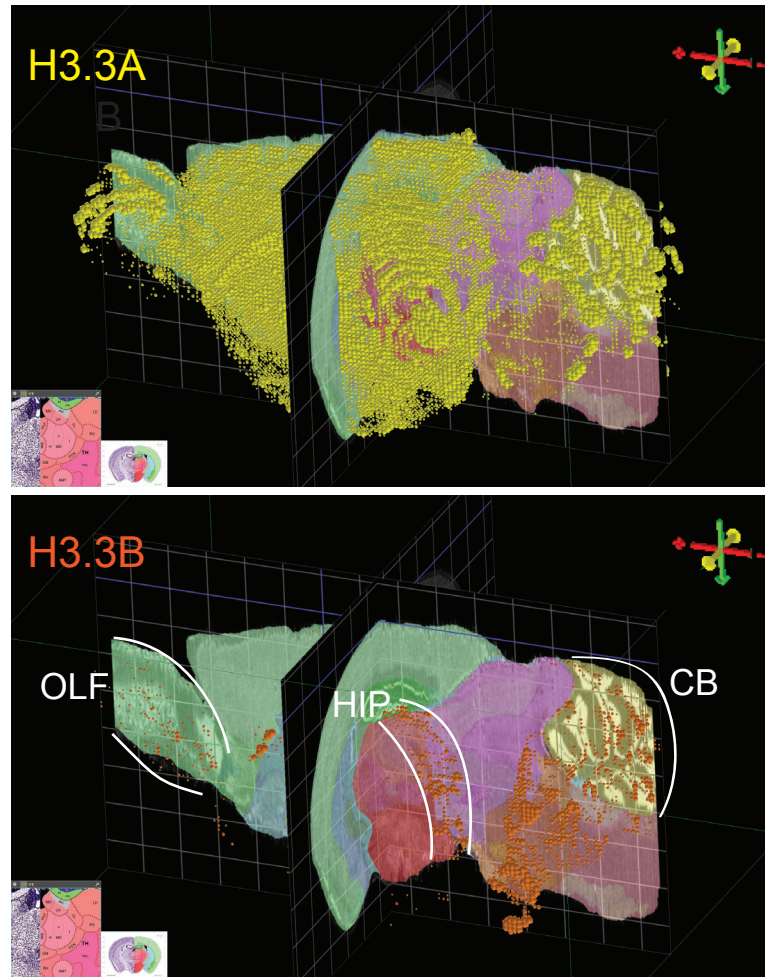
Figure 1.21: Patterns of H3.3A and H3.3B expression in the adult mouse brain

A. In situ hybridization of H3.3A and H3.3B RNA in the adult mouse brain reveals distinct, but partially overlapping patterns of expression. Three-dimensional oblique views of the mouse brain are shown, with structures depicted on the sagittal and coronal planes. H3.3A RNA expression is shown in yellow (top), and H3.3B RNA expression is shown in orange (bottom). While H3.3A expression is relatively ubiquitous, H3.3B expression is limited to the cerebellum (CB), hippocampus (HIP), and olfactory bulb (OLF). Data was downloaded from the Allen Brain Atlas and represented within the Allen Brain Atlas Brain Explorer.

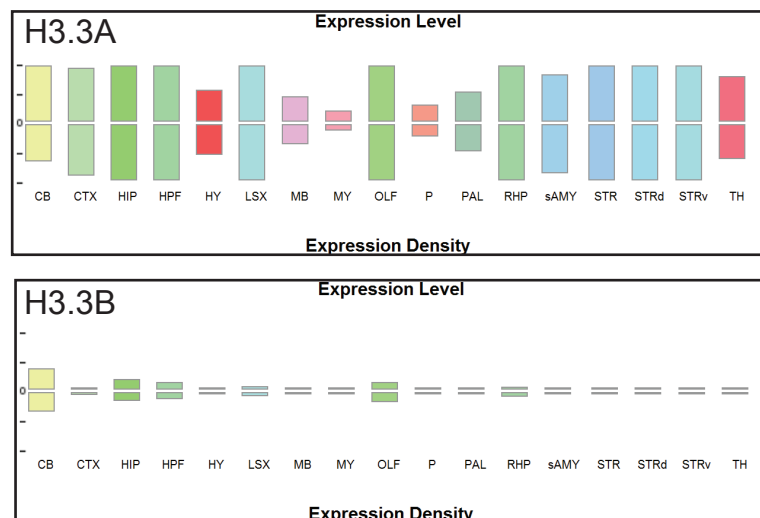
B. Relative quantification of H3.3A and H3.3B expression level and expression density in specific brain structures, as reported by the Allen Brain Atlas (Lein et al., 2007).

CB	Cerebellum
CTX	Cerebral cortex
HIP	Hippocampal region
HPF	Hippocampal formation
HY	Hypothalamus
LSX	Lateral septal complex
MB	Midbrain
MY	Medulla
OLF	Olfactory bulb
P	Pons
PAL	Pallidum
RHP	Retrohippocampal region
sAMY	Striatum-like amygdalar nuclei
STR	Striatum
STRd	Striatum dorsal region
STRv	Striatum ventral region
TH	Thalamus

A



B



Expression data downloaded from: Allen Mouse Brain Atlas [Internet]. Seattle (WA): Allen Institute for Brain Science. ©2009. Available from: <http://mouse.brain-map.org>.

Figure 1.21

into the first intron of murine H3.3A that created a hypomorphic mutation, resulting in partial neonatal lethality, stunted growth, neuromuscular deficits, and male sub-fertility (Couldrey et al., 1999). H3.3A gene trap hypomorphic mice had significantly fewer matings, and of these matings, only 6% of matings (1/16) of H3.3A hypomorphic mice resulted in pregnancy, while 74% (48/65) of matings in wild-type mice resulted in pregnancy (Couldrey et al., 1999). However, hypomorphic H3.3A homozygotes still retained expression of H3.3A detectable by RT-PCR (Couldrey et al., 1999), and the phenotype is also likely to have been rescued by the presence of H3.3B, which codes for an identical protein. Therefore, the precise function of H3.3 in mammalian development and physiology has yet to be established. More recently, full knockouts of both copies of H3.3 have been reported in *Drosophila* and *Tetrahymena*, and they show an absolute requirement of H3.3 for sexual reproduction (Cui et al., 2006; Hodl and Basler, 2009; Sakai et al., 2009).

In both *Drosophila* and *Tetrahymena*, replication-independent H3.3 or H3.3-like genes are not essential for growth, but are required for sexual reproduction (Cui et al., 2006; Hodl and Basler, 2009; Sakai et al., 2009). In *Tetrahymena*, cells can grow with expression of either replication-coupled or replication-independent H3 variants, but they show an absolute requirement for the minor H3 variants (H3.3 and H3.4) in the production of viable sexual conjugation progeny (Cui et al., 2006). Two groups recently reported that H3.3 was also absolutely required for fertility, but only partially required for developmental viability of *Drosophila* (Hodl and Basler, 2009; Sakai et al., 2009). In *Drosophila*, loss of both copies of H3.3 leads to partial but incomplete lethality (~42% viability), complete sterility, and transcriptional defects, particularly at highly expressed genes (Sakai et al., 2009). However, the survival of even less than half of H3.3 double null flies indicates that H3.3 is not absolutely required for transcription (Hodl and Basler, 2009; Sakai et al., 2009). Loss of both copies of H3.3 appears to be partially compensated by an increase

in the expression of the normally replication-dependent histone genes, and these upregulated histone gene transcripts were found to be polyadenylated (Sakai et al., 2009). In double null H3.3 flies, expression of the H3.3 rescue transgene (but not the H3.2 rescue transgene) suppresses overexpression of the endogenous H3 genes, suggesting that *Drosophila* cells can distinguish H3.3 expression from bulk H3 (Sakai et al., 2009). These experiments also showed that maintaining wild-type levels of H3K4 methylation is likely still important for transcriptional regulation, but is still partially dispensable for normal fly development (Hodl and Basler, 2009), as ~20% of H3.3K4R mutant flies remain viable (Sakai et al., 2009).

The absolute requirement of H3.3 for fertility appears to be due to its role in chromosome segregation during spermiogenesis, and remodeling of germline chromatin (Sakai et al., 2009). H3.3 double null flies fail to properly condense and segregate chromosomes during male meiosis, and they also show an abnormal arrangement of chromatin in postmeiotic nuclei, indicating that H3.3 is critical for remodeling of germline chromatin (Sakai et al., 2009). Intriguingly, in H3.3 double null flies, H3.3S31A, H3.3K4R, and H3.3K9R rescue fly fertility, while expression of H3.2 does not. Therefore, although known H3.3 modification sites are not required for fertility, the H3.3 specific residues at 87, 89, and 90 (**Figure 1.16**) are absolutely required, suggesting that these amino acids play an essential role in the remodeling and packaging of germline chromatin (Sakai et al., 2009).

Histone H3 deposition machinery

Progress has also been made on the molecular machinery that deposit different histones into chromatin (De Koning et al., 2007; Loyola and Almouzni, 2007). Nucleosome deposition or nucleosome replacement occurs in a variety of contexts, including DNA replication, DNA repair, and transcription, and elucidating the mechanisms responsible for recruitment of histone deposition factors to these

distinct genomic regions remains an intense area of research (De Koning et al., 2007; Groth et al., 2007b; Polo and Almouzni, 2006).

Histone H3.3 has been shown to be deposited into *Drosophila* chromatin in both a replication-coupled (RC) and replication-independent (RI) manner, while H3.2 deposition is coupled to replication (Ahmad and Henikoff, 2002). Biochemical purification of tagged-H3.1 and H3.3 from human HeLa cells by Nakatani, Almouzni and colleagues yielded complexes with distinct histone chaperones (Tagami et al., 2004). The chromatin assembly factor complex 1 (CAF-1) was associated with H3.1, while histone regulatory protein A (HIRA) was associated with H3.3 (Tagami et al., 2004). Intriguingly, CAF-1 has been shown to mediate DNA synthesis-dependent nucleosome assembly, while HIRA mediates DNA synthesis-independent nucleosome assembly (Ray-Gallet et al., 2002; Tagami et al., 2004).

The H3.1 associated CAF-1 is composed of three subunits, p150, p60, and p48, and this complex has been shown to promote chromatin assembly in the context of DNA replication or DNA repair (Polo et al., 2006; Smith and Stillman, 1989; Verreault et al., 1996). Biochemical studies have established that CAF-1 is required for DNA synthesis-dependent nucleosome assembly in *Xenopus* egg extracts, as well as for H3.1 deposition at sites of DNA damage (Polo and Almouzni, 2006; Polo et al., 2006; Tagami et al., 2004). CAF-1 colocalizes with the proliferating cell nuclear antigen (PCNA) on newly replicating DNA, and is also required for heterochromatin organization in pluripotent ES cells and for early embryonic development (Shibahara and Stillman, 1999; Verreault et al., 1996).

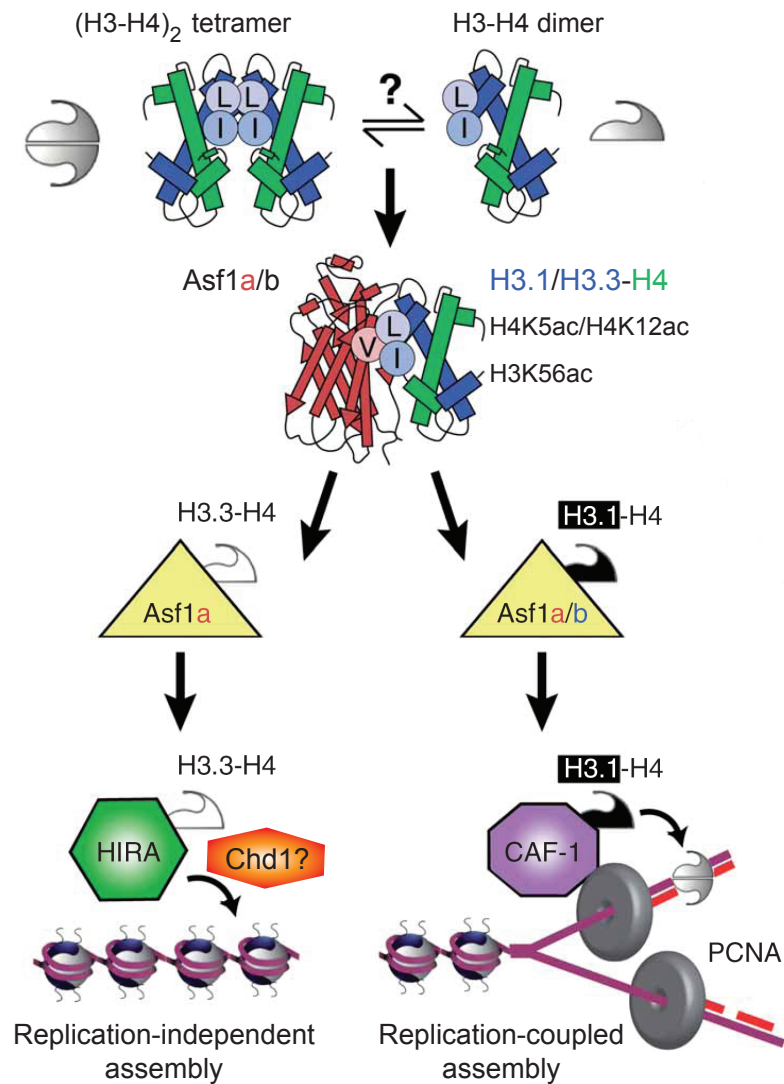
In addition to CAF-1 and HIRA, which appear to be specific in their preferences for H3.1 or H3.3, Nakatani, Almouzni, and colleagues identified a variety of proteins were associated with both H3.1 and H3.3 (Tagami et al., 2004). Such common partners included the histone H3-H4 chaperones anti-silencing function 1 (ASF1) A and B, the H1 binding protein NASP, the histone acetyltransferase HAT1, and the

common CAF-1 subunit p48, also known as retinoblastoma binding protein p48 (Tagami et al., 2004).

Deposition of H3-H4 as a heterodimer

H3 and H4 form a $(\text{H3-H4})_2$ tetramer in the context of the nucleosome, and this tetramer is stable even in the absence of DNA, leading to the longstanding suggestion that H3 and H4 are deposited into chromatin as a tetramer (Baxeavanis et al., 1991; De Koning et al., 2007; Luger et al., 1997). However, H3-H4 bound to HIRA or CAF-1 was recently shown to exist as a H3-H4 heterodimer rather than a tetramer (Tagami et al., 2004). Moreover, recent biochemical and crystallographic studies have demonstrated that both yeast and human ASF1 bind to and stabilize H3-H4 heterodimers, physically blocking assembly of $(\text{H3-H4})_2$ tetramers (Bao and Shen, 2006; English et al., 2006; Natsume et al., 2007) (**Figure 1.22**). Soluble H3-H4 heterodimers have also been shown to be heavily acetylated at H4K5 and H4K12 (Groth et al., 2007b; Loyola et al., 2006). In vertebrate cells, H4K5ac/H4K12ac is largely mediated by the two subunit enzyme HAT1-RbAp46, and these acetyl marks have been implicated in histone deposition in a variety of organisms (Barman et al., 2006; Sobel et al., 1994; Sobel et al., 1995; Verreault et al., 1996, 1998). HAT1 is surprisingly dispensable for replication-coupled chromatin assembly in vertebrate cells, though HAT1^{-/-} cells are sensitive to DNA damage, and the precise function of H4K5/K12 diacetylation remains unclear (Barman et al., 2006; Groth et al., 2007b).

Current models suggest that ASF1 serves as a donor for deposition of H3.1-H4 or H3.3-H4 heterodimers by CAF-1 or HIRA (De Koning et al., 2007) (**Figure 1.22**). CAF-1 and HIRA both bind to ASF1A, and ASF1A and ASF1B have recently both been shown to play a role in both RI and RC nucleosome assembly in human cells (Galvani et al., 2008; Tang et al., 2006). Interestingly, although



Adapted from De Koning 2007.

Figure 1.22: Current view of H3 variant deposition pathways

The current view of histone H3 variants in chromatin assembly pathways is depicted here. Histone H3 and H4 can exist in vitro as tetramers or dimers, and it remains unclear (question mark) whether the H3-H4 dimer utilized in histone deposition comes from a split tetramer (upper left) or a dimeric source (upper right). Asf1 (red) interacts with and stabilizes the H3-H4 dimer. In replication-coupled chromatin assembly (right), Asf1a and Asf1a act as histone donors for CAF-1. In replication-independent chromatin assembly (left), Asf1a and Hira act together to deposit H3.3 during transcription or specific developmental contexts. Chd1 may also cooperate with Hira to deposit H3.3 into chromatin (Konev et al., 2007). Notably, in current models, Hira is required for H3.3-specific deposition into chromatin (De Koning et al., 2007).

HIRA is reported to physically interact with ASF1A but not ASF1B, depletion of both ASF1A and ASF1B were required to inhibit G1-phase deposition of H3.3 into chromatin, indicating that H3.3 deposition during G1 could still occur in the absence of a HIRA-ASF1A interaction (Galvani et al., 2008; Zhang et al., 2005).

ASF1 can participate in both nucleosome assembly and disassembly (Groth et al., 2007b). In human cells, ASF1 interacts with the putative replicative helicase (MCM1-7) through a H3-H4 bridge, and ASF1A is required for vertebrate cellular viability and chromatin assembly during replication (Groth et al., 2007a; Sanematsu et al., 2006). In yeast and mammals, ASF1 is required for H3K56ac, a modification that has been associated with sites of newly deposited histones in replication-coupled, replication-independent, and DNA damage deposition pathways (Chen et al., 2008a; Fillingham and Greenblatt, 2008; Li et al., 2008; Recht et al., 2006; Yuan et al., 2009). Because of its common role in multiple H3 deposition pathways, H3K56ac has been suggested as a universal signature of newly assembled yeast chromatin (Fillingham and Greenblatt, 2008). Interestingly, despite its association with chromatin assembly, H3K56ac has also been shown to be required for normal promoter activation and chromatin disassembly in yeast (Williams et al., 2008). Genome-scale profiles of H3K56ac have recently been established in human ES cells, and this modification localizes to both active and inactive promoters, significantly overlapping with binding sites for key regulators of pluripotency (Xie et al., 2009). H3K56ac has been shown to increase the affinity of H3 for CAF-1 as well as the histone chaperone Rtt106 in yeast, although it remains unclear if H3K56ac can play a similar role in increasing affinity of H3.3 for HIRA in mammalian cells (Chen et al., 2008a; Fillingham and Greenblatt, 2008; Li et al., 2008).

HIRA – histone regulatory protein A

The H3.3-associated protein HIRA is required for replication-independent nucleosome assembly *in vitro*, and for *Drosophila* male pronuclear decondensation *in vivo* (Bonnefoy et al., 2007; Loppin et al., 2005; Ray-Gallet et al., 2002; Tagami et al., 2004). Although humans have a single HIRA protein, both budding and fission yeast have two homologs of HIRA that function together (Blackwell et al., 2004; Kanoh and Russell, 2000; Lamour et al., 1995; Spector et al., 1997). The yeast homologs of HIRA (Hir1 and Hir2 in *S. cerevisiae*) were initially identified as factors that repress the transcription of histone genes outside of S phase, and subsequent work has shown that these proteins are also required to maintain pericentromeric heterochromatin in the fission yeast *S. pombe* (Blackwell et al., 2004; Lamour et al., 1995; Spector et al., 1997). Consistent with its role in the promotion of repressive yeast chromatin, nucleosomes bound by the yeast HIR complex are resistant to chromatin remodeling by SWI/SNF (Prochasson et al., 2005). Intriguingly, a recent study demonstrated that the *S. pombe* HIRA complex functions as a global transcriptional repressor, and is particularly important for promoter silencing and the suppression of cryptic antisense transcripts (Anderson et al., 2009). HIRA has also been shown to play a role in human heterochromatin, as HIRA and ASF1A drive chromosome condensation during the formation of specialized heterochromatin domains known as senescence associated heterochromatin foci (SAHF) (Adams, 2007; Ye et al., 2007; Zhang et al., 2007; Zhang et al., 2005). However, despite this conserved functional association with heterochromatin and gene repression, the human homolog of HIRA was specifically isolated with H3.3 from HeLa cells, and as we have seen, H3.3 is largely associated with active chromatin (Ahmad and Henikoff, 2002; Tagami et al., 2004).

Biochemical experiments clearly link HIRA to replication-independent nucleosome assembly (Ray-Gallet et al., 2002; Tagami et al., 2004). HIRA is a

large WD repeat containing protein with seven WD repeats conserved between yeast and human HIR proteins, and both the mammalian and *Xenopus* homologs of HIRA interact with core histones (Lorain et al., 1996; Lorain et al., 1998; Magnaghi et al., 1998). Following this observation, the role of HIRA in nucleosome assembly has largely been studied in *Xenopus* egg extracts, which provides a powerful system for studying chromatin assembly (Ray-Gallet et al., 2002; Tagami et al., 2004). Proteins contained within egg extracts will assemble chromatin on plasmid DNA (Almouzni and Mechali, 1988; Gaillard et al., 1996; Gaillard et al., 1999). If the plasmid has been previously irradiated with UV, the *Xenopus* egg extract will perform chromatin assembly in the presence of DNA synthesis during the process of nucleotide excision repair (Gaillard et al., 1999). In *Xenopus* egg extracts, immunodepletion of HIRA did not effect nucleosome assembly of UV-irradiated plasmids coupled to nucleotide excision repair, but loss of HIRA severely impaired DNA synthesis independent nucleosome assembly (Ray-Gallet et al., 2002; Tagami et al., 2004). This defect in replication-independent nucleosome assembly could be rescued by addition of recombinant HIRA protein and (H3-H4)₂ tetramers, or HIRA enriched protein fractions from egg extract (Ray-Gallet et al., 2002). More interestingly, RI nucleosome assembly could also be rescued by purified H3.3 complexes from HeLa cells that contained HIRA, but not by purified H3.1 complexes that contained CAF-1, or by recombinant HIRA or CAF-1 (Ray-Gallet et al., 2002; Tagami et al., 2004). Conversely, CAF-1 was required for DNA synthesis-dependent nucleosome assembly, and this defect could be rescued by purified H3.1 complexes containing CAF-1, but not by H3.3 complexes (Tagami et al., 2004). In summary, these experiments demonstrated that H3.3 and H3.1 associate with distinct complexes containing HIRA and CAF-1, and these complexes are critical for replication-independent (HIRA) or replication-dependent (CAF-1) chromatin assembly (Ray-Gallet et al., 2002; Tagami et al., 2004).

Functional roles for HIRA and CHD1 in histone deposition, cellular differentiation, and development

In vivo studies in *Drosophila* show that Hira is required for chromatin assembly and H3.3 deposition in the male pronucleus (Loppin et al., 2005). Maternal HIRA and FLAG-tagged H3.3 accumulate in the male pronucleus following fertilization (Loppin et al., 2005). A specific point mutation in Hira (known as *ssm*) impairs H3.3 deposition to the male pronucleus and prevents male pronucleus formation, providing the first direct link between H3.3 deposition and Hira *in vivo* (Loppin et al., 2005). Without male pronuclear formation, embryos from eggs laid by *ssm* females are haploid, develop only with maternal chromosomes, and die before hatching (Loppin et al., 2005; Loppin et al., 2000). Interestingly, this same mutation in Hira enhanced variegation of a reporter transgene located between heterochromatin and a site for H3.3 replacement, suggesting that Hira-directed H3.3 replacement can help to counteract heterochromatin spreading in somatic cells (Nakayama et al., 2007). Surprisingly, however, later studies demonstrated that paternal chromatin assembly after fertilization is the only developmental process in flies that absolutely required HIRA, as HIRA null flies were viable, but females were sterile (Bonnefoy et al., 2007). Strikingly, HIRA is not required for global H3.3 deposition in *Drosophila* embryos or adult cells, providing intriguing hints that alternate pathways exist to mediate H3.3 nucleosome assembly (Bonnefoy et al., 2007).

Loss of Hira in mammals leads to a far more severe phenotype than the loss of Hira in flies (Bonnefoy et al., 2007; Roberts et al., 2002). Targeted mutagenesis of Hira in mice results in gastrulation defects and patterning abnormalities of mesendodermal derivatives prior to early embryonic lethality (Roberts et al., 2002). Hira ^{-/-} embryos displayed a range of defects, including a distorted primitive streak, delayed and abnormally displaced expression of the primitive streak marker *Brachyury*, severe axial and paraxial defects, abnormally patterned

neuroepithelium, abnormal placentation, and cardiovascular defects (Roberts et al., 2002). Inbred Hira $-/-$ embryos died by E10, and even Hira $-/-$ embryos from mice crossed to an outbred background all died by E11 (Roberts et al., 2002). Despite this severe early embryonic lethality of Hira $-/-$ embryos, Hira $-/-$ ES cells are surprisingly viable (Meshorer et al., 2006).

Hira $-/-$ ES cells proliferate normally and express the pluripotency marker Oct4, but show elevated levels of unbound histones (Meshorer et al., 2006). Hira $-/-$ ES cells display significantly reduced incorporation of both H3.3-YFP and H3.2-YFP as assessed by rapid fluorescence recovery after photobleaching (FRAP), and salt extraction also revealed an increased fraction of soluble H3 (Meshorer et al., 2006). Intriguingly, Hira $-/-$ ES cells underwent accelerated differentiation during the formation of embryoid bodies, despite normal regulation of differentiation specific genes in Hira $-/-$ ES cells (Meshorer et al., 2006). However, while they gave rise to early NPCs, Hira $-/-$ cells failed to progress further into differentiated neurons (Meshorer et al., 2006). Overall, these results indicate that Hira is dispensable for the survival of proliferating ES cells, but is required to maintain normal levels of histone incorporation, and is required for normal gene regulation and survival during cellular differentiation and early development (Meshorer et al., 2006; Roberts et al., 2002).

HIRA has also been implicated in genome-wide DNA demethylation and chromatin remodeling during establishment of the mammalian germ line (Hajkova et al., 2008). Primordial germ cells (PGCs) are specified during early mammalian development (Hayashi and Surani, 2009; Kurimoto et al., 2008; Ohinata et al., 2005). In mice, this germline specification takes place at E7.25, followed by migration of PGCs into the developing gonad (Saga, 2008). Between E11.5 and E12.5, PGCs show genome-wide DNA demethylation and erasure of genomic imprinting (Hajkova et al., 2002; Lee et al., 2002a), along with a concomitant genome-wide

transient or persistent loss of histone modifications such as H3K9me3, H3K9ac, 4/H2AR3me2, H3K27me3, and histone variants such as H1 (Hajkova et al., 2008). This specific timepoint also exactly coincides with the nuclear localization of Hira and the nuclear exclusion and cytoplasmic localization of CAF-1 (Hajkova et al., 2008). These results implicate Hira-mediated histone replacement and potentially H3.3 as likely players in mammalian germline chromatin remodeling, though the lack of H3.3 antibodies has made further study of this question difficult.

More recently, the chromatin remodeling factor CHD1 was also shown to be important for H3.3 deposition in *Drosophila* (Konev et al., 2007). Like Hira, maternal CHD1 is required for H3.3 deposition into the male *Drosophila* pronucleus, and for the participation of the male pronucleus in zygotic mitoses (Konev et al., 2007). Embryos from homozygous CHD1-null females develop with only maternally derived genomes, and die before hatching (Konev et al., 2007). CHD1 physically associates with Hira, and has been suggested to work together with Hira to mediate H3.3 incorporation into chromatin, with Hira proposed to be essential for histone delivery, and CHD1 thought to facilitate histone deposition (Konev et al., 2007).

Recent studies have established that CHD1 plays an important role in the regulation of chromatin state in mammalian cells (Gaspar-Maia et al., 2009; Sims and Reinberg, 2009). The chromodomain of human CHD1 recognizes H3K4me2 and H3K4me3, and CHD1 localizes to euchromatic promoter regions in mouse ES cells (Gaspar-Maia et al., 2009; Sims et al., 2005). CHD1 is critical for the maintenance of open chromatin in ES cells, as downregulation of CHD1 impaired the normally hyperdynamic chromatin state of ES cells and significantly increased the number of H3K9me3 nuclear foci (Gaspar-Maia et al., 2009). Surprisingly, CHD1 RNAi ES cells showed global maintenance of the ES cell transcriptome (Gaspar-Maia et al., 2009). However, CHD1 RNAi ES cells also have a high propensity for neuronal differentiation, and failed to give rise to primitive endoderm or beating cardiac

mesoderm, indicating that CHD1 is required for normal pluripotency (Gaspar-Maia et al., 2009). After the recent finding that CHD1 is required for H3.3 male pronuclear deposition in *Drosophila* (Konev et al., 2007), it has been speculated that CHD1 may be mediating its role in pluripotency and open chromatin through the deposition of H3.3 (Gaspar-Maia et al., 2009).

Localization and function of H2A.Z in undifferentiated and differentiated cells

As described above, the histone H2A variant H2A.Z co-occupies nucleosomes with H3.3 at specific regions of the genome in differentiated cells, including promoters and regulatory elements, and H3.3/H2A.Z nucleosomes have been shown to be less stable *in vitro* (Jin and Felsenfeld, 2006, 2007; Jin et al., 2009).

Recent ChIP-chip studies of H2A.Z have revealed that the genome-wide localization of this variant changes dramatically upon ES cell differentiation (Creyghton et al., 2008). Previous studies in differentiated cells have shown that H2A.Z is enriched at promoters in proportion to transcriptional activity, as well as at regulatory regions (Barski et al., 2007; Jin et al., 2009; Schones et al., 2008). However, in pluripotent ES cells, H2A.Z localizes to the promoters of transcriptionally repressed polycomb target genes, but changes its localization to the promoters of actively transcribed genes upon ES cell differentiation to neural precursor cells (NPCs) (Creyghton et al., 2008). Knockdown studies demonstrated that H2A.Z is required for the repression of polycomb target genes, and that the localization of H2A.Z and the PcG protein Suz12 at target promoters is interdependent, with loss of appropriate localization of one upon depletion of the other (Creyghton et al., 2008). Embryoid body differentiation and tetraploid complementation studies revealed that H2A.Z depleted ES cells were defective in differentiation and lineage commitment, failing to repress pluripotency genes and failing to express differentiation specific genes

in an appropriate spatiotemporal manner (Creyghton et al., 2008). Overall, these studies indicate that H2A.Z plays an unexpected role in gene repression in ES cells, and is required for ES cell differentiation.

Concluding remarks – setting the stage for the study of H3.3 in mammalian embryonic stem cells

As we have seen, the H3.3 *Drosophila* deposition factor Chd1 and the H3.3 nucleosome destabilizing partner H2A.Z are both required for mouse ES cell pluripotency (Creyghton et al., 2008; Gaspar-Maia et al., 2009). In addition, the *Drosophila* H3.3 chaperone HIRA is also required for normal ES cell differentiation (Meshorer et al., 2006; Roberts et al., 2002). Given these findings, as well as the conserved association between H3.3 and active chromatin, H3.3 itself has therefore been speculated to play an important role in establishing and maintaining the pluripotent chromatin state (Creyghton et al., 2008; Gaspar-Maia et al., 2009). However, as of yet, no genome-wide studies in pluripotent cells distinguish between H3 variants, nor do they examine the genome-wide role of Hira or other histone chaperones in specifying H3.3 deposition at specific genomic regions. Although studies in non-mammalian systems point to a role for H3.3 in transcriptional regulation and an absolute requirement for H3.3 in sexual reproduction (Hodl and Basler, 2009; Sakai et al., 2009), the genomic localization and precise function of H3.3 in mammalian cells remains an open question.

With this foundation, I set out to generate tools to examine the genome-wide localization of histone H3.3 in mammalian cells. In Chapter Two, I will describe a zinc finger nuclease technology that enables rapid targeting and point mutagenesis of the endogenous histone H3.3B gene in mouse ES cells. After differentiating these targeted ES cells into neural precursor cells (NPCs), I have generated the first genome-wide profiles of H3.3 localization in pluripotent and

differentiated cells. In Chapter Three, I use wild-type and Hira $-/-$ ES cells to determine the genome-wide dependence of H3.3 localization on the known H3.3 chaperone Hira. Surprisingly, my data show that while Hira is required for H3.3 deposition at active and repressed genes, Hira is not essential for H3.3 deposition at telomeres, ribosomal DNA (rDNA), and specific TFBS. This striking result opened up an entirely new area of investigation for my thesis, and for the field. To identify H3.3 interacting proteins that might mediate Hira-independent deposition at these elements, I purified H3 variant associated proteins from wild-type and Hira $-/-$ ES cells, and my collaborator Dr. Ileana Cristea performed mass spectrometry. I found that the chromatin remodeling protein Atrx and the death domain associated protein Daxx specifically associate with H3.3 in the presence and absence of Hira. Finally, I show that Atrx is required for deposition of H3.3 at telomeres and rDNA, and for the repression of telomeric and ribosomal transcription. In Chapter Four, I will describe some additional collaborative studies in progress, and discuss the implications of my work for the field. Chapter Five provides the materials and methods used in my studies. In addition, I have included an appendix with a more detailed discussion of ChIP and ChIP-seq, as well as the ChIP-seq protocols that I have used throughout my work.

CHAPTER 2

GENOME-WIDE PATTERNS OF HISTONE H3.3 IN MAMMALIAN CELLS ARE DEPENDENT ON ENDOGENOUS AMINO ACID SEQUENCE AND CELLULAR STATE

Summary

As described in Chapter One, recent landmark studies have established genome-wide maps of covalent histone H3 modifications in undifferentiated mammalian embryonic stem (ES) cells and differentiated cells (Barski et al., 2007; Mikkelsen et al., 2007). However, mammalian cells express multiple, distinct genetically encoded variants of histone H3 proteins (Ahmad and Henikoff, 2002). When I began this project, there were no genome-wide or genome-scale profiles of histone H3 variants in any mammalian cell type. Moreover, existing H3 antibodies used by prior studies did not distinguish between these variants. As we have seen in Chapter One, histone H3 variants are extremely similar in sequence, and most of the amino acids that differ between them are inaccessible in the context of the nucleosome (see **Figure 1.16**).

To distinguish histone H3 variants, I began a collaboration with Sangamo Biosciences with the goal of rapidly editing the genomic sequence of an endogenous mammalian histone H3.3 gene. I used designed zinc finger nucleases (ZFN) and short, drug selection-less targeting constructs to rapidly knock-in epitope tags into the endogenous histone H3.3B gene of mouse embryonic stem (ES) cells. To establish the dependence of H3 variant localization on amino acid sequence, I created point mutations to mutate endogenous H3.3B towards H3.2 or H3.1. Using chromosome spreads and chromatin immunoprecipitation assays coupled to

high-throughput sequencing (ChIP-seq), I have established the first genome-wide profiles of histone H3 variants in mammalian ES cells. To facilitate the analysis of this genome-wide ChIP-seq data, I began a collaboration with Assistant Professor Deyou Zheng at the Albert Einstein College of Medicine. My collaborators and I found enrichment of H3.3 in specific genic and intergenic regions and in specific repeats, including telomeres. Surprisingly, we determined that H3.3, like H3K4 methylation, is enriched around transcription start sites of high CpG content promoters in ES cells at both transcriptionally active and transcriptionally repressed genes, as well as the transcription start sites of H3K4me3/H3K27me3 bivalent genes. We also found H3.3 enriched within the bodies of actively transcribed protein coding genes as well as large intervening non-coding RNAs (lincRNAs), correlated with H3K36 methylation. We observed that the transcription end sites of highly transcribed genes are marked by peaks of H3.3 and phosphorylated RNA polymerase II, a phenomenon that has not been previously described in mammalian cells. We also found that H3.3, like H3K4me1, is enriched at all known classes of transcription factor binding sites (TFBS) in mouse ES cells, including intergenic sites. We established that all specific patterns of H3.3 enrichment in mammalian cells are dependent on H3.3-specific amino acid sequence. Following differentiation of ES cells into neuronal precursor cells, we showed that the localization of H3.3 changes with cellular differentiation at specific genic and intergenic sites. The work described in this chapter provides a foundation for integrating histone H3 variants into our understanding of the epigenetic regulation of gene expression in mammalian cells.

Results

Designed zinc finger nucleases enable highly efficient gene targeting of histone H3.3B

In mammals, histone H3.3 is encoded by two different genes (H3.3A and H3.3B)

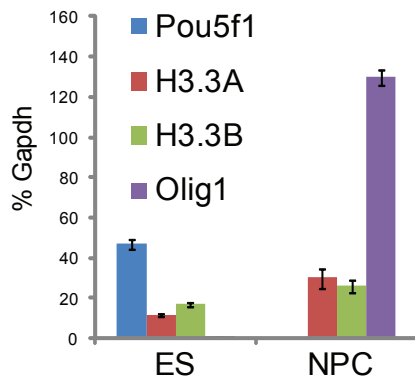


Figure 2.1: Expression of H3.3A and H3.3B in embryonic stem (ES) cells and neural precursor cells (NPCs)

Real-time PCR indicates expression of H3.3A and H3.3B mRNAs in mouse ES cells and differentiated NPCs, as a percentage of Gapdh expression. The pluripotency gene Pou5f1 (Oct4) and the neural precursor marker Olig1 serve as controls for cellular differentiation state.

that code for the same amino acid sequence, but differ in their 5' and 3' UTRs, and also in their patterns of tissue-specific expression (Bramlage et al., 1997). Using RT-PCR, I found that the mRNA of H3.3A and H3.3B are both highly expressed in ES cells and neural precursor cells, with H3.3B slightly more highly expressed than H3.3A in ES cells (**Figure 2.1**).

To establish a rapid system for editing the endogenous H3.3B gene, I began a collaboration with Sangamo Biosciences, which has recently developed a technology for highly efficient targeting of mammalian genes (Urnov et al., 2005).

Gene editing using conventional technology has been very successful (Hogan, 2007), but is generally inefficient (<~0.0025% of electroporated cells) due to the low frequency of homologous recombination, (Carroll, 2008; Hasty et al., 1991; te Riele et al., 1992). Despite the low efficiency of homologous recombination, the introduction of a selectable antibiotic resistance marker (neomycin phosphotransferase) and the use of antibiotics (neomycin, G418) has made conventional gene targeting a viable and successful technology (Hogan, 2007; te Riele et al., 1992). Zinc finger nuclease (ZFN) mediated gene targeting is based on the principle that a targeted DNA double-strand break can significantly enhance the efficiency of homologous recombination at a target site by engaging

the cell's endogenous system of homologous recombination and homology-directed repair (Carroll, 2008; Urnov et al., 2005). The C2H2 zinc finger is the most common DNA binding motif in all metazoa (Tupler et al., 2001; Vaquerizas et al., 2009), and recent work has enabled the engineered design of linked zinc fingers to recognize particular DNA sequences with high specificity (Klug, 2005; Moore et al., 2001). Linked and engineered zinc fingers can be coupled to the nonspecific DNA restriction enzyme *FokI* to target a DNA double strand break to a desired location, producing homologous recombination with specific targeting rates as high as 20% for the introduction of simple point mutations, even in the absence of selection (Urnov et al., 2005).

Taking advantage of the highly efficient technique of targeted gene editing mediated by ZFNs (Carroll, 2008), my collaborators and I used an archive of validated two-finger modules (Doyon et al., 2008; Urnov et al., 2005) to generate designed ZFNs specific to the last coding exon of the H3.3B gene (**Figure 2.2A**). I designed two general strategies to knock-in epitope tags fused to the C-terminus of H3.3B. One strategy was designed to knock-in the open reading frame (ORF) of the enhanced yellow fluorescent protein (EYFP) (**Figure 2.2B**). To address possible concerns regarding the large size of the EYFP tag (~25 kDa), I also designed a parallel strategy to knock-in a shorter C-terminal hemagglutinin (HA) tag (~1 kDa), along with an internal ribosomal entry site (IRES) and separately translated EYFP tag to provide a fluorescent reporter of H3.3B gene expression (**Figure 2.2C**). Promoter-less and selection-less donor constructs were designed and cloned out of mouse B6 genomic DNA to include approximately 0.6kb of genomic sequence at the 5' end of H3.3B, an EYFP sequence with a short linker or a HA-IRES-EYFP sequence fused to the final H3.3B coding exon, and approximately 1.3kb of the H3.3B 3'UTR (**Figure 2.2B-C**). Because of the high efficiency of ZFN mediated gene targeting (Carroll, 2008; Urnov et al., 2005), our donor constructs are far

Figure 2.2: Zinc finger nuclease targeting scheme

A. Schematic of zinc finger nucleases (ZFNs) recognizing endogenous H3.3B (h3f3b) sequence near the end of the last coding H3.3B exon on chromosome 11.

B-C. Designed zinc finger nucleases (ZFN) target the endogenous sequence of the final coding exon of the mouse histone H3.3B gene on chromosome 11. Exon coding sequences are thick rectangles, while non-coding untranslated sequences are thin rectangles. ZFN targets double-strand break (DSB) formation at H3.3B, with H3.3B-EYFP or H3.3B-HA-IRES-EYFP donors generating a new H3.3B allele with a C-terminal EYFP tag (**B**), or a C-terminal HA tag (red) and an internal ribosomal entry site (IRES, blue) followed by an EYFP reporter (**C**). Location is indicated for genomic forward primer outside donor sequence and reverse primer within EYFP sequence used for PCR amplification and sequencing (**2.6**) of modified H3.3B alleles.

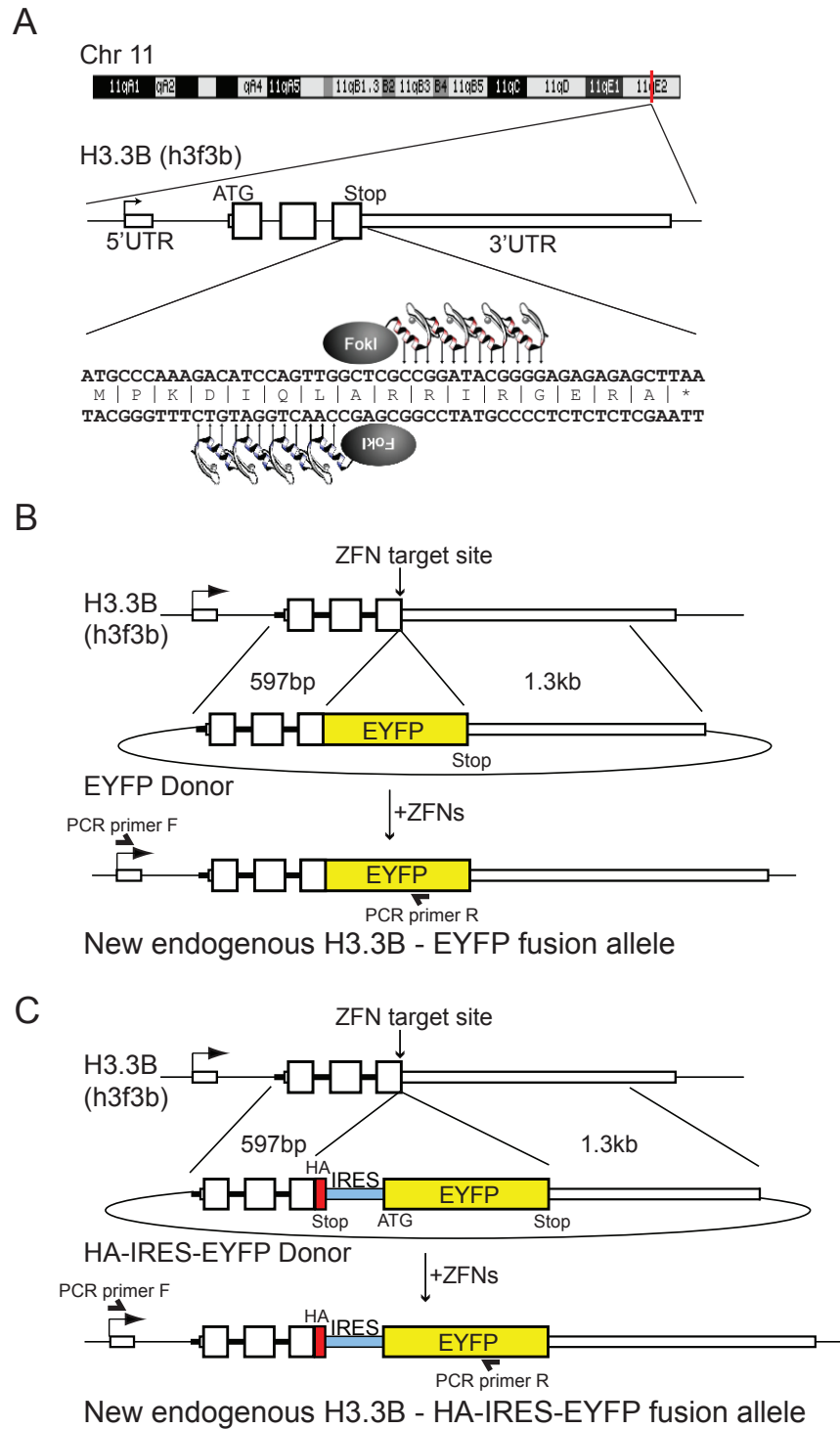


Figure 2.2

shorter than traditional mouse targeting constructs (Deng and Capecchi, 1992; Hasty et al., 1991).

I found ZFN-mediated targeting to be highly efficient at driving homology-directed addition of EYFP or HA epitope tags, as well as targeted editing of the endogenous H3.3B gene. To evaluate the efficiency of ZFN-mediated targeting, the donor construct was transfected in the presence and absence of ZFN. 5 days post-transfection with H3.3-EYFP donor construct in the absence of ZFN, less

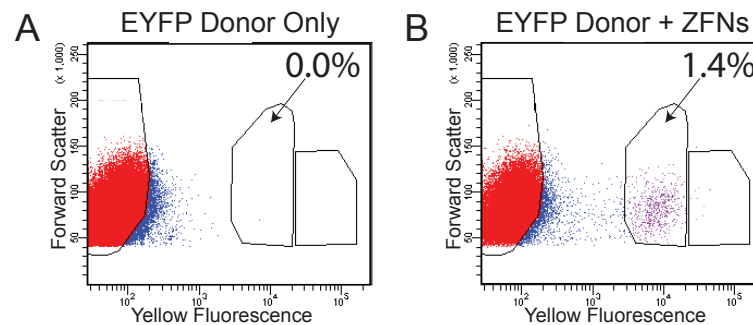


Figure 2.3: Zinc finger nucleases increase the efficiency of H3.3B gene targeting in mouse embryonic stem cells by over 1000 fold.

A-B. Mouse ES cells were transfected with H3.3-EYFP donor alone (**A**) or H3.3-EYFP donor + zinc finger nucleases (ZFN) (**B**) and analyzed by flow cytometry 3 days post-transfection. Percentage EYFP positive are indicated.

than 1/100,000 cells were EYFP+ (**Figure 2.3A**), while in the presence of ZFN, between 1-2% of cells were EYFP+ (**Figure 2.3B**), an over 1000 fold increase in the frequency of EYFP+ cells.

To evaluate the dependence of H3.3 localization on amino acid sequence, we also modified our donor constructs to knock-in point mutations at H3.3 codons 31, 87, 89, 90, and 96 and simultaneously mutate endogenous H3.3B towards H3.2 or H3.1 via marker co-conversion (**Figure 2.4**). While greater than 1% of ES cells were correctly targeted with H3.3-EYFP or H3.3-HA-IRES-EYFP at one H3.3B allele, we found that the introduction of point mutations decreased the efficiency of gene targeting (**Figure 2.5**). Less than 0.5% of cells became EYFP+ following

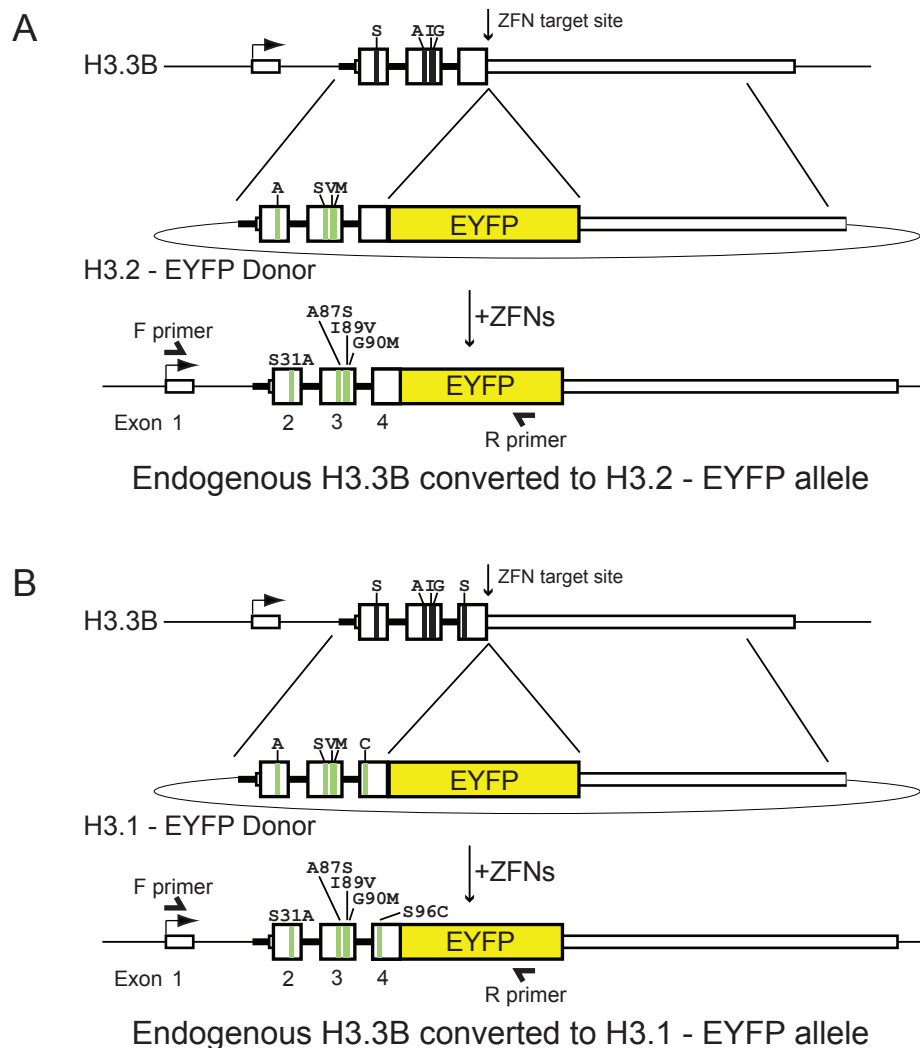


Figure 2.4: Gene conversion of endogenous H3.3B to epitope-tagged H3.2 or H3.1

A-B. Schematic of ZFN-mediated targeting of H3.3B EYFP tag and simultaneous co-conversion of H3.3B sequence to H3.2 (**A**) or H3.1 (**B**). A similar ZFN-mediated co-conversion strategy was used with H3.2-HA-IRES-EYFP and H3.1-HA-IRES-EYFP donor constructs. Point mutations in H3.3 sequence are indicated in green. F and R primers used for PCR (**2.6**) are indicated as in **2.2B-C**.

transfection of H3.2-HA-IRES-EYFP or H3.1-HA-IRES-EYFP donor constructs, and less than 1/5 of these EYFP⁺ clones were fully corrected to H3.2 or H3.1 codons at the modified allele (**Figure 2.5**, and data not shown). Notably, these percentages are still far higher than traditional gene targeting in the absence of ZFNs (te Riele et al., 1992).

Fluorescent cells were manually picked or sorted using flow cytometry (**Figure 2.3**, **2.5**), and screened by PCR (**Figure 2.6A**), sequencing of both H3.3B

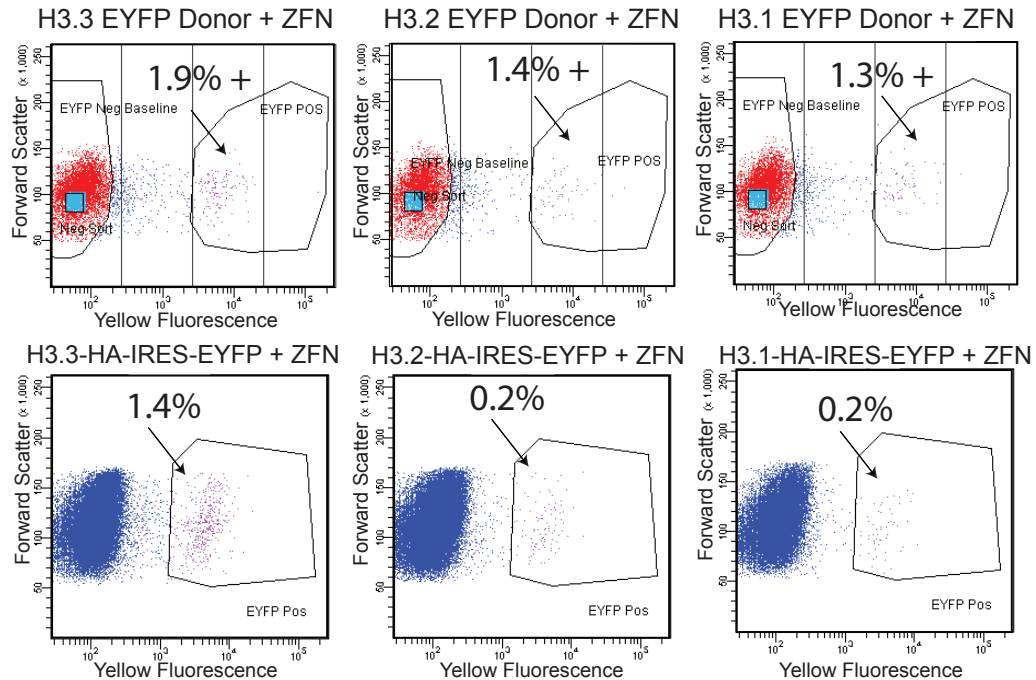


Figure 2.5: Efficiency of H3.3B gene targeting is dependent on donor sequence homology. Introduction of point mutations decreases efficiency of targeting for H3-EYFP donors (top row) and H3-HA-IRES-EYFP donors (bottom row). Mouse ES cells were transfected with H3-EYFP donor + ZFN or H3-HA-IRES-EYFP donor + ZFN and analyzed by flow cytometry 3 days post-transfection. % EYFP positive cells are indicated.

alleles (**Figure 2.6B**), and Southern blotting (**Figure 2.6C**). Nearly all (>90%) of the EYFP+ ES lines targeted with H3.3B-EYFP were correctly targeted at the H3.3B allele, as confirmed by genomic PCR and sequencing. Multiple heterozygous lines of ES cells were isolated with one targeted allele expressing H3.3-EYFP, H3.2-EYFP, H3.1-EYFP, H3.3-HA-IRES-EYFP, H3.2-HA-IRES-EYFP, or H3.1S31-HA-IRES-EYFP (**Figure 2.6B-C**). Initial screening was unable to isolate a line fully converted to H3.1-HA; hence, lines denoted H3.1S31-HA retain the H3.3 serine at position 31. All heterozygously tagged ES cells retain three “wild-type” copies of H3.3 genes, including one copy of unmodified H3.3B (**Figure 2.6C**), and two copies of H3.3A. Western blots demonstrate comparable protein expression in the wild-type and mutated epitope-tagged H3.3B ES lines (**Figure 2.6D**).

Although ZFN-mediated gene targeting represents a far more efficient method of mutagenesis than conventional methods (Carroll, 2008; te Riele et al.,

Figure 2.6: Generation and validation of H3.3B targeted ES lines

A. Genomic DNA from wild-type and targeted ES lines were subject to PCR amplification, using primers indicated in **2.2** and **2.4**.

B. Nucleotide sequence of wild-type and modified H3.3B alleles. PCR fragments shown in **A** were sequenced and aligned to sequence from endogenous wild-type H3.3B allele. These nucleotide sequences correspond to the amino acids shown in **Figure 1.16**. Note that in cell lines targeted with H3.1-HA-IRES-EYFP, mutagenesis was incomplete, with H3.3B mutated to H3.1 sequences at positions 87, 89, 90, and 96, but retaining a serine at position 31. These cell lines are hereafter referred to as H3.1S31.

C. Genomic DNA from wild-type and targeted ES cells was digested with BsrBI and analyzed by Southern blot, using a 638bp AvaII fragment of the H3.3B donor as probe.

D. Whole cell extracts from wild-type ES and targeted ES cell lines were immunoblotted with GFP, HA, and C-terminal (untagged) H3 antibodies as indicated.

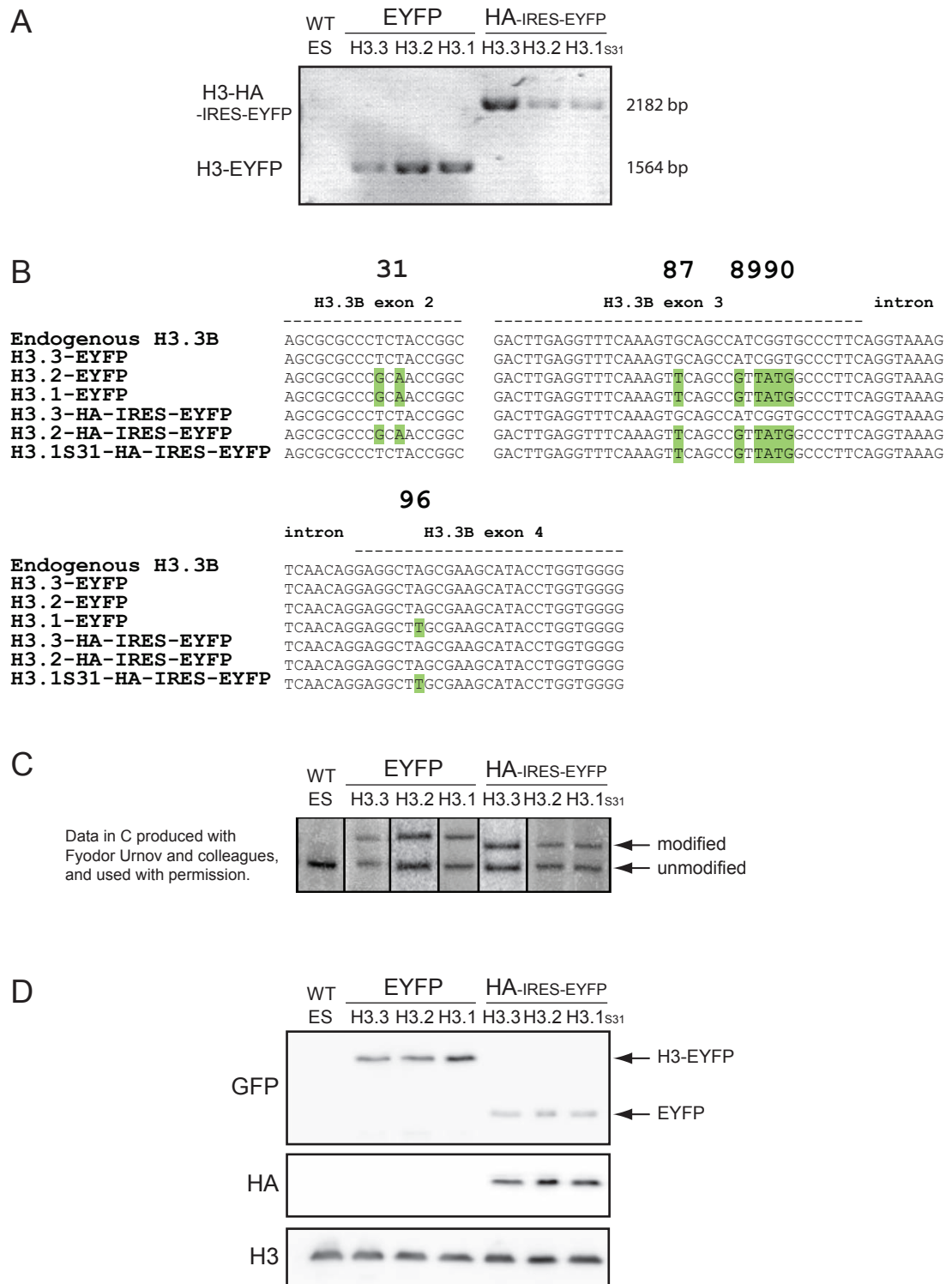
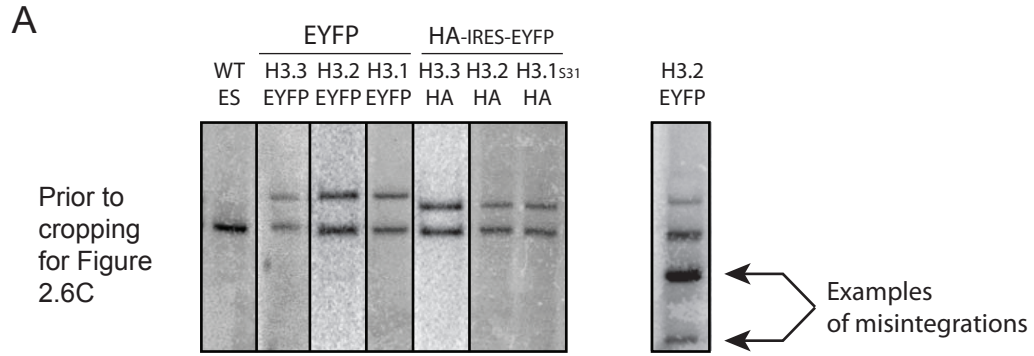


Figure 2.6

1992; Urnov et al., 2005), it is still necessary to carefully screen targeted cells to ensure desired genomic editing. For example, upon Southern blotting, my collaborators and I found some clones (<20%) where the donor sequence had misintegrated elsewhere in the genome (**Figure 2.7A, right**). In addition, I also found that both the 'modified' and 'unmodified' allele had occasional (~10% of sequenced clones) mutations and small deletions at the ZFN target site (**Figure 2.7B**), a phenomenon likely to reflect incomplete DNA repair by non-homologous end joining (Carroll, 2008). It is important to note that the only 'selection' in this case is EYFP fluorescence. Although the frequency of undesired mutations (~10-20%) following ZFN-mediated gene targeting point to the necessity of careful screening, the frequency of desired clones remains far higher than conventional targeting, where several hundred clones are often screened even following antibiotic selection (te Riele et al., 1992).

Specific enrichment of H3.3 in gene-rich euchromatic chromosome arms and telomeres is dependent on amino acid sequence

To assess the incorporation and chromosomal distribution of H3 variants in mouse ES cells, I initially examined chromosome spreads from metaphase-arrested ES cells. Mouse chromosomes are acrocentric, with visible regions of DAPI-dense heterochromatin around their centromeres (Martens et al., 2005) (see **Figure 1.19**). Analysis of H3.3-EYFP and H3.3-HA ES lines demonstrates that H3.3 is specifically localized to euchromatic chromosome arms in mouse ES cells, and is markedly depleted in DAPI-dense pericentric heterochromatin (**Figure 2.8A-B, 2.9**). In addition, telomeric foci of H3.3 are visible on the Y-chromosome in mouse ES cells (**Figure 2.8A-B, 2.9B-C, K-L, inset**). This result is in agreement with recent studies that found H3.3 serine 31 phosphorylation (H3.3Ser31p) enriched at telomeres in mouse ES cells (Wong et al., 2009). However, it is notable that



Data in A produced with Fyodor Urnov and colleagues, and used with permission.

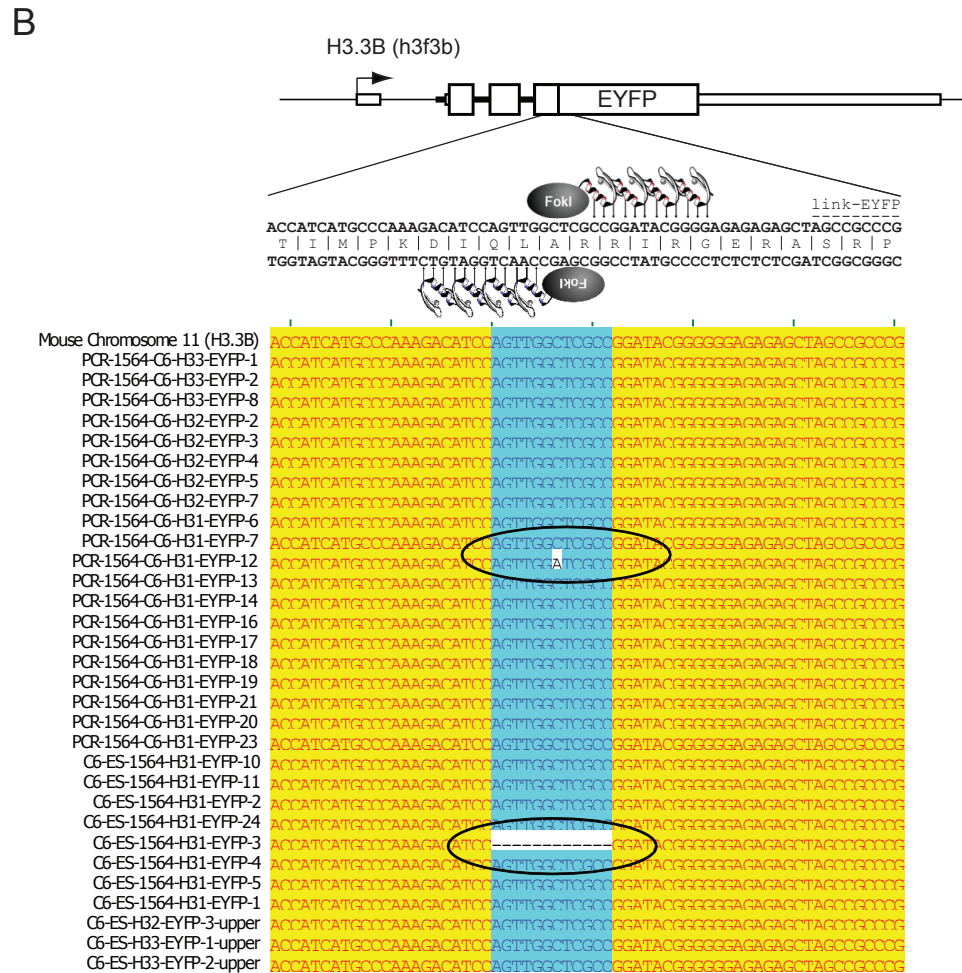


Figure 2.7: Misintegrations and mutations in some ZFN targeted lines

A. Uncropped Southern blots as in 2.6C. Some clones (for example, right) show additional genomic integrations of donor construct.

B. Alignment of modified H3.3B alleles. Note examples of point mutation and deletion at the ZFN target site. A slightly higher frequency (~10%) of similar deletions and point mutations were also observed upon sequencing the 'unmodified' H3.3B allele (not shown).

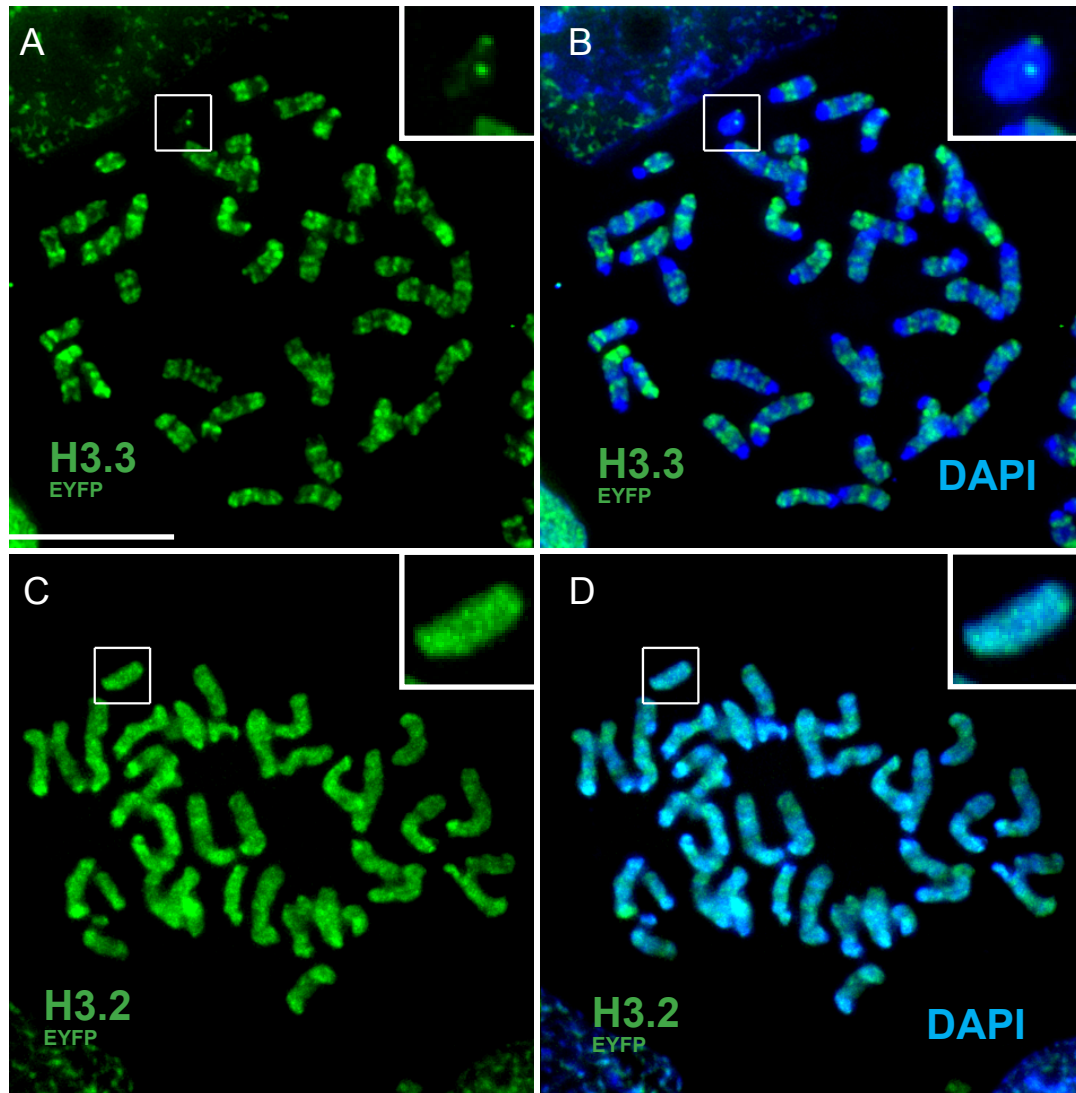


Figure 2.8: Genomic localization of histone H3.3 is dependent on amino acid sequence

A-D. H3.3-EYFP and H3.2-EYFP ES cells were treated with 0.2 $\mu\text{g/mL}$ colchicine for 2 hrs at 37°C. Samples were prepared by cytopsin, fixed, stained with DAPI, and visualized using DeltaVision deconvolution microscopy. EYFP and DAPI overlays of maximum-intensity projections from multiple Z-stacks are shown. Bar: 10 μm . Panels **A** and **C** show EYFP fluorescence. Panels **B** and **D** show EYFP and DAPI overlays. Inset shows a higher magnification image of the Y-chromosome, which was identified by its size and uniform DAPI density. For complete chromosome spreads including DAPI alone, as well as H3.1-EYFP and H3.3-HA, please see the following **Figure 2.9**.

Figure 2.9: Complete chromosome spreads of H3.3-EYFP, H3.2-EYFP, H3.1-EYFP, and H3.3-HA ES cells

A-F. Full representative metaphase chromosome spreads as described in **2.8**, including DAPI fluorescence (**A,D**).

G-I. H3.1-EYFP localization is widespread throughout metaphase chromosomes, similar to H3.2-EYFP.

J-L. H3.3-HA localization is similar to H3.3-EYFP localization. Chromosome spreads were performed using anti-HA (HA.11) for immunofluorescence of H3.3-HA (green). The Y-chromosome in all spreads was identified by its size and DAPI pattern, and is shown in the inset of **A-L**.

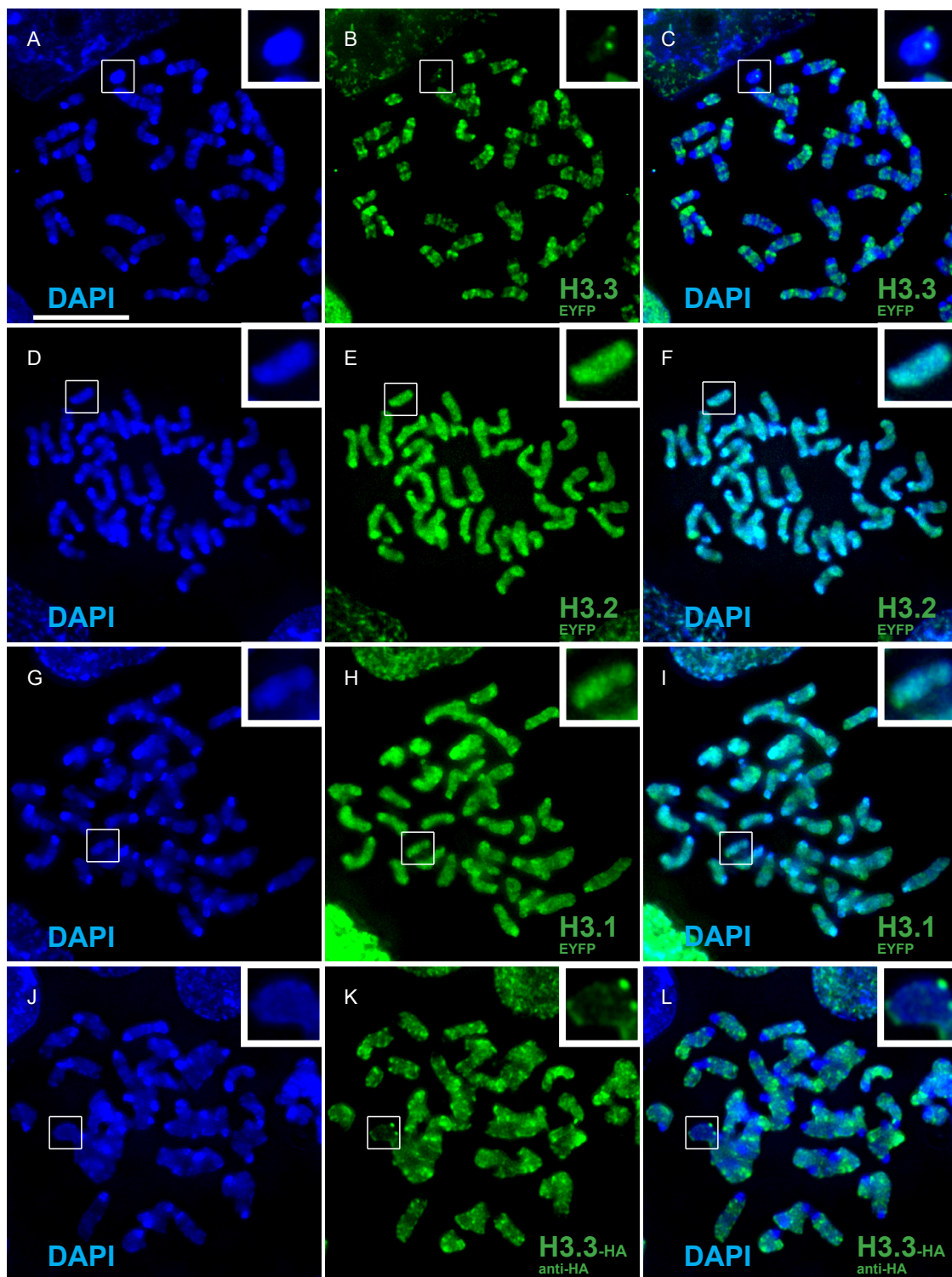


Figure 2.9

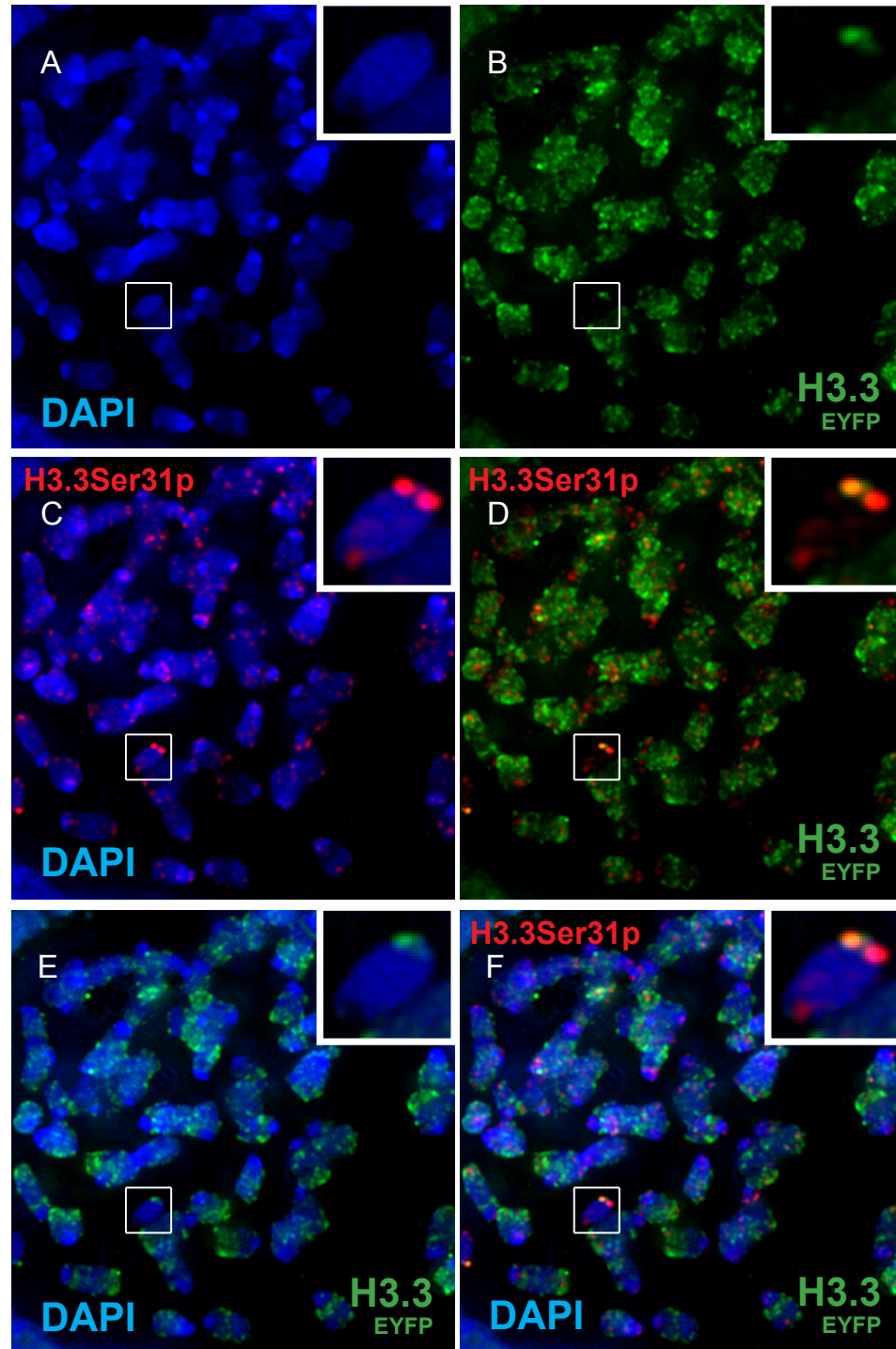


Figure 2.10: Foci of H3.3 serine 31 phosphorylation represent a small subset of overall H3.3 on metaphase chromosomes

A-F. Metaphase chromosome spreads of H3.3-EYFP ES cells were performed as in previous figures, and stained with DAPI (blue), anti-GFP (green), and H3.3Ser31p (red). Inset shows the Y-chromosome. Note partial co-localization of H3.3 and H3.3Ser31p on Y-chromosome telomeric foci.

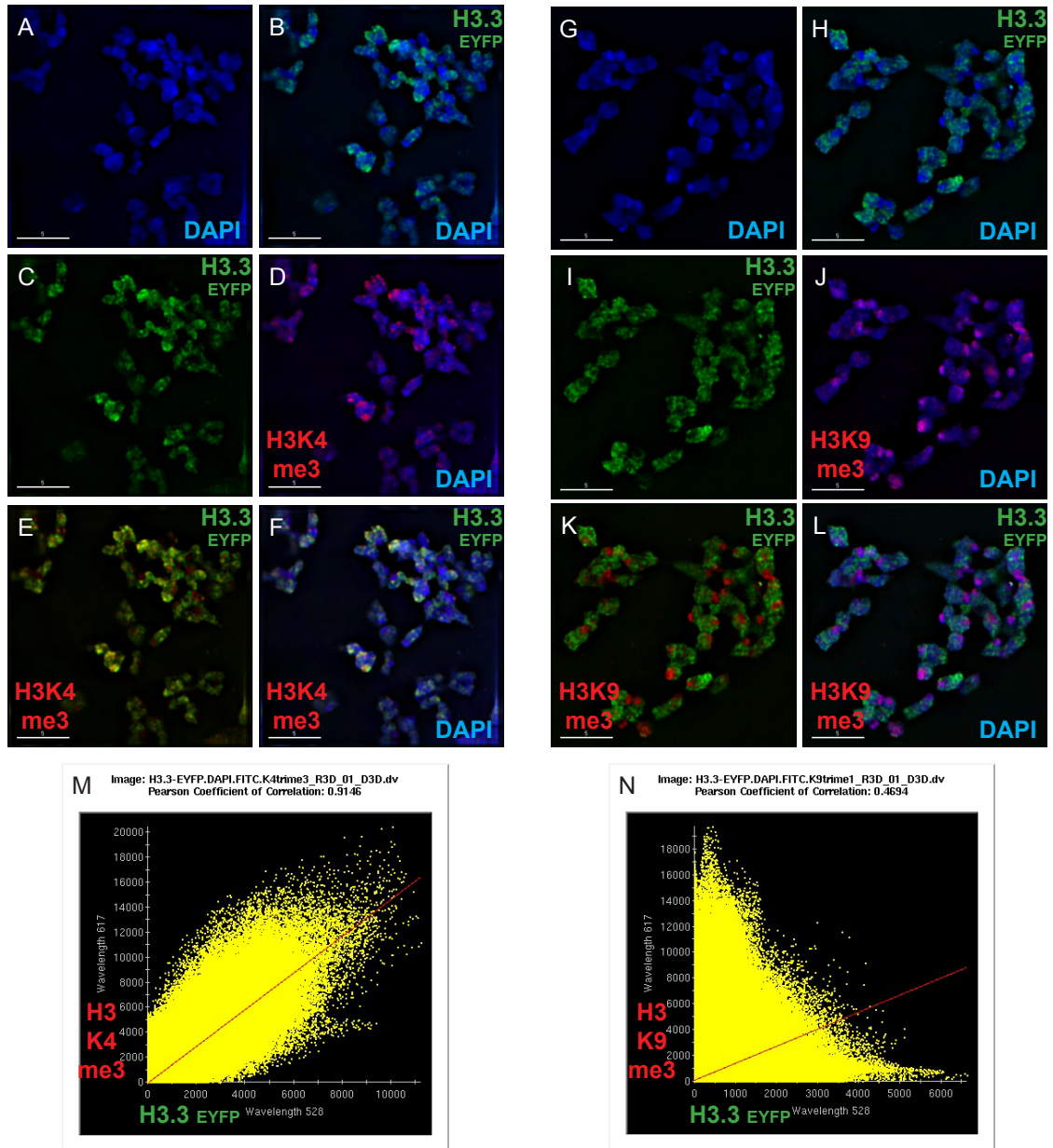


Figure 2.11: H3.3 significantly colocalizes with H3K4me3, and does not colocalize with H3K9me3 pericentric heterochromatin on metaphase chromosomes

A-F. Metaphase chromosome spreads were stained with DAPI (blue), anti-GFP (green), and anti-H3K4me3 (red). Note overlap of H3.3 and H3K4me3 (**E**).

G-L. Metaphase chromosome spreads were stained with DAPI (blue), anti-GFP (green), and anti-H3K9me3 (red). Note lack of overlap of H3.3 and H3K9me3 (**K**).

M-N. Scatterplots depicting colocalization of H3.3-EYFP and H3K4me3 (**M**) and H3.3-EYFP and H3K9me3 (**N**), as analyzed with DeltaVision software through multiple Z-stacks. Each dot represents a pixel, with red or green signal intensity plotted on the Y or X axes. The relative intensities of H3.3-EYFP and H3K4me3 are significantly correlated (Pearson Coefficient of Correlation 0.9146), while H3.3-EYFP and H3K9me3 are poorly correlated.

most H3.3 is found incorporated on chromosome arms, while telomeric foci of H3.3Ser31p represent only a small subset of overall H3.3 incorporation, and most clearly co-localize with H3.3 at the Y-chromosome (**Figure 2.10**).

To determine if H3.3 colocalizes with euchromatic or heterochromatic histone modifications, I co-stained H3.3-EYFP ES cell metaphase chromosomes with an antibody to H3K4me3 or H3K9me3. H3K4me3 is associated with transcription start sites of active and inactive genes (Guenther et al., 2007; Ruthenburg et al., 2007a), while H3K9me3 is associated with repetitive pericentric heterochromatin in areas such as major and minor satellite repeats (Martens et al., 2005) (see **Chapter One**). Colocalization analysis revealed that H3.3 is significantly colocalized with H3K4me3 on a chromosome-wide level (**Figure 2.11, A-F, M**, note yellow color in **E**). In addition, H3.3 does not colocalize with H3K9me3, further indicating that H3.3 is largely excluded from H3K9me3-rich pericentric heterochromatin (**Figure 2.11, G-L, N**). These data provide further evidence that H3.3 is enriched in euchromatic chromosome arms, and is largely excluded from H3K9me3-rich pericentric heterochromatin.

To establish whether the specific patterns I observed for H3.3 are dependent on amino acid sequence, I also performed chromosome spreads from ES cells in which one allele of H3.3B was mutated to H3.2-EYFP or H3.1-EYFP (**Figure 2.8A-D, 2.9**). Metaphase chromosome spreads demonstrate that while each of our EYFP-tagged H3 variants are incorporated into chromatin in ES cells, their chromosomal localization is dependent on amino acid sequence (**Figure 2.8A-D**). Despite expression of H3.2 or H3.1 from the endogenous H3.3B promoter, with identical C-terminal epitope tags, specific patterns of H3.3 enrichment were abolished upon mutation of H3.3 to H3.2 or H3.1, and exclusion from pericentric heterochromatin is lost (**Figure 2.8C-D, 2.9**). In contrast to the telomeric localization of H3.3 on the Y-chromosome, H3.2 and H3.1 are incorporated throughout the

largely heterochromatic Y-chromosome (**Figure 2.8C-D**, inset, **2.9**). Notably, I did not observe significant differences in localization between H3.2 and H3.1, both of which are broadly distributed throughout metaphase chromosomes (**Figure 2.9**). These results demonstrate that the sequence of an endogenous H3.3 gene determines its genomic localization in mammalian cells.

ChIP-seq provides genome-wide maps of histone H3 variants

To extend my analysis at higher resolution, I next used ChIP-seq to analyze the genome-wide localization of histone H3 variants in mouse ES cells. As described in **Chapter One**, ChIP-seq provides an enormously powerful, relatively unbiased technique for determining genome-wide localization and profiling chromatin state (Barski and Zhao, 2009; Mendenhall and Bernstein, 2008). ChIP-seq provides digital, quantitative data regarding genome-wide patterns of enrichment, allows for the interrogation of repetitive sequences, and enables mononucleosome level resolution (Barski and Zhao, 2009; Mendenhall and Bernstein, 2008).

As ChIP-seq is a relatively new technique, I performed multiple controls to validate these findings, including comparison of native versus crosslinking ChIP-seq, comparison of HA and EYFP tagged variants, sequencing of input ChIP DNA, sequencing of anti-HA ChIP samples from untagged wild-type ES cells, ChIP-seq of control histone modifications, and real-time PCR validation of independent ChIP samples (see **Chapter Five** and **Appendix** for details). I also found remarkably similar genome-wide patterns for H3.3 in mouse ES cell lines (C57BL/6J ES and F1 hybrid ES) of significantly distinct genetic backgrounds (Mikkelsen et al., 2007).

Following alignment of our sequenced ChIP DNA to the reference mouse genome, I observed that gene-rich regions of the genome show greater enrichment of H3.3 than H3.2 or H3.1S31 (**Figure 2.12, 2.13**). I also profiled the genome-wide localization of H3K4me1, H3K4me3, H3K27me3, and H3K36me3, and compared

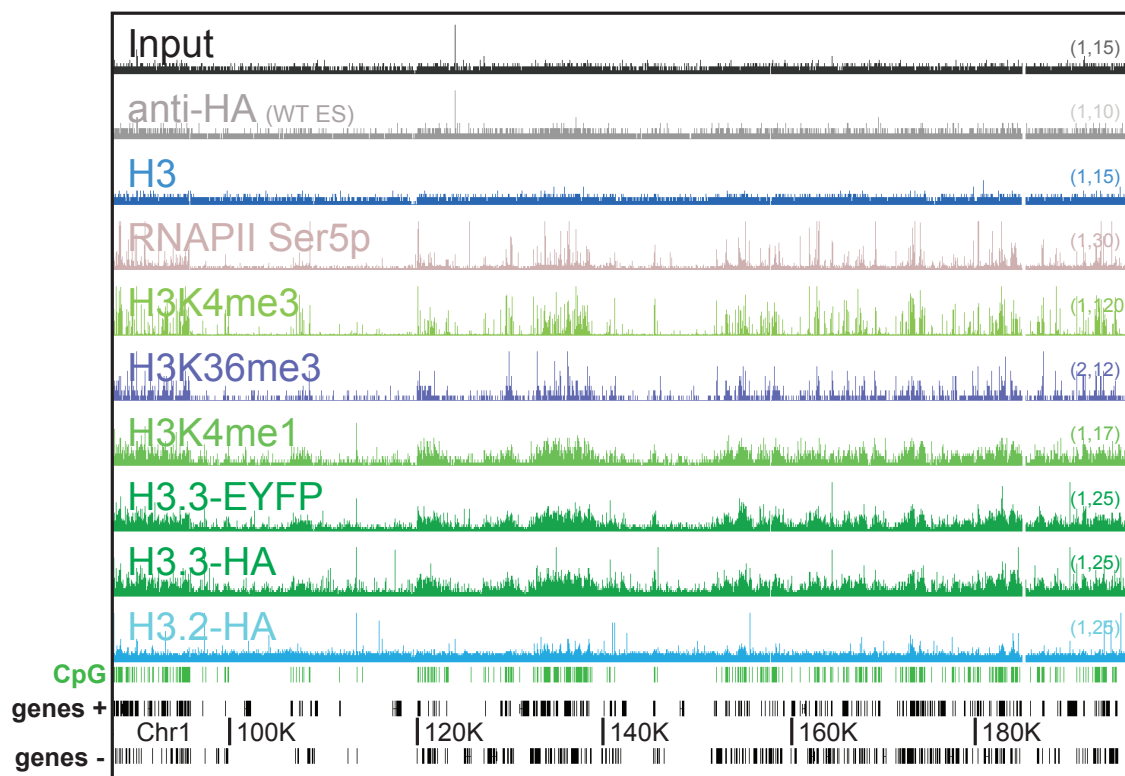


Figure 2.12: Chromosome-scale map of ChIP-seq data reveals genic enrichment of H3.3

ChIP-seq profiles across a large representative genomic region (120 MB of chromosome 1) in mouse ES cells, using antibodies as indicated. Y-axis represents the number of reads spanning a genomic position, with scale indicated on the right (baseline, maximum). CpG islands (green) and genes (positive strand (+) and negative strand (-)) along with chromosome coordinates are shown below plots. Note the enrichment of H3.3 with genes, CpG islands, and markers of active chromatin, including RNAPII Ser5p, H3K4me3, H3K36me3, and H3K4me1. Upon mutation of H3.3-HA to H3.2-HA, enrichment in genic regions is significantly reduced. See later figures for higher resolution analyses. Data for general anti-H3 in ES cells is from (Mikkelsen et al., 2007). Data for H3.1S31-HA is nearly identical to H3.2-HA, and is not shown.

these results to established profiles for these modifications in ES cells and NPCs (Meissner et al., 2008; Mikkelsen et al., 2007). In all cases, histone modification profiles from tagged ES cells were nearly identical to previously established datasets (**Figure 2.14**, **Figure 2.15**). The remarkable similarity between my multiple controls and previous datasets from other laboratories underscores the reliability and reproducibility of ChIP-seq, and gives me greater confidence in my findings. Moreover, this result suggests that heterozygous mutation of H3.3B to

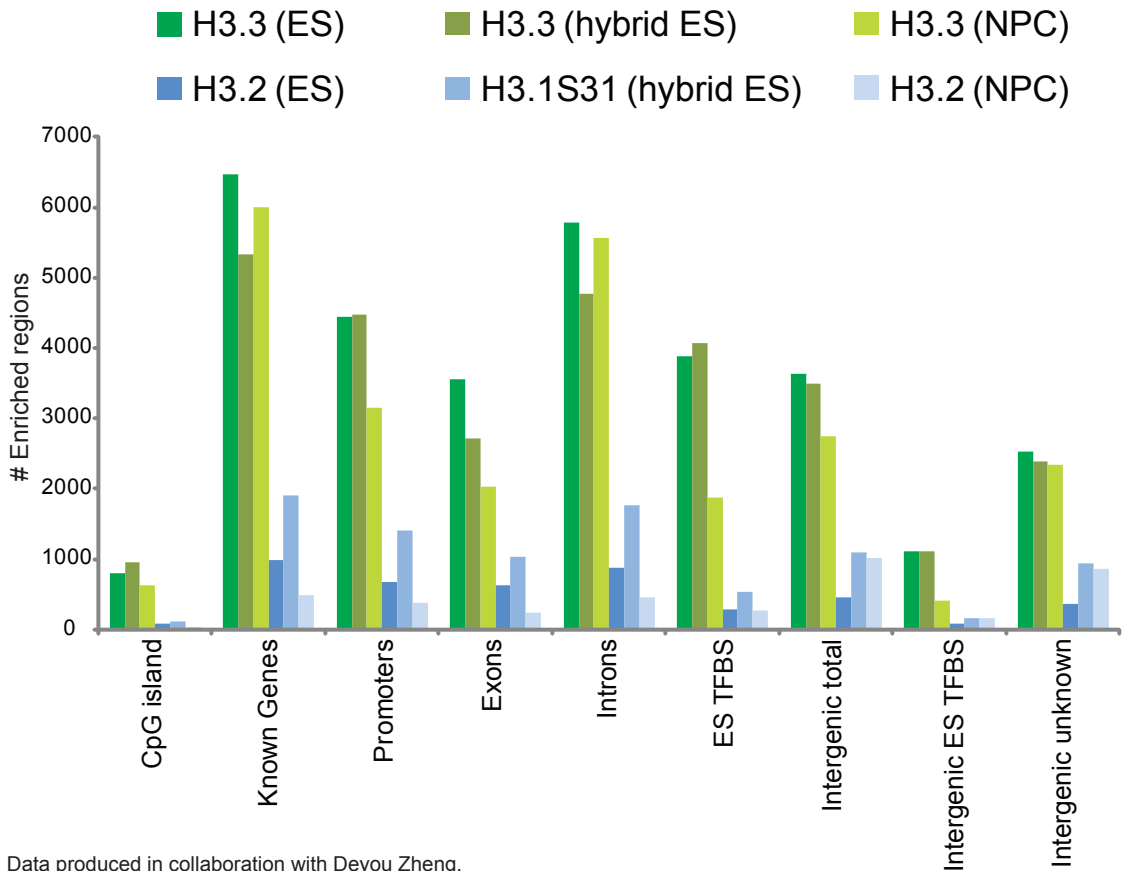


Figure 2.13: Genomic locations of H3 variant enriched regions

Enriched regions were identified for H3 variants as described in Chapter Five, and their relationship with annotation is plotted. Y-axis represents the number of enriched regions, while the X-axis represents different categories of gene annotation and transcription factor binding sites (TFBS). Known ES TFBS are from (Chen et al., 2008). “Hybrid ES” experiments were performed in a genetically divergent F1 hybrid (castaneus/129) ES line, the same ES line described in previous published ChIP-seq studies (Mikkelsen et al. 2007). I have found that genome-wide patterns of H3.3 enrichment are extremely similar in all analyzed ES lines, regardless of genetic background. See Chapter Five for further details.

H3.2 or H3.1S31 does not significantly alter genome-wide patterns of histone modifications.

My collaborators and I found 10,099 total genic and intergenic regions highly enriched for H3.3 in crosslinking ChIP-seq of ES cells, in contrast to 1,442 regions enriched for H3.2, and we found that gene-rich regions of the genome show greater enrichment of H3.3 than H3.2 or H3.1S31 (**Figure 2.12., 2.13**). On a chromosomal scale map, H3.3 enrichment correlated with markers of transcription, including

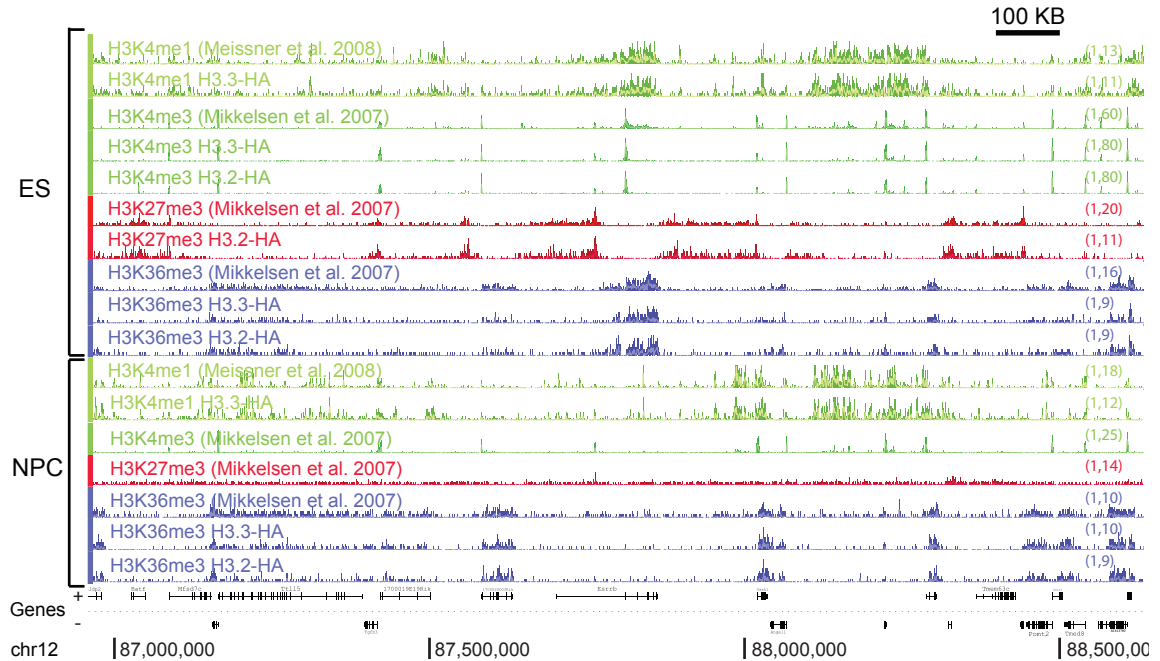


Figure 2.14: Genome-wide patterns of histone modifications are not affected by heterozygous conversion of H3.3B to H3.3-HA or H3.2-HA

Data is shown for a representative region of chromosome 12 containing *Esrrb*. H3 variant patterns for this gene are shown in **2.26A**. Y-axis represents the number of reads spanning a genomic position, with scale indicated on the right. Genes (+ and - strand) along with chromosome 12 coordinates are shown below plots. Data for previously established datasets in mouse ES cells and NPC are indicated (Meissner et al., 2008; Mikkelsen et al., 2007), and were downloaded from <http://www.ncbi.nlm.nih.gov/geo/>. For patterns of histone modifications in H3.3-HA vs H3.2-HA ES cells, see also **2.24**.

RNAPII, H3K4me3, H3K36me3, and H3K4me1 (**Figure 2.12**). Unbiased clustering analyses confirmed the genome-wide correlation of H3.3 with active histone modifications, particularly H3K4me1 (**Figure 2.15**). Despite different epitope tags, we found extremely similar profiles of H3.3-HA and H3.3-EYFP (**Figure 2.12**). In addition to genes, genome-wide analysis revealed specific enrichment of H3.3 at previously identified genic and intergenic ES TFBS (Chen et al., 2008b), as well as peaks of H3.3 in specific unannotated intergenic regions (**Figure 2.13**). These chromosome-wide patterns of H3.3 were largely eliminated upon mutation of H3.3B towards H3.2 or H3.1 (**Figure 2.12**). Our data demonstrate that the amino acid sequence of an endogenous H3.3 gene determines its genomic localization in

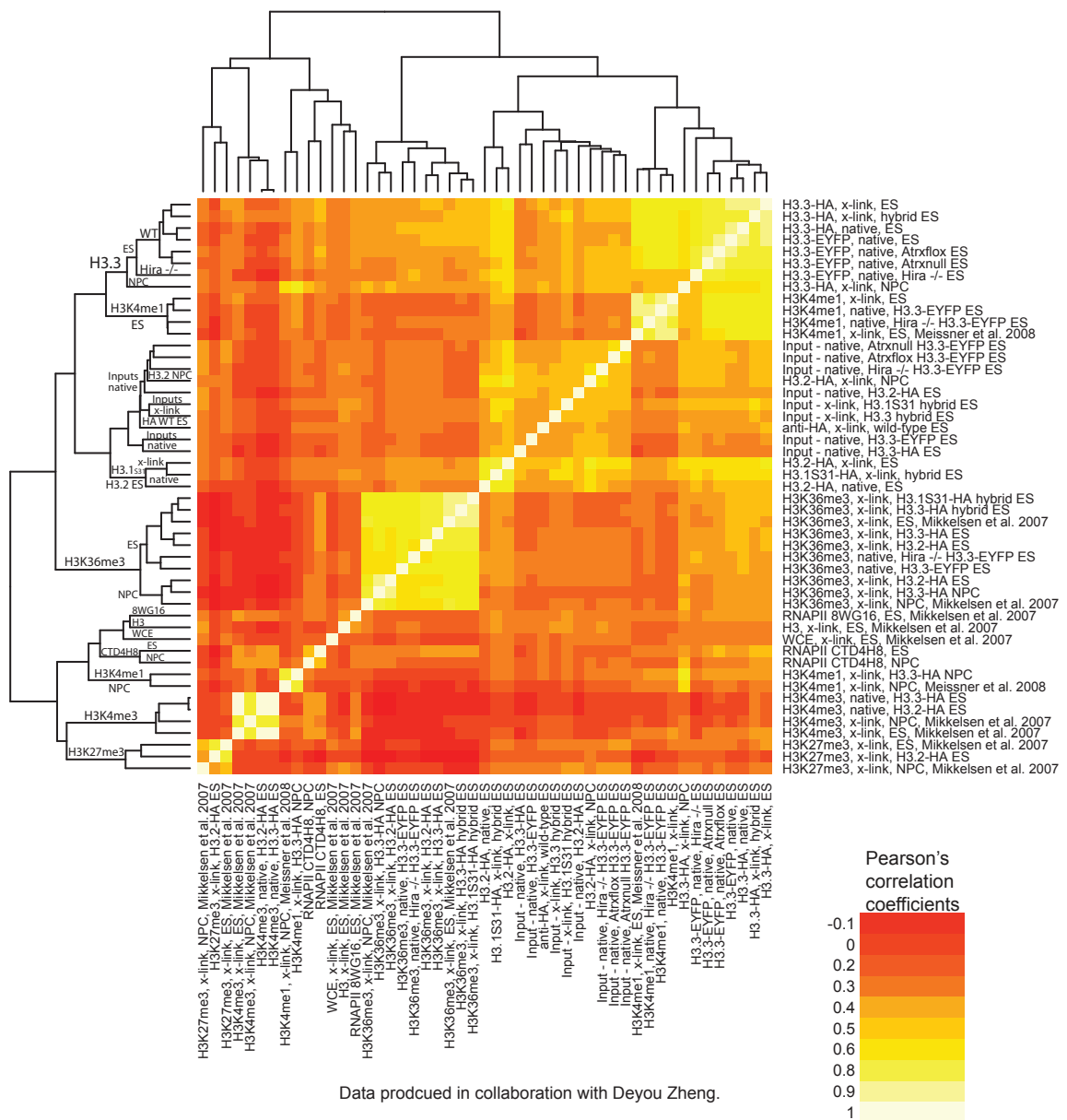


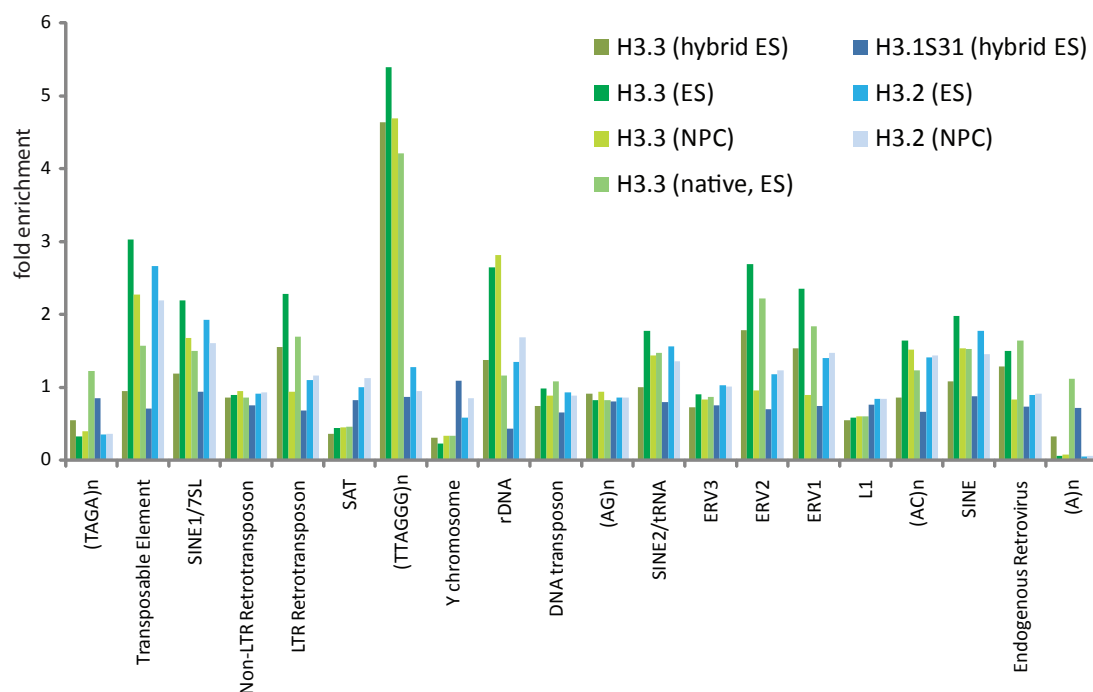
Figure 2.15: Genome-wide correlation of ChIP-seq datasets

Mouse chromosomes were divided into non-overlapping 2-kb regions and then the number of ChIP-Seq reads in each region was counted to generate a vector representing the genome-wide distribution of each ChIP-Seq experiment. Pearson's correlation coefficients were computed between every pair of ChIP-Seq data and the resulting correlation matrix was clustered and shown as a heat map.

mammalian cells.

Genome-wide patterns of H3 variants at repetitive elements

Unlike ChIP-chip technologies, which are limited by probe design, ChIP-seq enables the localization of proteins to repetitive regions of the genome (Mikkelsen et al., 2007). To analyze H3 variant enrichment in different classes of repeats, we determined the relative enrichment of specific repeat sequences from our ChIP-seq experiments and compared it to that from control input DNA (see **Chapter Five**). H3.3 was reproducibly depleted in satellite repeat sequences and in Y-chromosomal repeat DNA, while enriched in rDNA repeats (**Figure 2.16**). Our ChIP-seq results are consistent with the visible exclusion of H3.3 from DAPI-dense pericentric heterochromatin (**Figure 2.8A-B, 2.9, 2.11**), as pericentric heterochromatin is



Data produced in collaboration with Deyou Zheng.

Figure 2.16: Enrichment of H3 variants in different classes of repeats

Analysis of H3 variants in repetitive elements. Reads from HA-tagged H3 variants were aligned directly to a library of mouse consensus repetitive sequences (<http://www.girinst.org>). Note particular enrichment of H3.3 in the canonical (TTAGGG)_n vertebrate telomeric repeat sequence (Meyne et al., 1989). Data are represented as fold enrichment over input.

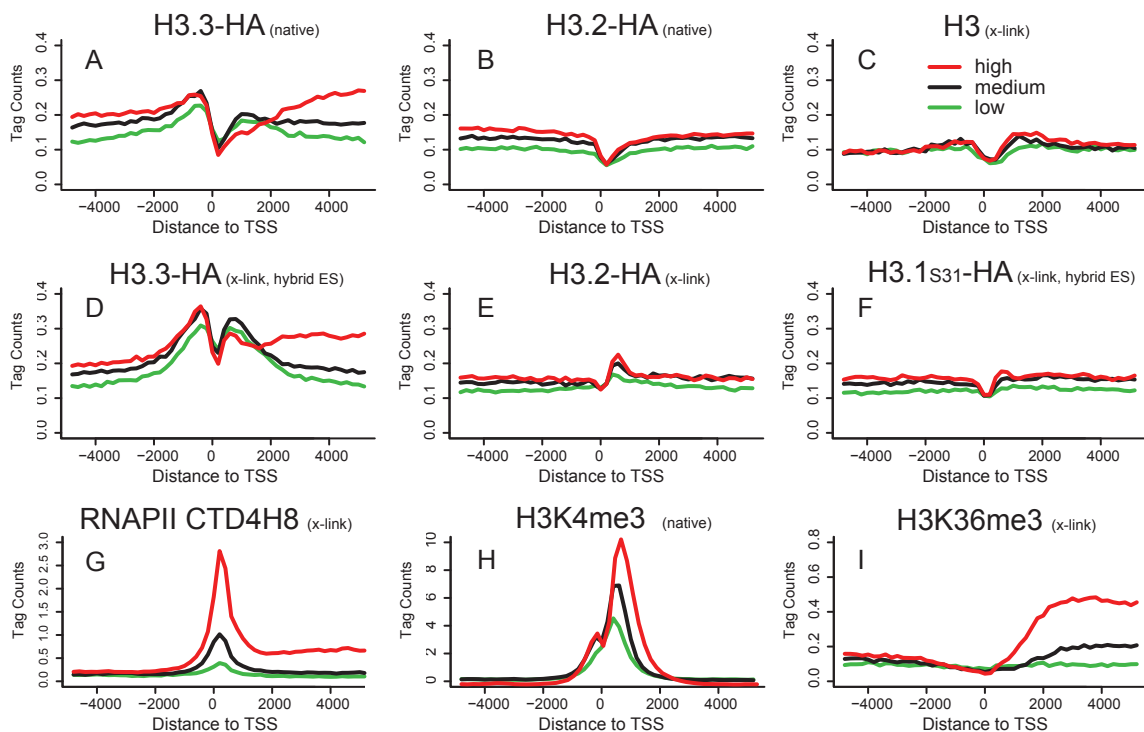
enriched in satellite repeats (Martens et al., 2005), and the relative depletion of H3.3 from the Y-chromosome. Notably, we found the most significant (5-fold) enrichment of H3.3 in the (TTAGGG)_n repeat that is the conserved telomeric DNA sequence among vertebrates (Meyne et al., 1989) (**Figure 2.16**). This telomeric enrichment of H3.3 is consistent with the telomeric foci of incorporation visible on the largely heterochromatic Y-chromosome (**Figure 2.8A-B, 2.9B-C, K-L, 2.10**, inset). Mutation of H3.3B to H3.2 or H3.1S31 demonstrates that repeat-specific patterns of H3.3 enrichment or depletion are dependent on amino acid sequence (**Figure 2.16**).

H3.3 is enriched around transcription start sites of both active and repressed genes, but is enriched in the bodies of only active protein-coding and non-coding RNA genes

As there has been some disagreement regarding the patterns of H3.3 at active versus repressed genes in vertebrates (Chow et al., 2005; Daury et al., 2006; Jin and Felsenfeld, 2006; Sutcliffe et al., 2009; Tamura et al., 2009; Wirbelauer et al., 2005), we used ChIP-seq to address the genome-wide patterns of H3 variant enrichment around gene transcription start sites (TSS) in mouse ES cells. Surprisingly, we find that H3.3 is not exclusively a marker of transcriptionally active genes.

Mammalian genes can be divided into classes based on the CpG dinucleotide content of their promoters. High CpG content promoters (HCP) are over-represented in nuclear, metabolic, and transcriptional regulatory genes, and HCP genes are likely to be either ubiquitously expressed housekeeping genes or to have complex, developmentally regulated patterns of expression (Mikkelsen et al., 2007; Saxonov et al., 2006). In contrast, genes with low CpG content promoters (LCP) are frequently involved in specific physiological processes and

are often highly tissue-specific, such as olfactory receptor genes and keratin genes (Saxonov et al., 2006). Previous studies have found that the majority of high CpG content promoters in both ES cells and differentiated cells are marked by histone H3K4me3 and the presence of RNAPII, regardless of whether the gene is active or repressed (Barski et al., 2007; Guenther et al., 2007; Mikkelsen et al., 2007). When we divide HCP genes into low, medium, and high expression in ES cells (see **Materials and Methods**), we find that H3.3, like H3K4me3 and RNAPII, is enriched around the TSS of both active, highly expressing genes, as well as repressed, low expressing genes (**Figure 2.17A, D**). In both native and crosslinking ChIP-seq, we find H3.3 less enriched at the TSS itself (**Figure 2.17A, D, 2.18**). This depletion of



Data produced in collaboration with Deyou Zheng.

Figure 2.17: H3.3 is enriched around transcription start sites of both active and inactive genes with high CpG content promoters, and into the body of active genes

A-I. Profiles of H3 variants, H3, H3 modifications, or RNAPII as indicated above each panel across the TSS for highly active (red), medium expressing (black), or low expressing (green) CpG rich genes. Y-axis represents the average number of tags per gene per 200 bp per 1 million mapped reads. Data for H3 in ES cells is from (Mikkelsen et al., 2007).

Figure 2.18: Mononucleosome resolution profiles of H3 variants and H3K4me3 at TSS

Y-axis represents the average number of tags per gene per 10 bp per million mapped reads. Red represents sequence tags from the sense strand, and green represents sequence tags from the antisense strand. The color bars above profiles mark the estimated nucleosomes around TSSs. The data for H3.3 in panel G (dark blue = sense H3.3, light blue = antisense H3.3, red = sense H3K4me3, green = antisense H3K4me3) are scaled by a factor of 50 to match the profile of H3K4me3.

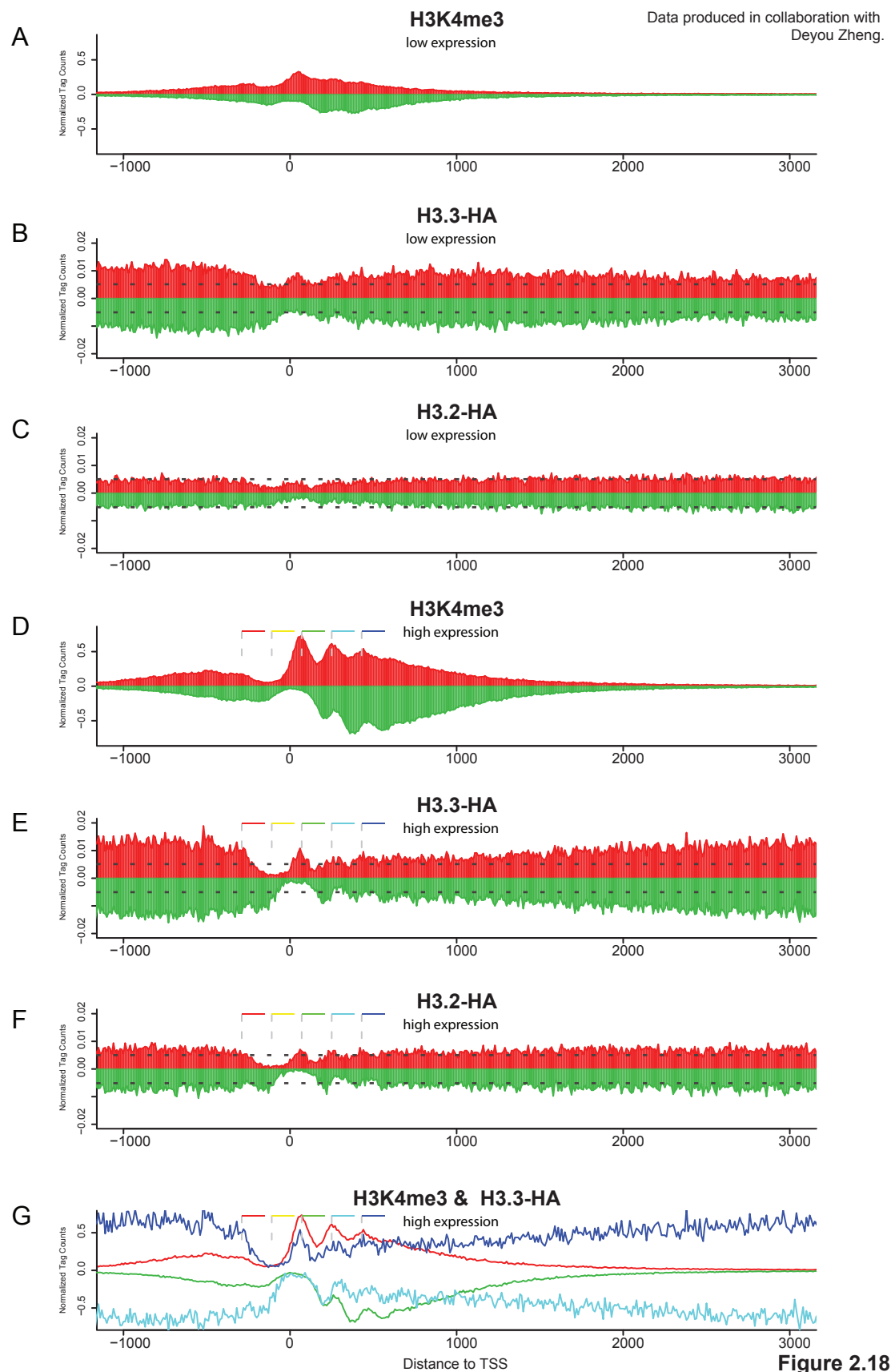


Figure 2.18

the -1 nucleosome at the TSS of active and inactive genes has been extensively described (Schones et al., 2008), and has been recently attributed both to the instability of nucleosome core particles containing H3.3 and H2A.Z, and to the presence of specific sequences at CpG promoters that directly reduce nucleosome stability (Jin et al., 2009; Ramirez-Carrozzi et al., 2009).

While H3.3 is incorporated around the TSS of both active and repressed genes, we find that H3.3 is enriched in the body of active genes, but not that of repressed genes (**Figure 2.17A, D**). Mononucleosome resolution analysis indicates that H3.3 is incorporated into the +1 nucleosomes in both activated and repressed genes, but up to +3 nucleosomes and further into the coding regions in activated genes (**Figure 2.18**). The level of H3.3 deposition in gene bodies is correlated with gene expression, particularly at highly expressed genes (Spearman's rank correlation coefficient $\rho = 0.54$, $p < 2.2e-16$; see **Figure 2.17A, D, 2.19**). Highly active genes (**Figure 2.17A, D**, red) have higher levels of H3.3 in their gene bodies, while medium (black) and lower expressed genes (green) have lower levels of H3.3. Strikingly, upon mutation of H3.3 to H3.2 or H3.1S31, H3.3 specific patterns of enrichment around the TSS and gene body are lost, generating patterns similar to general H3 (**Figure 2.17B, C, E, F**).

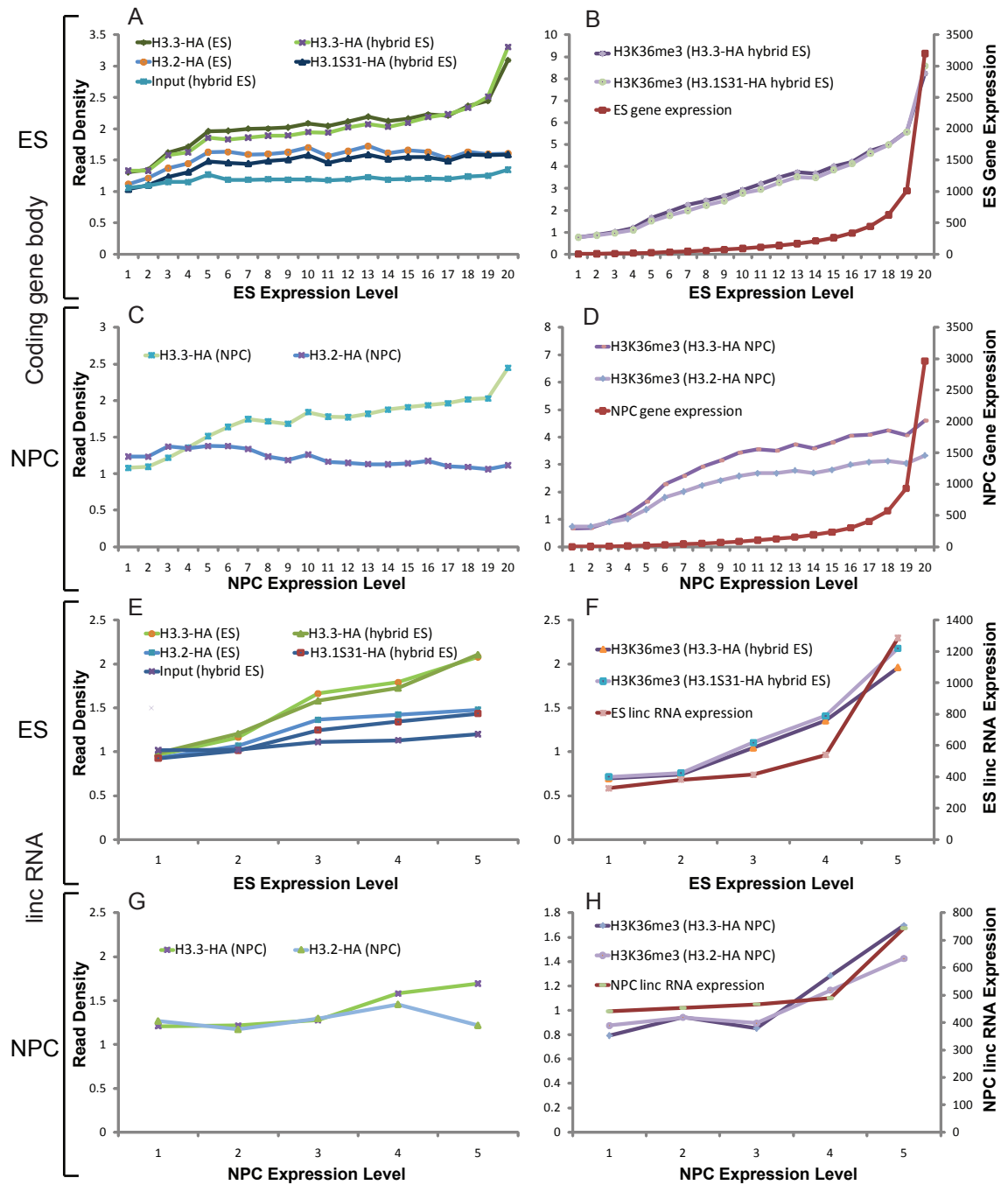
Genome-wide studies recently demonstrated that more than 20% of HCP genes in ES cells carry both H3K4me3 and H3K27me3 in their promoters (Bernstein et al., 2006a; Mikkelsen et al., 2007). These 'bivalent' genes are transcriptionally repressed in ES cells, and many encode key developmental TFs and cell surface proteins, leading to the model that they represent repressed genes that are poised for activation following cell differentiation (Bernstein et al., 2006a; Mikkelsen et al., 2007). While bivalent genes are defined by their pattern of H3 modifications, the relative distribution and identity of the H3 variants that carry these marks has not been established. When we analyzed the pattern of H3 variants at bivalent

Figure 2.19: Correlation of H3.3 with gene expression

The left Y-axis represents read density that was computed as average number of ChIP-Seq reads per 2-kb genomic sequence per million mapped reads.

A-D. Densities of H3.3 and H3K36me3 reads in coding gene bodies are correlated with gene expression, particularly for highly expressed genes, in both ES (A-B) and NPC (C-D). Gene expression data are from (Mikkelsen et al., 2007). For the X-axis, gene expression values were uniformly divided into 20 groups to yield genes with low to high levels of expression. Mean gene expression values for each group shown on the right Y-axis in B and D. H3.3 deposition in gene bodies is correlated with gene expression, particularly at highly expressed genes (Spearman's rank correlation coefficient $\rho = 0.54$, $p < 2.2e-16$). By comparison, H3K36me3 shows a greater correlation ($\rho = 0.76$), but H3.1S31 ($\rho = 0.31$) and H3.2 ($\rho = 0.25$) show significantly smaller correlations.

E-H. Densities of H3.3 and H3K36me3 in lincRNAs are correlated with gene expression at highly expressed lincRNAs. lincRNAs were divided into 5 groups on the basis of lincRNA expression data (Guttman et al., 2009), downloaded from <http://www.ncbi.nlm.nih.gov/geo/>, (GSE13765). lincRNA expression values are shown on the right Y-axis in F and H.



Data produced in collaboration with Deyou Zheng.

Figure 2.19

TSS by ChIP-seq, we found that H3.3 is enriched around the TSS of H3K4me3/H3K27me3 bivalent genes in ES cells, while mutation of H3.3 towards H3.2 or H3.1 abolishes this enrichment (**Figure 2.20A, 2.21E-F**). In contrast to repressed bivalent H3K4me3/H3K27me3 genes, H3.3 is significantly enriched around the TSS as well as into the coding region of active monovalent H3K4me3 genes (**Figure 2.20B**).

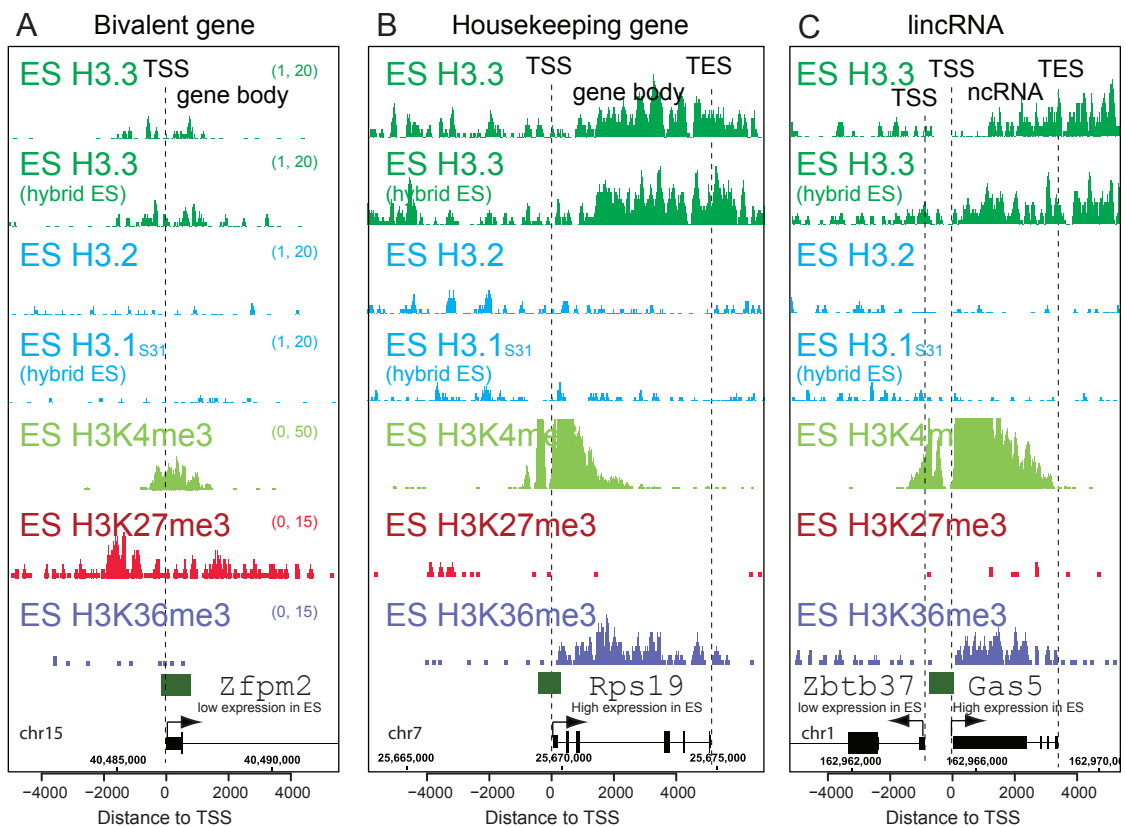


Figure 2.20: Specific examples of H3.3 incorporation at the transcription start sites of active and repressed coding genes, and actively transcribed non-coding RNA

A. H3.3 is enriched around the TSS, but not into the gene body of the H3K4me3/H3K27me3 bivalent and transcriptionally repressed *Zfp202* gene in ES cells.

B. H3.3 is enriched around the TSS and also in the gene body of the highly expressing ribosomal protein S19 gene *Rps19*.

C. H3.3 is enriched around the TSS and into the body of the non-coding RNA gene *Gas5*, but less enriched in the gene body of the neighboring low expressing *Zbtb37* gene in ES cells. The Y-axes in J-L are identical (indicated in the right side of panel J), and represent the number of reads spanning a genomic position. Gene is shown to scale below plot, with start and direction of transcription indicated by arrow. Green rectangles above the gene represent CpG islands.

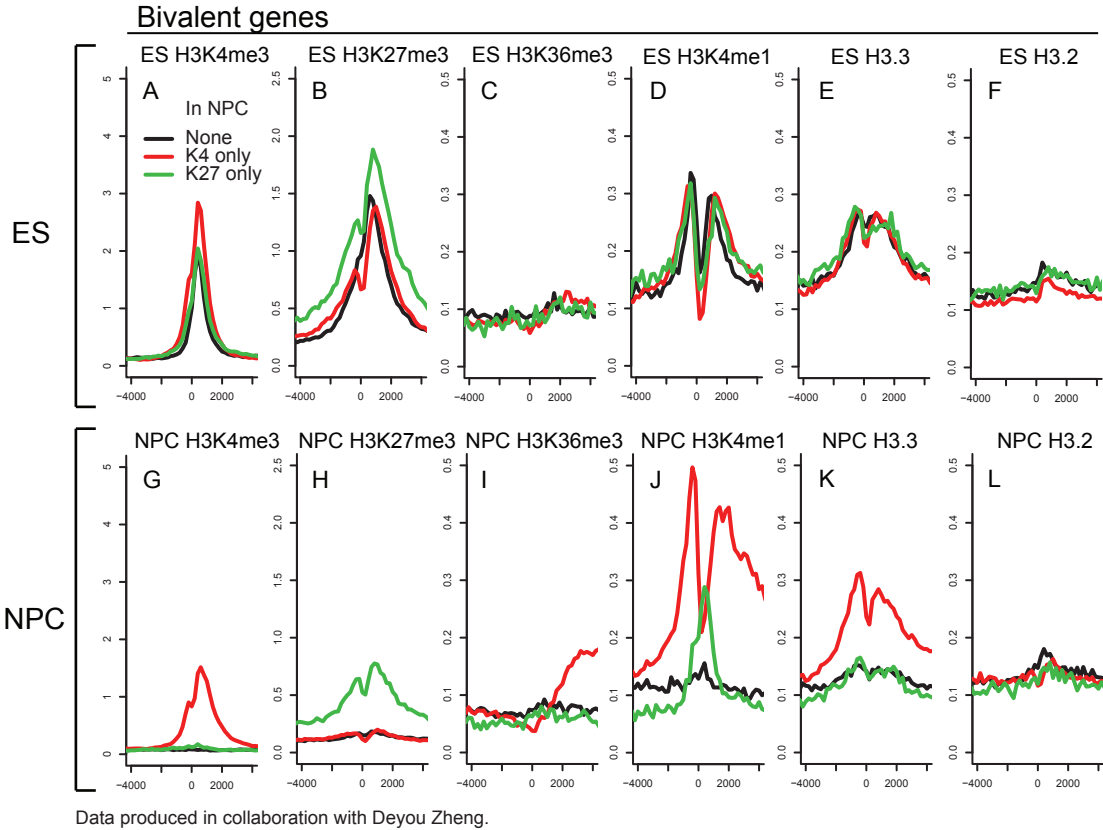


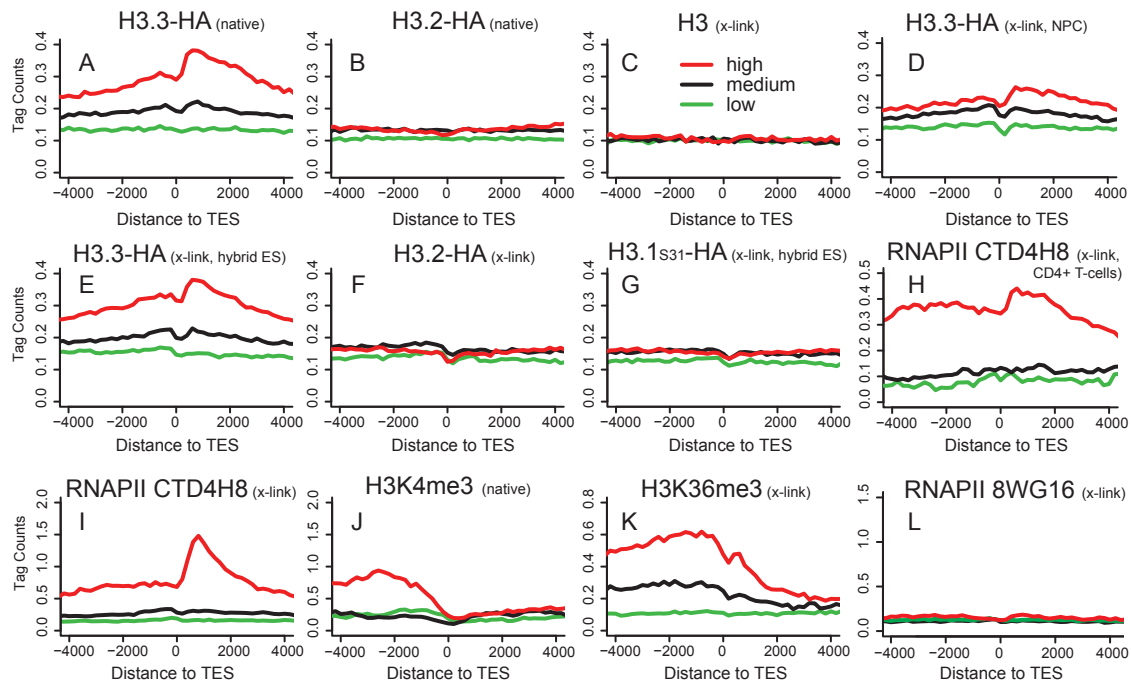
Figure 2.21: Genome-wide profiles of H3.3 at transcription start sites change with cellular differentiation and resolution of bivalent domains

A-L. Profiles of H3 variants or H3 modifications as indicated above each panel around the TSS of H3K4me3/H3K27me3 bivalent genes in ES cells. ES bivalent genes are classified by resolution in NPCs to H3K4me3 only (red), H3K27me3 only (green), or no mark (black). H3K4me3/H3K27me3 data and bivalent gene definitions are from (Mikkelsen et al., 2007). Y-axes are as in **2.17**.

Recent studies have also shown that mammalian cells produce over 1600 large intergenic non-coding RNA (lincRNA) transcripts (Guttman et al., 2009). When we examined the genomic loci of lincRNA transcripts, we found that lincRNAs known to be expressed in ES cells show enrichment of H3.3 around their TSS, as well as into their non-coding RNA gene bodies, similar to the pattern of H3.3 found at active genes. Again, these patterns of H3.3 at lincRNAs are dependent on H3.3 amino acid sequence (**Figure 2.19, 2.20C**).

Transcription end sites of highly expressed genes are marked by peaks of H3.3 and phosphorylated RNA polymerase II

When we examined the patterns of H3.3 incorporation at transcriptionally active HCP genes, we noticed that enrichment of H3.3 often extended beyond the gene body and past the transcriptional end site (TES), particularly at highly active genes (**Figure 2.20B-C**). To determine whether this is a genome-wide phenomenon, we used the same division of HCP genes into high, medium, and low expression, and analyzed the ChIP-seq patterns of H3 variants and histone modifications around TES. In accordance with previous studies in *Drosophila* and human cells (Henikoff et al., 2009; Jin et al., 2009; Mito et al., 2005), we find that the enrichment of H3.3 beyond the TES is a general phenomenon at highly expressed genes in ES cells



Data produced in collaboration with Deyou Zheng.

Figure 2.22: H3.3 and RNA polymerase Ser5p are enriched after the transcriptional end site of highly expressed genes

A-L. Profiles of H3 variants, H3, H3 modifications, or RNAPII as indicated above each panel across the TES for highly active (red), medium expressing (black), or low expressing (green) CpG rich genes, with data represented as in Figure 3. All data except for D and H are for mouse ES cells. Panel D shows data from differentiated NPCs. The data in panel H are for human CD4+ T cells (Barski et al., 2007). The data in C and L are from a previous study (Mikkelsen et al., 2007).

(Figure 2.22A, E) and in differentiated NPCs (Figure 2.22D).

At highly expressed genes such as beta-actin, H3.3 enrichment increases immediately after the TES, and peaks at approximately +700 bp \pm 200 bp after the TES (Figure 2.22A, E, 2.23). This post-TES enrichment of H3.3 was also dependent on amino acid sequence, as we did not observe significant enrichment of H3.3 mutated to H3.2 or H3.1S31, or of general H3 (Figure 2.22B, C, F, G, 2.23). Control ChIP-seq experiments revealed the expected patterns of incorporation and lack of post-TES enrichment for H3K4me3, H3K4me1, and H3K36me3, in

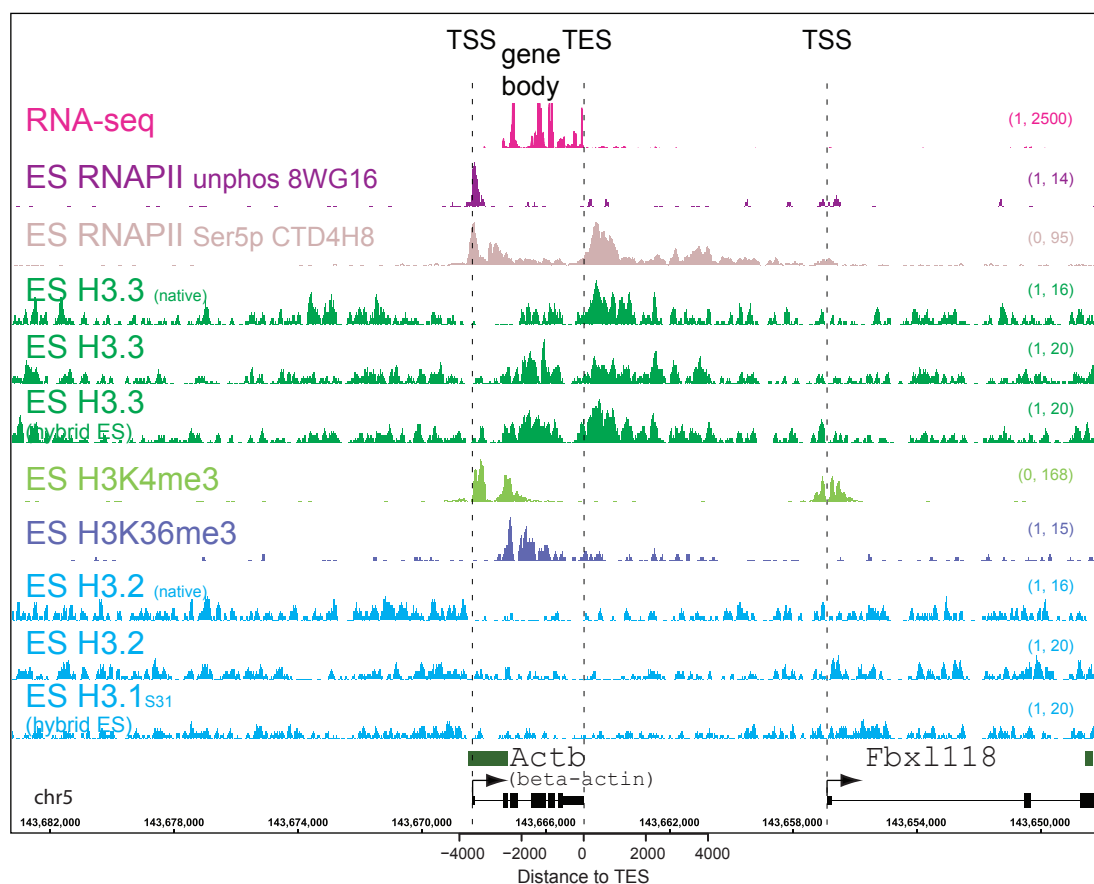


Figure 2.23: Profiles of H3 variants, H3 modifications, and RNA polymerase around the highly expressed housekeeping gene beta-actin

H3.3 and RNAPII Ser5p are enriched in the gene body and after the TES of the highly expressed gene *Actb* (beta-actin) in ES cells. Data for the RNAPII unphos 8WG16 track are from (Mikkelsen et al., 2007). Y-axis is indicated on the right as in Figure 2 and 3. RNA-seq data are from (Cloonan et al., 2008). The neighboring *Fbx1118* gene is not highly expressed in ES cells, and does not show gene body enrichment of H3.3.

Figure 2.24: TSS and TES enrichment of H3 variants, histone modifications, and RNAPII

Profiles of H3 variants, H3 modifications, or controls as indicated above each panel across the TSS and TES for highly active, medium expressing, or low expressing CpG rich genes as shown. Y-axis represents the average number of tags per gene per 200 bp per 1 million mapped reads.

A,C,E,G. H3.3 profiles are similar across the TSS and TES for both HA and EYFP-tagged H3.3.

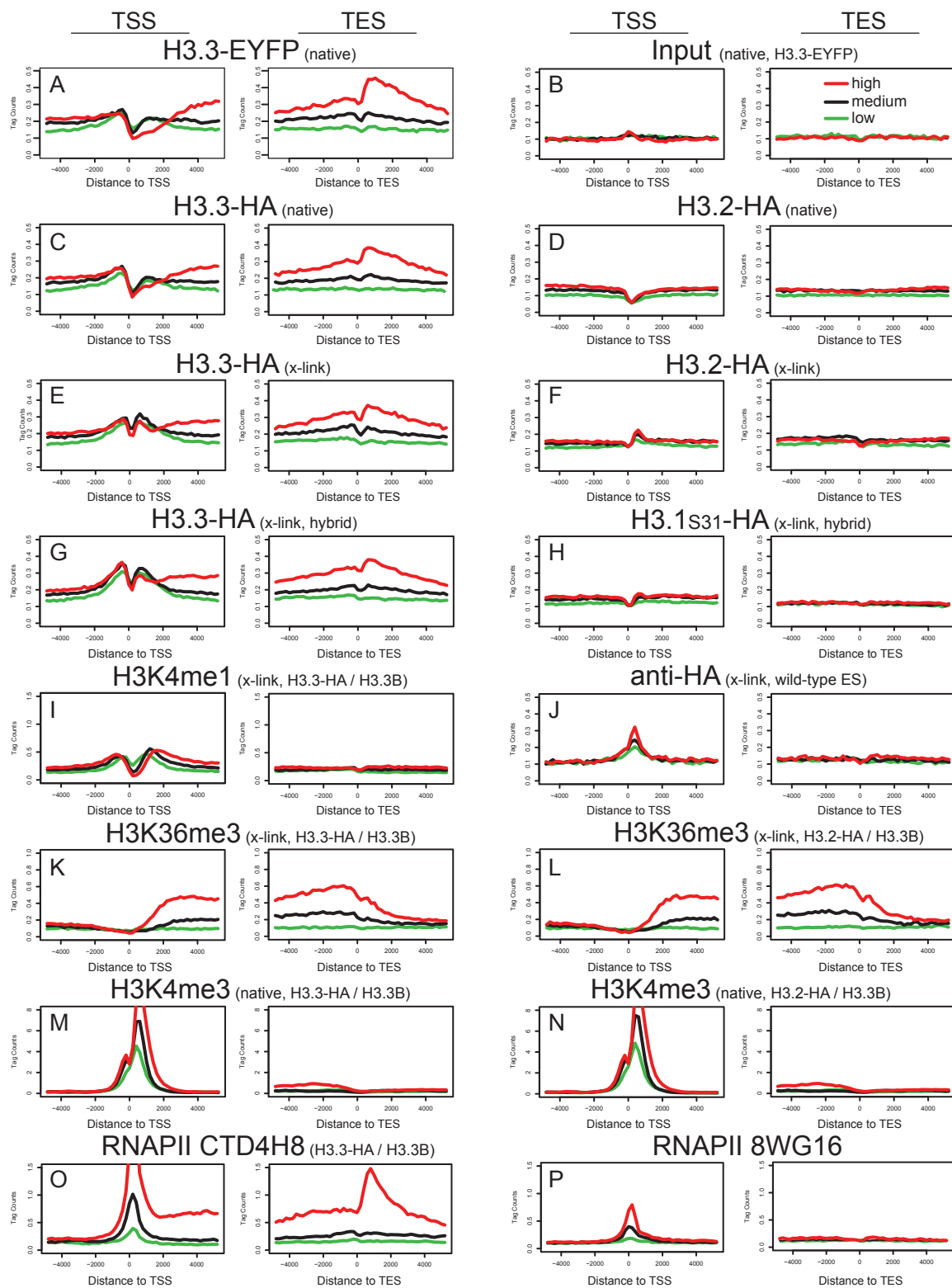
D,F,H. Mutation of H3.3B towards H3.2 or H3.1 abolishes H3.3 patterns of enrichment.

I. H3K4me1 patterns are similar to H3.3 around the TSS, though levels of H3K4me1 decline into the gene body of highly expressed genes, while H3.3 levels increase and peak beyond the TES.

J. Background of anti-HA antibody. Sequencing of anti-HA ChIP from untagged wild-type ES cells shows enrichment at the TSS, likely representing the enrichment of nucleosome-poor DNA following cross-linking and sonication.

K-N. Profiles of H3K4me3 and H3K36me3 in H3.3-HA and H3.2-HA ES cells. Heterozygous mutation of H3.3 to H3.2 does not alter genome-wide profiles of H3K4me3 or H3K36me3 around TSS and TES.

O. Profiles of RNAPII Ser5p in H3.3-HA ES cells at TSS and TES. **P)** Profiles of unphosphorylated RNAPII at TSS and TES in ES cells. Data for P are from (Mikkelsen et al., 2007).



Data produced in collaboration with Deyou Zheng.

Figure 2.24

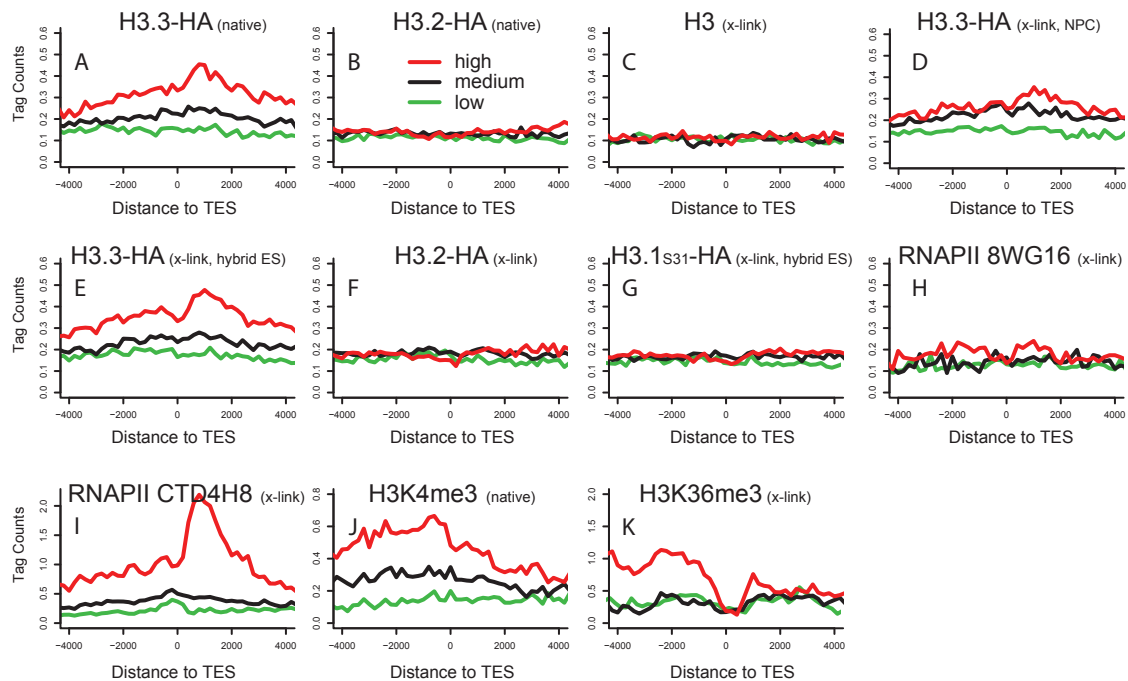
both H3.3-HA and H3.2-HA ES cells (**Figure 2.22J, K, 2.23, 2.24**). H3K36me3 has been associated with transcriptional elongation in a variety of eukaryotes, from yeast to mammals (Bannister et al., 2005; Barski et al., 2007; Mikkelsen et al., 2007). H3.3 and H3K36me3 are both enriched in the gene body of active genes, with levels of both positively correlated with gene expression (**Figure 2.17A, D, I, Figure 2.19, Figure 2.22A, E, K**). Unlike H3.3, however, levels of H3K36me3 at active genes begin to decline prior to the TES, and decline more significantly following the TES (**Figure 2.22K, 2.23**). While H3.3 is correlated with H3K36me3 in active gene bodies, H3.3 and H3K36me3 maintain distinctive patterns near the TES of active genes.

As a control, we performed ChIP-seq with a mouse monoclonal antibody (CTD4H8) to the Ser-5 phosphorylated C-terminal domain (CTD) of RNAPII. To our surprise, we observed significant peaks of enrichment of phosphorylated RNAPII not only at the TSS of high and low expressing genes (**Figure 2.17G**), but also after the TES of high expressing genes (**Figure 2.22I, 2.23**). Interestingly, the distributions of these post-TES peaks of phosphorylated RNAPII at active genes are closely co-localized with the peaks of H3.3, with the phosphorylated RNAPII peak slightly downstream (~200 bp) from the H3.3 peak.

Previous studies have shown that the unphosphorylated CTD form of RNAPII is enriched in the promoter of both active and inactive genes, while differentially phosphorylated CTD forms of RNAPII are associated with elongating polymerase in the gene body of actively transcribed protein coding genes and small nuclear RNA genes (Guenther et al., 2007; Koch et al., 2008). When we analyzed the patterns of unphosphorylated RNAPII CTD (Mikkelsen et al., 2007) in mouse ES cells, we observed a peak of unphosphorylated RNAPII CTD at the TSS of high and low expressing HCP genes (**Figure 2.23, 2.24**). However, we did not find any significant peak of unphosphorylated RNAPII CTD around the TES of high or low

expressing HCP genes (**Figure 2.21L, 2.22**), providing further evidence that this post-TES enrichment of RNAPII is specific to the phosphorylated form of RNAPII.

To determine if the enrichment of H3.3 and the phosphorylated CTD of RNAPII is specific to ES cells, we also analyzed their distributions in other cell types. Our results indicate that post-TES enrichment of H3.3 and phosphorylated RNAPII CTD is not specific to ES cells. When we generated profiles of H3.3 at high, medium, and low-expressing genes in differentiated NPCs, we observe that H3.3 enrichment at highly active genes continues after the TES (**Figure 2.22D**). While it has not previously been used in ChIP-seq studies of ES cells, the monoclonal CTD4H8 antibody has also been used in genome-wide studies in human CD4+T cells (Barski et al., 2007). When we analyze these datasets, we observe a similar enrichment of phosphorylated RNAPII beyond the TES of highly active genes in



Data produced in collaboration with Deyou Zheng.

Figure 2.25: Analysis of H3.3 post-TES at HCP genes without repeats in their TES

HCP genes without repeats (as annotated by the RepeatMasker in the UCSC table browser data) in their TES regions (-500 bp to +2kb) were separated into three groups based on their expression values, and then the distribution of ChIP-Seq reads around TSSs for these three groups of genes were generated as described above in **2.23**.

differentiated human CD4⁺ T cells (**Figure 2.22H**).

To ensure that the extension of phosphorylated RNAPII and H3.3 did not represent a previously unrecognized extension of transcription beyond the TES of high-expressing genes, we analyzed a published RNA-seq dataset from mouse ES cells (Cloonan et al., 2008). We observed peaks of transcribed RNA in exons, as expected, and did not find significant extension of transcription beyond the TES of active genes such beta-actin (**Figure 2.23**). We were also concerned that the H3.3 peak downstream of TESs could be due to the presence of Alu or other active repetitive elements in some highly expressed genes. However, this post-TES peak remained prominent when we focused our analysis on genes without repeats in their TES regions (-500 bp to + 2kb) (**Figure 2.25**). From these analyses, we conclude that both H3.3 and phosphorylated RNAPII are significantly enriched beyond the TES of highly expressed genes in undifferentiated and differentiated mammalian cell types.

The profile of H3.3 at cell-type specific genes changes with cell differentiation

To determine how the genome-wide patterns of H3.3 change with cell differentiation, we used an established protocol to differentiate both H3.3-HA and H3.2-HA ES cells to NPCs (Conti et al., 2005). To confirm that both H3.3-HA and H3.2-HA NPCs were similar in their genome-wide profiles of histone modifications to previously analyzed NPCs (Meissner et al., 2008; Mikkelsen et al., 2007), we performed control H3K36me3 ChIP-seq experiments from both H3.3-HA and H3.2-HA NPCs and found profiles that were extremely similar to established datasets (**Figure 2.14, Figure 2.15**).

Our results demonstrate that the profile of H3 variants changes in undifferentiated ES cells and differentiated NPCs at cell-type specific genes. In ES

cells, H3.3 was enriched in the gene body of expressed pluripotency genes such as *Esrrb*, *Nanog*, and *Oct4*, correlated with H3K36me3 (**Figure 2.26A, 2.27**). As with other high expressing genes, enrichment of H3.3 in the gene body of pluripotency genes is dependent on amino acid sequence, as mutation to H3.2 or H3.1S31 abolishes this enrichment (**Figure 2.26A, 2.27**). Upon differentiation of ES cells to NPCs, the expression of most pluripotency genes is lost (Conti et al., 2005). Accordingly, the enrichment of H3.3 and H3K36me3 in the bodies of pluripotency

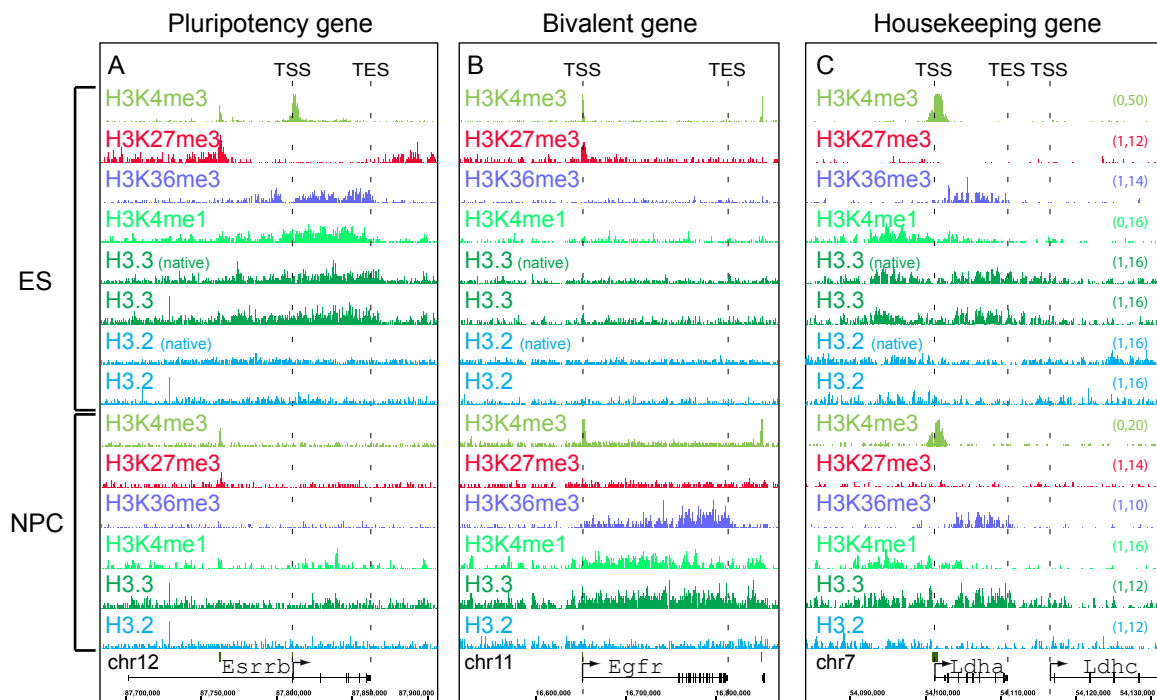


Figure 2.26: H3.3 localization changes at cell-type specific genes with cell differentiation and resolution of bivalent domains

A. H3.3, H3K4me1, and H3K36me3 are enriched in the gene body of the highly expressed pluripotency gene *Esrrb* in ES cells, and this enrichment is largely lost upon differentiation to NPCs.

B. The epidermal growth factor receptor gene *Egfr* is H3K4me3/H3K27me3 bivalent and transcriptionally repressed in ES cells (Mikkelsen et al. 2007). Upon differentiation to NPCs, H3.3, H3K4me1, and H3K36me3 are enriched in the *Egfr* gene body.

C. H3.3, H3K4me1, and H3K36me3 are enriched in the gene body of the highly expressed housekeeping gene lactate dehydrogenase A *Ldha* in both ES cells and NPCs, and are not enriched in the low expressed neighboring *Ldhc* gene. The Y-axes in panels A-C are identical (indicated in the right side of panel C).

genes largely disappeared upon cell differentiation (**Figure 2.26A, 2.27**).

We next sought to examine how the pattern of H3 variants changes at H3K4me3/H3K27me3 bivalent genes (Bernstein et al., 2006a) that are transcriptionally repressed in ES cells, but poised for expression in later developmental lineages (see **Figure 1.14**). Analysis of 2,525 HCP genes defined as bivalent in mouse ES cells (Mikkelsen et al., 2007) demonstrated that H3.3 is enriched around the TSS, but not within the gene body of bivalent differentiation-specific genes in ES cells (**Figure 2.20A, 2.21E**). Again, this enrichment of H3.3 around the TSS of bivalent ES genes is dependent on amino acid sequence, as mutation of H3.3 to H3.2 eliminated TSS enrichment at ES bivalent genes (**Figure 2.21F**).

Following differentiation of ES cells to NPCs, the profile of H3.3 changes at

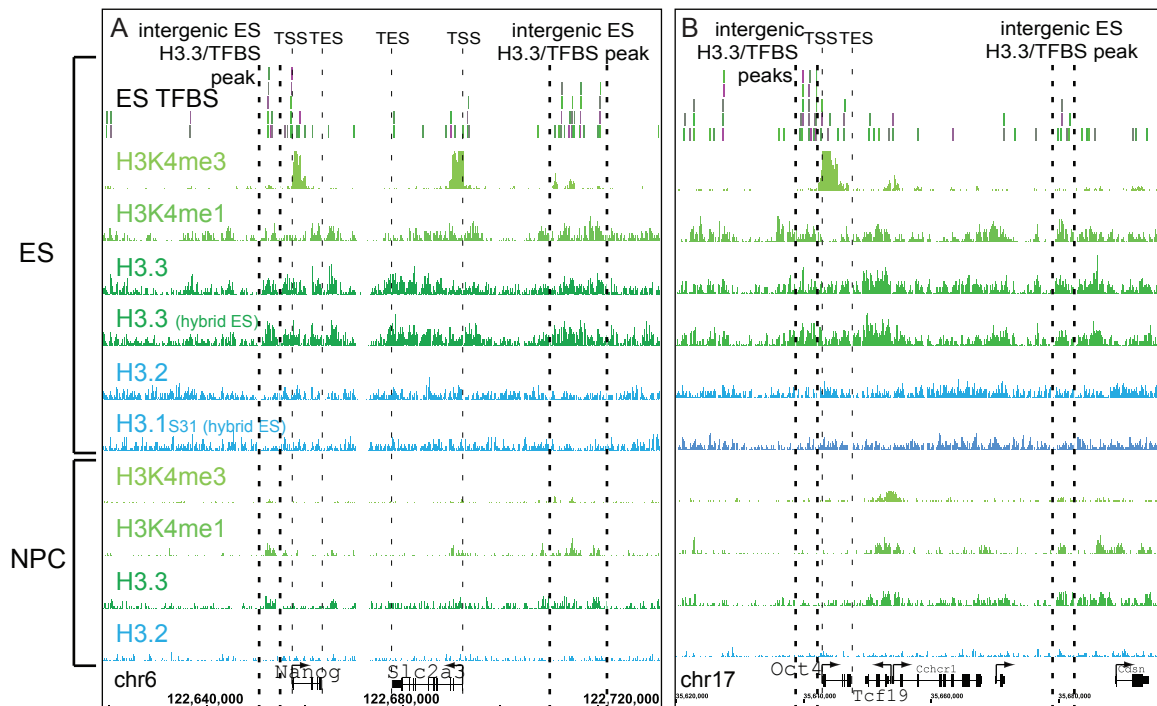
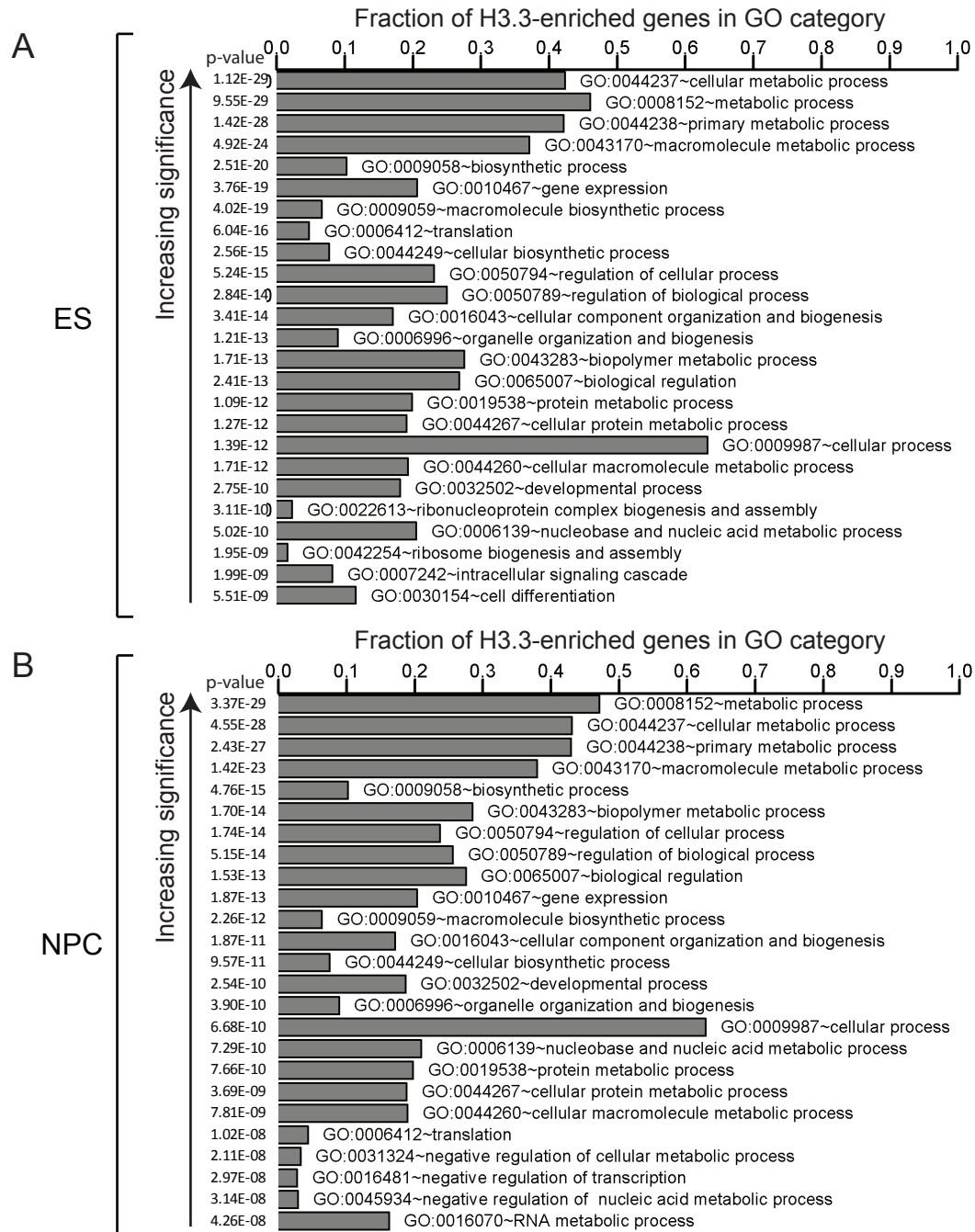


Figure 2.27: Cell-type specific enrichment of H3.3 at pluripotency genes and transcription factor binding sites

A. H3.3 and H3K4me1 are enriched around intergenic and genic TFBS near the Nanog gene in ES cells. In all cases, H3.3-specific enrichment is lost upon mutation of H3.3B to H3.2 or H3.1S31.
B. H3.3 and H3K4me1 are enriched around intergenic and genic TFBS near the Oct4 (Pou5f1) gene in ES cells.

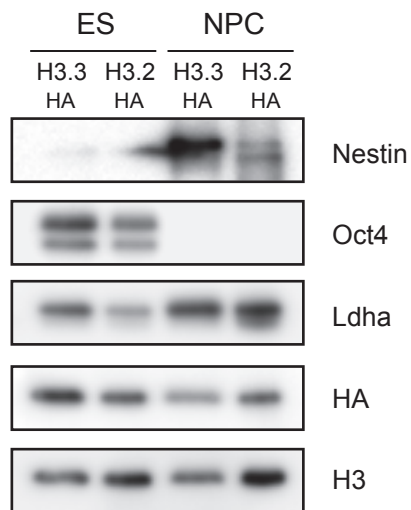


Produced in collaboration with Deyou Zheng.

Figure 2.28: H3.3 is enriched at genes involved in cellular and metabolic processes in both ES cells and NPC cells

Genes with H3.3 peaks in their promoter regions were analyzed for enrichment of individual GO terms using the program DAVID (Huang da et al., 2009). The top 25 GO categories are shown here with p-values indicated, and the x-axis representing the fraction of H3.3 genes in a GO category.

poised TSS with the resolution of bivalent domains. 92% of H3K4me3/H3K27me3 bivalent genes in ES cells have been shown to resolve to either H3K4me3 only (**Figure 2.21**, red), H3K27me3 only (green), or lose both modifications (black) upon differentiation to NPCs (Mikkelsen et al., 2007). In bivalent ES genes that resolve to H3K4me3 and become transcriptionally active in NPCs, H3.3 was maintained around the TSS and also incorporated into the gene body, in correlation with H3K36me3 and H3K4me1 (red, **Figure 2.21G, I-K**). For example, the epidermal growth factor receptor gene *Egfr* is bivalent and transcriptionally repressed in ES cells, yet resolves to H3K4me3 and is expressed in NPCs. Upon differentiation to NPCs, H3.3 extended into the gene body of *Egfr*, correlating with H3K36me3 and



Data produced in collaboration with Laura Banaszynski.

Figure 2.29: Expression of pluripotency and differentiation-specific genes in H3.3-HA and H3.2-HA ES cells and NPCs

Whole cell extracts of H3.3-HA and H3.2-HA ES cells and NPCs were immunoblotted with antibodies as indicated.

H3K4me1 (**Figure 2.26B**). In contrast, for bivalent genes that remain transcriptionally repressed and resolve to either H3K27me3 or no mark in NPCs, H3.3 enrichment was reduced at the TSS upon cell differentiation to NPCs (**Figure 2.21K**). As expected, H3.2 remains unenriched at ES bivalent TSS following differentiation to NPCs (**Figure 2.21L**).

While the pattern of H3.3 changes at cell-type-specific genes, housekeeping genes that remain highly expressed in both ES cells and NPCs retain similar enriched patterns of H3.3 incorporation. In both ES cells and NPCs, H3.3 is most highly enriched in genes associated with cellular and metabolic processes (**Figure 2.28**).

For example, the housekeeping gene lactate dehydrogenase A *Ldha* is expressed in both ES and NPCs (**Figure 2.29**). In contrast to the cell-type specific genes

Essrb and *Egfr*, H3.3 remains enriched around the TSS and within the gene body of *Ldha* in both ES cells and NPCs (**Figure 2.26C**). Our data show that the overall relationship between H3.3 incorporation and gene expression is retained in ES and NPCs, with H3.3 and H3K4 methylation marking the TSS of active and poised HCP genes, and H3.3 enriched in the body and TES of active HCP genes. However, at cell-type specific genes, H3.3 localization changes with cellular state.

Genome-wide enrichment of H3.3 at transcription factor binding sites is dependent on amino acid sequence and cellular differentiation state

To determine enriched regions of H3 variants genome-wide, we used a modified segmentation algorithm (see **Chapter Five**). Unexpectedly, we found that 36% (3,627) of the enriched regions of H3.3 in mouse ES cells were located outside of known annotated genes (**Figure 2.13, 2.27, 2.30**). Genome-scale studies have previously demonstrated that in addition to genes (Mito et al., 2005), H3.3 replacement marks the boundaries of cis-regulatory elements in *Drosophila* (Mito et al., 2007). Therefore, we considered the possibility that H3.3 might mark genic and intergenic TF binding sites in mammalian cells genome-wide.

Significant progress has been made regarding the core transcriptional regulatory networks that regulate pluripotency in mammalian cells. We therefore compared our genome-wide distributions for histone H3 variants in ES cells and NPCs with established genome-wide maps for 13 distinct sequence-specific TFs in ES cells (Chen et al., 2008b). Remarkably, 3,878 (38%) of the overall regions enriched for H3.3 in ES cells (**Figure 2.13**) correspond to previously identified genic or intergenic TF binding sites (TFBS) (Chen et al., 2008b). As with specific patterns of H3.3 at genes, peaks of H3.3 at TFBS are dependent on H3.3 amino acid sequence, as mutation of H3.3B to H3.2 or H3.1S31 eliminated these peaks of enrichment (**Figure 2.27, 2.30, 2.31**).

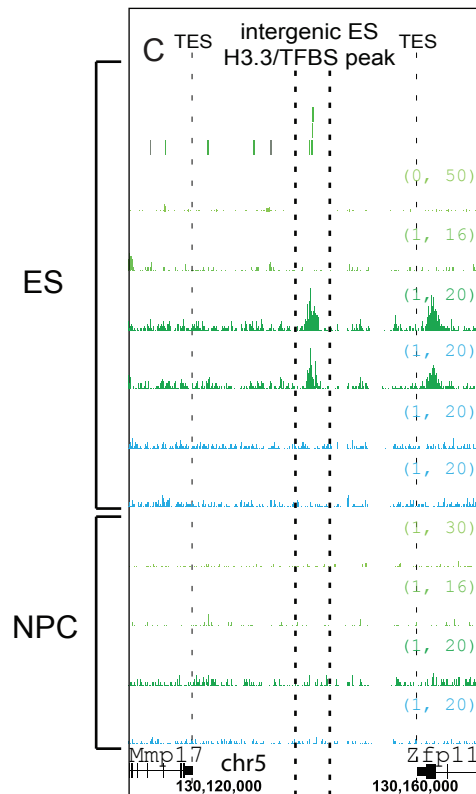
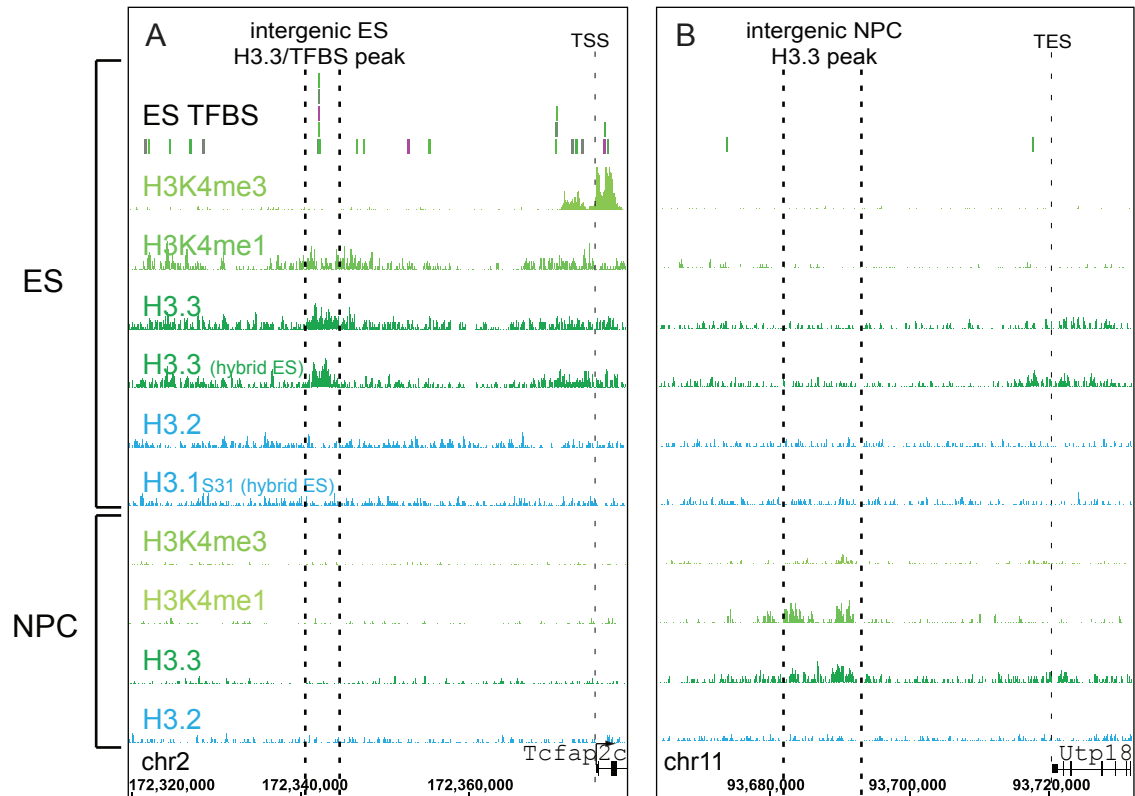


Figure 2.30: Cell-type specific enrichment of H3.3 at transcription factor binding sites

A. H3.3 and H3K4me1 are enriched at an intergenic region bound by multiple transcription factors (multiple transcription factor binding locus, MTL) (Chen et al., 2008) specifically in ES cells.

B. H3.3 and H3K4me1 are enriched at an intergenic region specifically in NPCs, which may represent an uncharacterized TFBS.

C. H3.3 is enriched at an intergenic MTL specifically in ES cells. In this case, H3K4me1 does not appear enriched at this location. The Y-axes in panels **A-C** are identical and indicated in the right side of panel **C**. TFBS identified in mouse ES cells by previous ChIP-seq studies (Chen et al., 2008) are indicated above the plots. NPC H3K4me3 data is from (Mikkelsen et al., 2007).

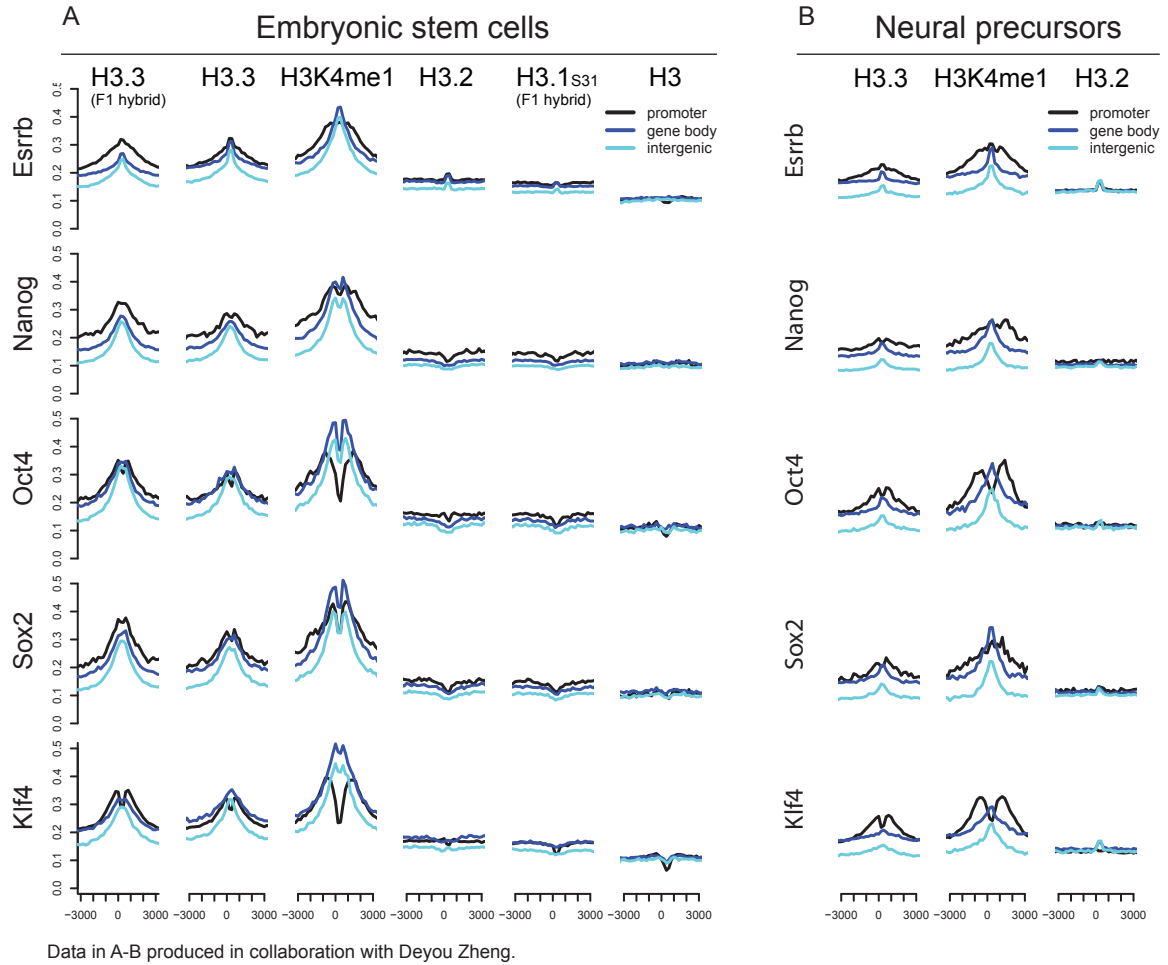


Figure 2.31: Genome-wide patterns of H3.3 at transcription factor binding sites are dependent on H3 amino acid sequence and cellular state

A-B. H3.3 and H3K4me1 are enriched genome-wide around ES cell TFBS. Genome-wide profiles of H3 variants (crosslinking ChIP-seq, HA tagged variants) and H3K4me1 as indicated around ES binding sites for Esrrb, Nanog, Oct4, Sox2, and Klf4. TFBS in mouse ES cells were identified by a previous study using ChIP-seq (Chen et al., 2008), and classified as promoter (black), gene body (dark blue), or intergenic (light blue). Panel A represents data from ES cells, while panel B represents data from NPCs at identical regions. Y-axis represents number of tags per binding site per 200 bp per 1 million mapped reads. Data for H3 in ES cells is from (Mikkelsen et al., 2007).

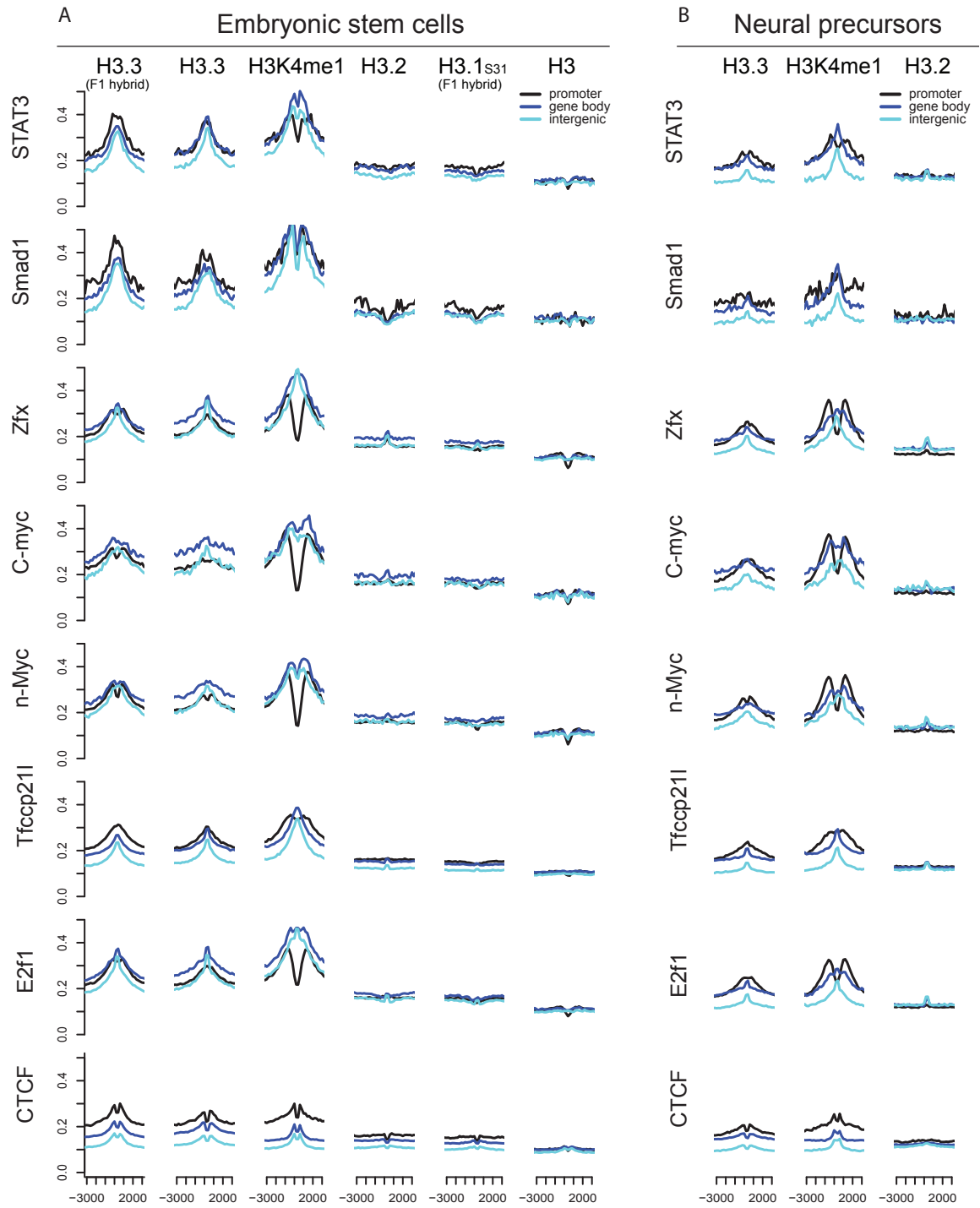
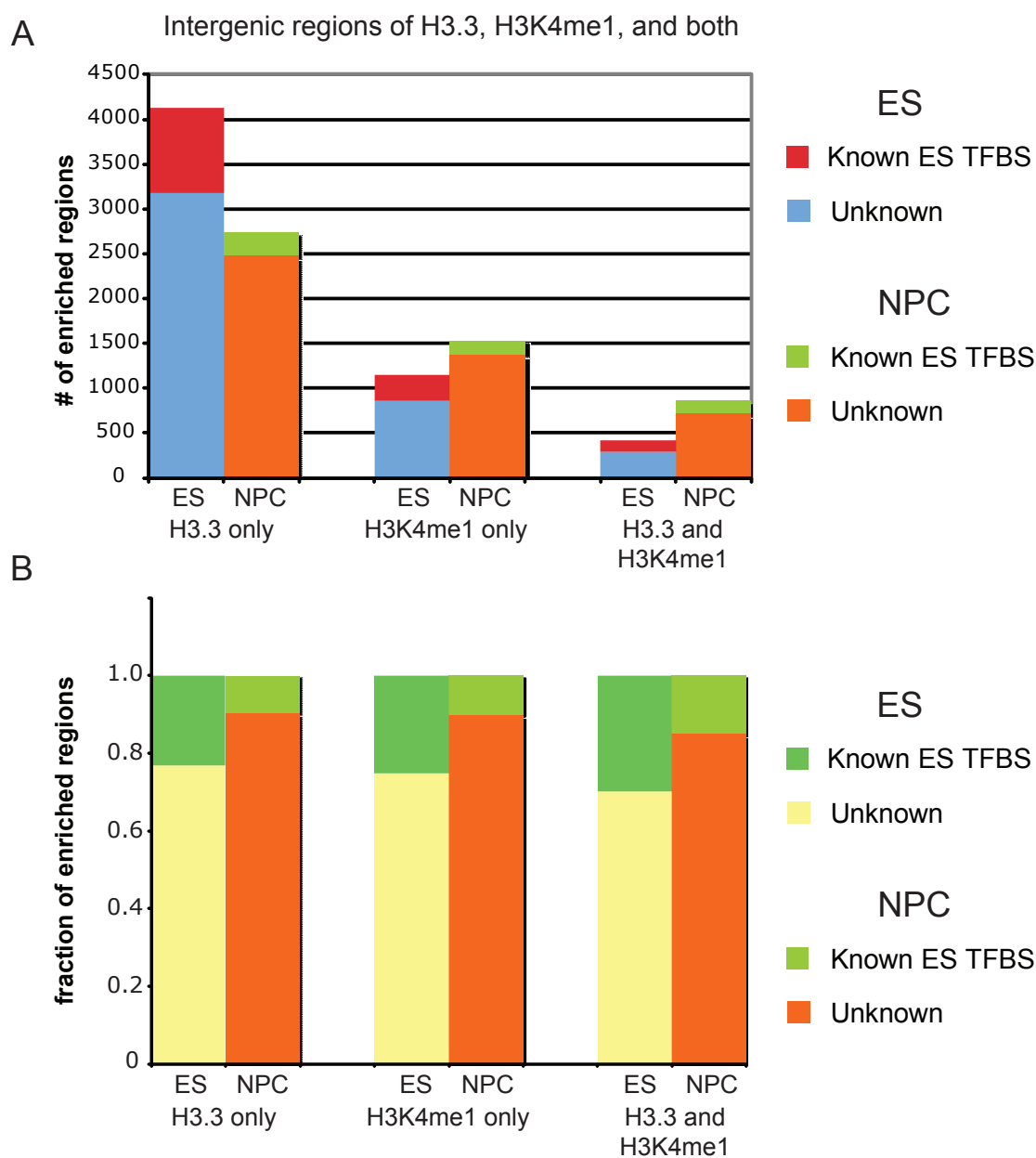


Figure 2.31

Many of the specific peaks of H3.3 at TFBS are cell-type specific, especially those located in intergenic regions. 976 intergenic peaks of H3.3 that co-localize with known binding sites for pluripotency-regulating TFs in ES cells are lost upon differentiation to NPCs (**Figure 2.30A, C**). For example, the locus for the pluripotency gene *Nanog* is highly regulated and is a target for multiple transcription factors in ES cells (Chen et al., 2008b). Patterns of H3.3 enrichment at the *Nanog* locus correspond with known TFBS in ES cells (Chen et al., 2008b), as well as previously identified DNaseI hypersensitive sites (Levasseur et al., 2008) (**Figure 2.27A**). These peaks of H3.3 incorporation around *Nanog* are significantly reduced following ES cell differentiation to NPCs (**Figure 2.27A**). In addition to ES cell specific peaks of H3.3 (**Figure 2.27, 2.30A, C**), we also found peaks of H3.3 that are specific to differentiated NPCs (**Figure 2.30B**). Intriguingly, 2478 (~90%) of the intergenic and NPC-specific H3.3 peaks do not overlap with known ES TFBS (Chen et al., 2008b) (**Figure 2.30B**).

Our data suggest that overlapping peaks of H3.3 and H3K4me1 may identify known and novel TFBS. H3K4me1 has recently been shown to have a strong genome-wide co-localization with TFBS, enhancers, and other regulatory elements (Heintzman et al., 2007; Robertson et al., 2008; Wang et al., 2008). Our analysis found that 30% of the intergenic peaks enriched with both H3.3 and H3K4me1 overlap with known TFBS (**Figure 2.30A, B, D, 2.31**), compared with 23.7% for all intergenic H3.3 peaks and 26.5% for all intergenic H3K4me1 peaks. Genome-wide patterns of H3.3 and H3K4me1 are similar (**Figure 2.15**). Indeed, we note that 26% of H3K4me1 peaks overlap with H3.3 peaks, and peaks of H3.3 are 5 times more likely to be associated with peaks of H3K4me1 than H3K4me3 ($p < 0.00001$). Notably, we found 724 intergenic peaks of both H3.3 and H3K4me1 in NPCs that do not correspond to any known TFBS (**Figure 2.30B, 2.32**), and it is possible that these represent uncharacterized, cell-type specific regulatory



Data in A-B produced in collaboration with Deyou Zheng.

Figure 2.32: Genome-wide changes in intergenic peaks of H3.3 and H3K4me1

H3.3 and H3K4me1 intergenic peaks were identified as described in Data Analysis. Data is for crosslinking ChIP-seq with H3.3-HA. Those intergenic peaks were then defined as 'known ES TFBS' if they overlap with an intergenic TFBS identified in ES cells (Chen et al., 2008), or 'unknown' if they do not overlap with known TFBS. **A** plots the total number of peaks in each category while **B** plots the relative fraction of peaks in each category. Note the relative increase in enriched unknown intergenic regions in NPCs for H3.3, H3K4me1, and H3.3 and H3K4me1.

regions. However, there are exceptions to the co-occupancy of TFBS by H3.3 and H3K4me1, such as TFBS marked only by peaks of H3.3 but not H3K4me1 (**Figure 2.30C**).

To examine the relationship between H3.3 incorporation and TFBS genome-wide, we took the TFBS identified by ChIP-seq in mouse ES cells for Nanog, Oct4, STAT3, Smad1, Sox2, Zfx, c-Myc, n-Myc, Klf4, Esrrb, Tcfcp2l1, E2f1, and CTCF (Chen et al., 2008b), and classified those binding sites as gene promoter, gene body, or intergenic. Remarkably, we find that H3.3 and H3K4me1 are enriched genome-wide at TFBS for all 13 characterized TFs in ES cells, whether these binding sites are in the gene promoter, gene body, or intergenic (**Figure 2.31A**). Upon mutation of H3.3B to H3.2 or H3.1S31, the enrichment of H3.3 at all 13 TFBS is lost, demonstrating the genome-wide dependence of H3.3 TFBS localization on amino acid sequence in mammalian cells (**Figure 2.30A**).

When we examine specific peaks of H3.3 at TFBS in ES cells, we find that many of them are co-localized with regions bound by multiple TFs (**Figure 2.27, 2.30A, C**). Genomic locations bound by more than four TFs have been described as multiple transcription factor-binding loci (MTL), and a subset of MTL bound by Oct4-Sox2-Nanog have been shown to serve as ES cell-specific enhanceosomes (Chen et al., 2008b). Our genome-wide analysis shows that the density of H3.3 in ES cell MTL increases significantly with the number of bound TFs (**Figure 2.33A**). Again, this relationship between transcription factor binding and H3.3 incorporation is dependent on amino acid sequence, as mutation of H3.3 to H3.2 eliminates this correlation (**Figure 2.33A**). Surprisingly, we do not find a correlation between TF binding in MTL and levels of H3K4me1 (**Figure 2.33A**), despite the strong concurrence between H3K4me1 and individual TFBS (**Figure 2.31A**). These data suggest that the level of H3K4me1 at MTL may saturate with lower numbers of TFBS, while levels of H3.3 continue to increase with increasing TF binding.

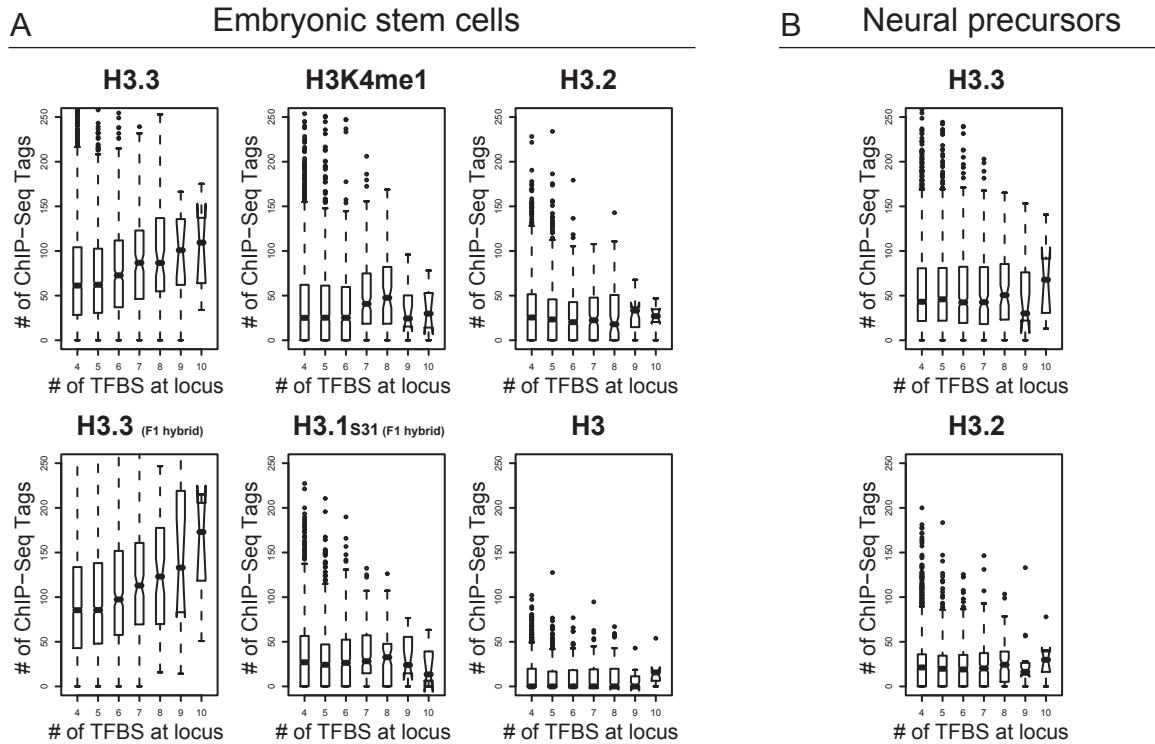


Figure 2.33: Cell-type specific enrichment of H3.3 at multiple transcription factor binding loci

A-B. In ES cells (A) and NPCs (D), levels of H3.3 at ES MTL are dependent on H3 amino acid sequence and cellular state. Box-and-whisker plots showing the ChIP-Seq read densities at MTL separated by # of TFBS at a locus (x-axis). Y-axis represents # of normalized ChIP-Seq tags at a group of MTL with the same # of TFBS, where the central bar is median, box shows 25th and 75th percentiles, and whiskers mark the 1.5x interquartile range.

Following differentiation to NPCs, the genome-wide relationships between H3.3, H3K4me1, and TFBS are reduced, but not entirely eliminated (**Figure 2.31B, 2.33B**). At some specific TFBS (**Figure 2.30A, C**), H3.3 and H3K4me1 enrichment is lost upon differentiation to NPCs. However, at other TFBS, such as those 5' and 3' of *Nanog*, H3.3 and H3K4me1 enrichment is reduced, but partially maintained (**Figure 2.27A**). At the genome-wide level, we observe that the overall enrichment of H3.3 at TFBS is significantly reduced, but not eliminated, upon differentiation from ES to NPCs (**Figure 2.31B**). At ES MTL, the relationship between multiple transcription factor binding and H3.3 enrichment is clearly dependent on cellular differentiation state (**Figure 2.33A-B**). Following differentiation to NPCs, MTL for

known ES cell TFs do not show increased levels of H3.3 (**Figure 2.33B**).

Specific patterns of H3.3 at genes, regulatory elements, and repeats

The data that we have seen in this chapter provide the first genome-wide maps of histone H3 variants in mammalian pluripotent cells and NPCs. I have established the normal patterns of histone H3.3 at active and repressed genes, TFBS, and specific repeats such as telomeres, and have demonstrated that these genome-wide patterns are critically dependent on endogenous H3.3 amino acid sequence. At cell-type specific genes and regulatory regions, the localization of H3.3 changes with cell differentiation. Therefore, genome-wide patterns of H3.3 enrichment are also dependent on cellular state. These findings raise many questions. Although H3.3 is clearly enriched at a variety of regions, what is the function of H3.3 at these elements? Does H3.3 play a similar role in TSS and gene bodies as it does after TES, at TFBS, and at telomeres? The work that I will describe in Chapter Three touches upon a potentially repressive function for H3.3 at telomeres and rDNA. In Chapter Four, I will also discuss ongoing experiments that attempt to address the function of H3.3 in mammalian development and cellular differentiation.

In addition to the function of H3.3, another critical question emerges from the data presented in this chapter. What are the factors that are responsible for H3.3 deposition at these various elements? The histone chaperone HIRA associates with H3.3 in HeLa cells, and it has been described to play a critical role in H3.3 deposition during *Drosophila* male pronuclear decondensation (Loppin et al., 2005; Tagami et al., 2004). Is HIRA required for H3.3 deposition at all classes of sites, or are there site-specific mechanisms that deposit H3.3 at transcribed sequences, TFBS, and telomeres? After establishing the 'wild-type' patterns of H3.3 in mammalian cells, I next sought to determine how these genome-wide patterns of H3.3 are altered in the absence of Hira. As we shall see, the results in Chapter Three indicate that the

deposition of H3.3 at different genomic regions is significantly more complicated than existing models proposed in the literature.

CHAPTER 3

HIRA AND ATRX ARE REQUIRED FOR H3.3 DEPOSITION AT DISTINCT GENOMIC REGIONS

Summary

In Chapter Two, I described the first genome-wide profiles of histone H3 variant localization in mammalian embryonic stem (ES) cells and differentiated neuronal precursor cells (NPCs). In this chapter, I establish the dependence and independence of these patterns on previously identified as well as novel factors involved in H3.3 deposition. To determine the dependence of genome-wide H3.3 patterns on the putative H3.3 chaperone Hira, I compared the genome-wide localization of H3.3 in wild-type and Hira $-/-$ ES cells. These studies are the first to demonstrate that Hira is required for genome-wide H3.3 enrichment at active and repressed genes. Surprisingly, I found that H3.3 enrichment at specific transcription factor binding sites (TFBS), ribosomal DNA, and telomeres is Hira-*independent*. To identify factors that might mediate Hira-independent deposition of H3.3, I used a combined immunoprecipitation and mass spectrometry approach. My work identifies the chromatin remodeling protein Atrx (alpha thalassemia and X-linked mental retardation) and Daxx (death-domain associated protein) as proteins that specifically associate with H3.3 in both pluripotent and non-pluripotent cells, both in the presence and in the absence of Hira. After generating genome-wide profiles of H3.3 in *Atrx*^{flox} and *Atrx*^{null} mouse ES cells, I demonstrate that Atrx is not required for H3.3 deposition at genes or TFBS. Finally, I show that Atrx is specifically required for enrichment of H3.3 at telomeres and rDNA in ES cells, and for the repression of telomeric and ribosomal RNA.

Results

Hira is required for enrichment of H3.3 at active and repressed genes

After establishing the genome-wide patterns of H3.3 in undifferentiated ES cells and differentiated NPCs, I sought to determine whether all of these patterns were dependent on the putative H3.3 chaperone Hira. Hira has been shown to associate with H3.3 in human cells, and to be required for assembly of H3.3 nucleosomes on paternal chromatin during male pronucleus formation in *Drosophila* (Loppin et al., 2005; Tagami et al., 2004). Moreover, as described in Chapter One, Hira is believed to be required for H3.3-specific deposition, perhaps working in combination with CHD1 to engage in replication-independent (RI) chromatin assembly (De Koning et al., 2007; Konev et al., 2007) (see **Figure 1.22**). Hira $-/-$ mice die during early embryogenesis following gastrulation defects and significant patterning abnormalities (Roberts et al., 2002). Although Hira $-/-$ cells are incapable of terminal differentiation into functional neurons, Hira $-/-$ ES cells are surprisingly viable, proliferate normally in culture, and display normal ES colony morphology (Meshorer et al., 2006). Hira $-/-$ mice and ES cells were generated in the laboratory of Dr. Peter Scambler (Meshorer et al., 2006; Roberts et al., 2002).

For the next portion of my thesis work, I therefore began a collaboration with Dr. Scambler's laboratory, who generously provided me with wild-type W9.5 and Hira $-/-$ ES cells. Using the techniques described in the previous chapter, I used zinc finger nucleases (ZFNs) and my H3.3B-EYFP knock-in donor constructs to tag one allele of the endogenous H3.3B gene in Hira $-/-$ ES cells with a C-terminal EYFP tag. Hira $-/-$ ES cells express H3.3B at slightly lower, but comparable levels to wild-type ES cells (**Figure 3.1**).

To determine the distribution of H3.3 in wild-type and Hira $-/-$ ES cells, I initially examined chromosome spreads from metaphase arrested cells. In agreement with previous studies in *Drosophila* embryos and adult cells (Bonney et al., 2007), I

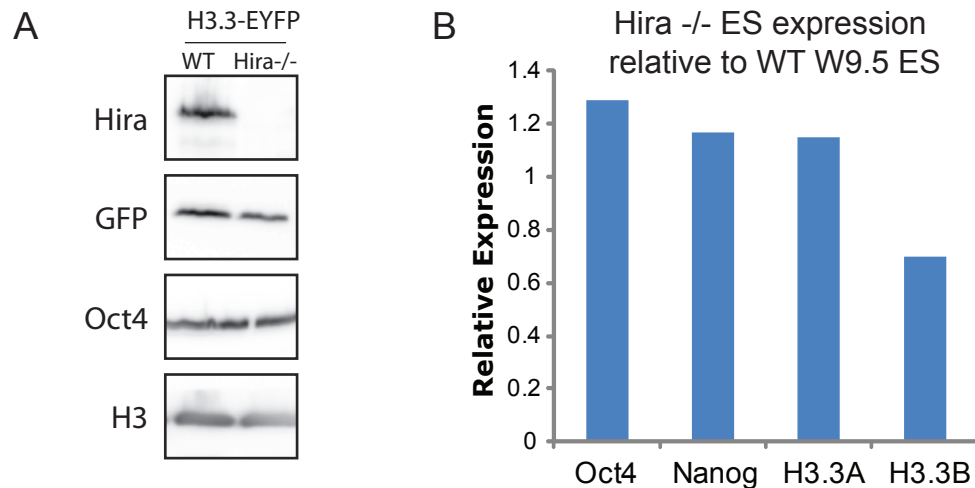


Figure 3.1: Expression of H3.3 and pluripotency genes in wild-type and Hira^{-/-} ES cells

A. Immunoblots from whole-cell extracts showing expression of H3.3-EYFP in wild-type and Hira^{-/-} ES cells, with antibodies as indicated.

B. Hira^{-/-} ES cells maintain expression of Oct4, Nanog, H3.3A, and H3.3B. Real-time PCR was performed following RNA extraction and cDNA synthesis. Data were normalized to Gapdh, and plotted as expression in Hira^{-/-} relative to W9.5 ES cells.

find that H3.3 is still globally incorporated into chromatin in the absence of Hira (**Figure 3.2**). In Hira^{-/-} ES cells, H3.3 has a somewhat weaker signal, and is concentrated in foci near the end of chromosomes (**Figure 3.2**).

I also sought to determine if H3.3 post-translational modifications were altered in the absence of Hira. For example, H3.3 serine 31 phosphorylation (H3.3S31P) is a post-translational modification that is enriched in mitosis and is specific to H3.3 (Hake et al., 2005). H3.3S31P co-localizes with telomeres in ES cells, but becomes more localized to pericentromeric heterochromatin following cell differentiation (Hake et al., 2005; Wong et al., 2009) (see **Figure 1.19**). I found that H3.3S31P is also Hira-independent, as mitotic enrichment of H3.3S31P enrichment is fully maintained in Hira^{-/-} ES cells (**Figure 3.3**).

To compare the genome-wide distribution of H3.3 in wild-type and Hira^{-/-} ES cells at higher resolution, I again performed native ChIP-seq. Although Hira has been suggested to play a critical role in deposition of mammalian H3.3 (Meshorer et

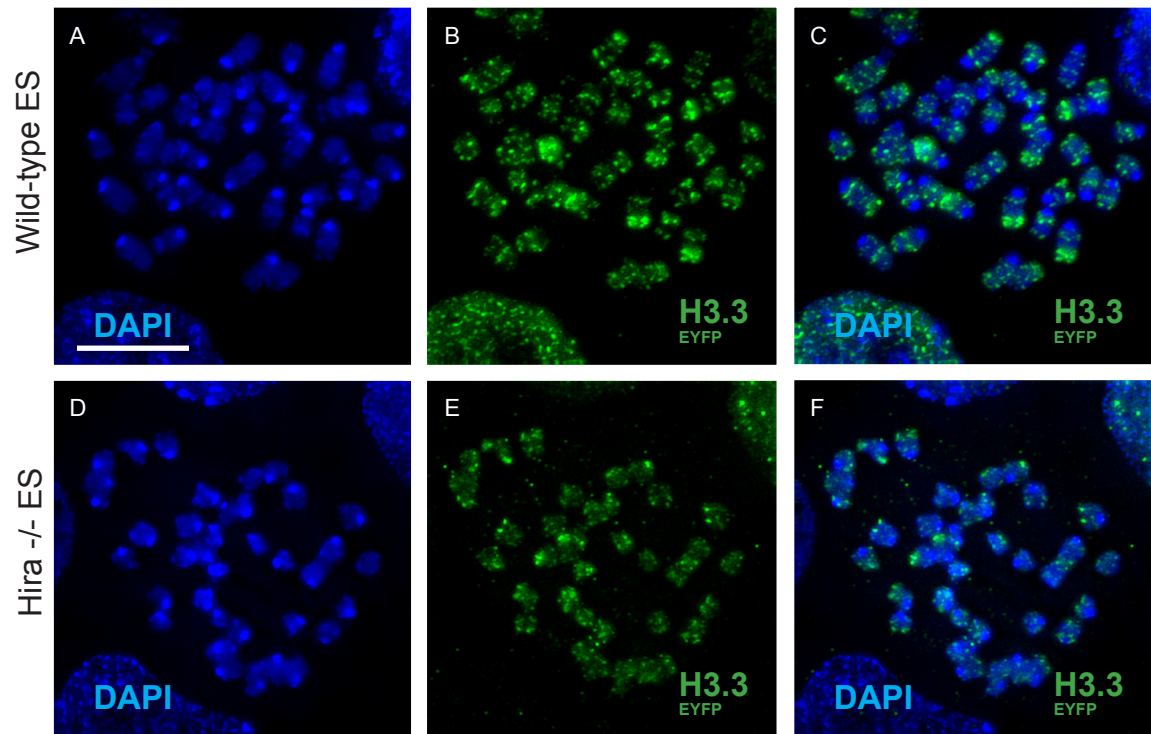


Figure 3.2: Chromosomal localization of H3.3 in wild-type and Hira $-/-$ ES cells

A-F. Metaphase chromosome spreads from WT H3.3-EYFP and Hira $-/-$ H3.3-EYFP ES cells, using anti-GFP. Bar: 10 μ m.

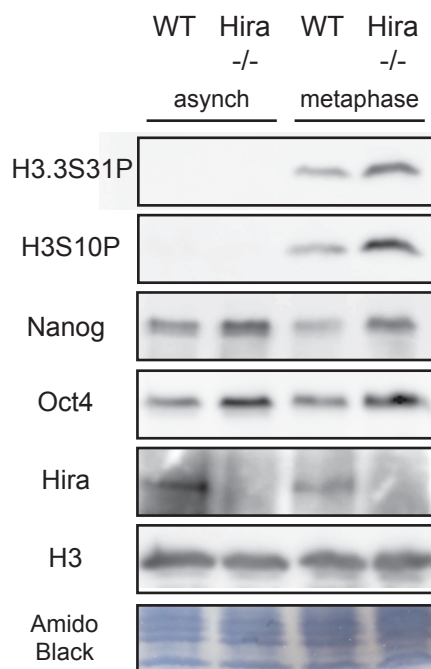


Figure 3.3: H3.3 serine 31 phosphorylation is Hira-independent

Immunoblots of whole cell extracts from WT and Hira $-/-$ ES cells grown asynchronously (left two lanes) or blocked in metaphase with nocodazole (right two lanes).

al., 2006; Tagami et al., 2004), this hypothesis has never been rigorously examined through ChIP, likely due to the difficulty of raising an antibody to distinguish H3.3 from the other H3 variants. As might be expected, my data show that Hira is specifically required for genome-wide H3.3 enrichment at active and repressed high CpG content promoter (HCP) genes (**Figure 3.4A-D**). In the absence of Hira, the pattern of H3.3 around active and repressed genes (**Figure 3.4D**) resembles H3.2 (**Figure 3.4C**), and general H3 (**Figure 2.17C**), with no enrichment surrounding the TSS, and depletion of H3.3 at the TSS itself. For example, in Hira $-/-$ ES cells, the enrichment of H3.3 around the TSS of repressed bivalent genes such as *Pparg* is abolished (**Figure 3.4E**). At highly expressed housekeeping genes such as the ribosomal protein-coding gene *Rps19*, the enrichment of H3.3 in the gene body and after the TES is mostly eliminated in the absence of Hira (**Figure 3.4F**). These data demonstrate that the vast majority of H3.3 enrichment at both active and repressed genes in ES cells is Hira-dependent, as expected from the existing literature.

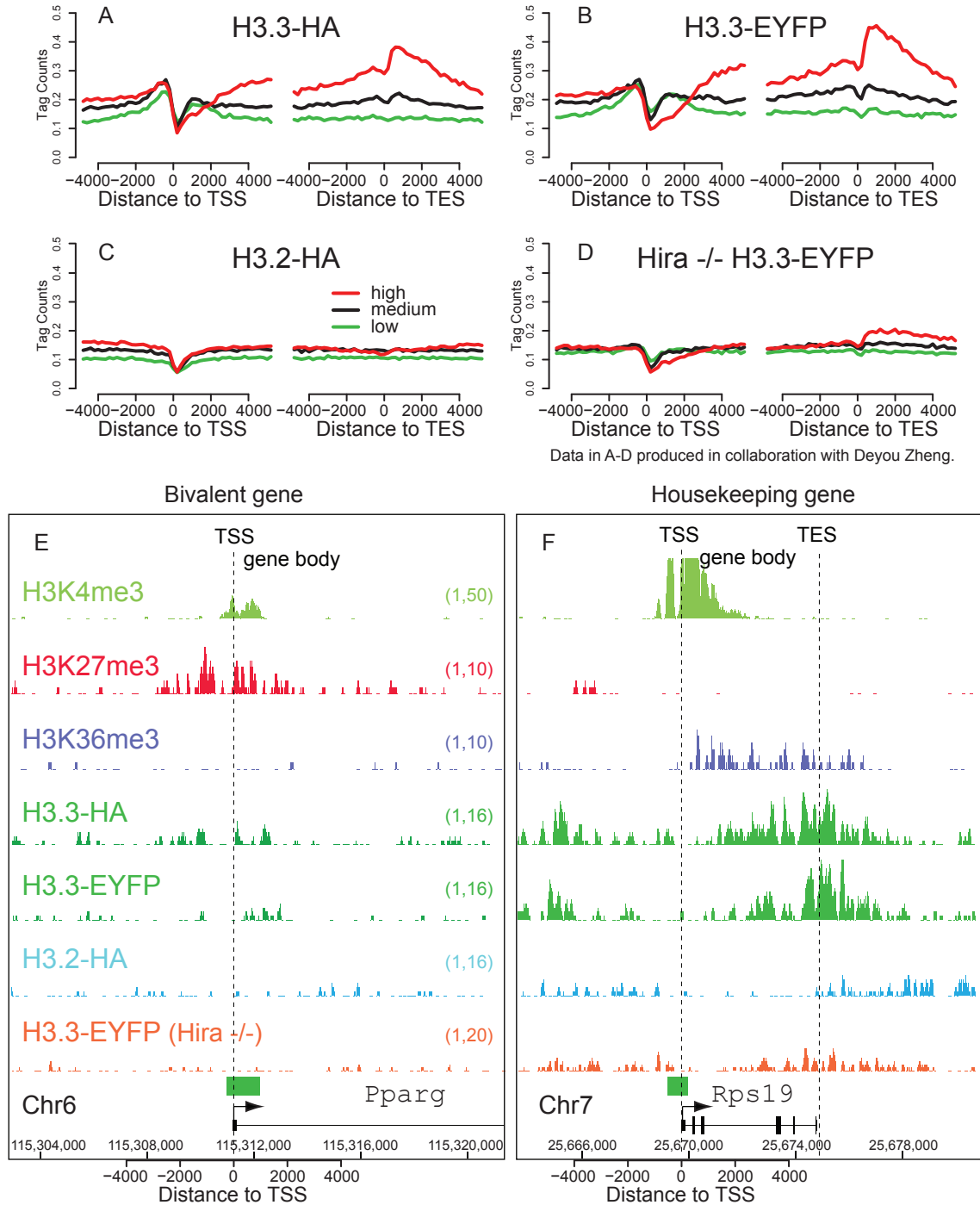


Figure 3.4: Enrichment of H3.3 at active and repressed genes is Hira-dependent

A-D. Native ChIP-seq profiles of H3 variants, in wild-type ES cells (B-D) and Hira^{-/-} ES cells (E) across the TSS and TES for highly active (red), medium expressing (black), or low expressing (green) CpG rich genes, with data represented as in Figure 2.

E. H3.3 enrichment at the TSS of repressed bivalent gene Pparg is Hira-dependent.

F. H3.3 enrichment at the active, ribosomal-protein coding gene Rps19 is Hira-dependent.

Hira $-/-$ ES cells maintain global patterns of ES cell gene expression

Given the lack of genic H3.3 enrichment in the absence of Hira, one possible explanation for the alteration in genome-wide patterns of H3.3 at active and repressed genes is a significant alteration in patterns of gene expression. Therefore, my collaborators and I sought to determine if Hira $-/-$ ES cells show genome-wide alterations in gene expression. Remarkably, microarray analysis reveals that global patterns of ES cell gene expression are maintained in Hira $-/-$ ES cells (**Figure 3.5A-C**). Moreover, genome-wide patterns of H3K36me3 are also extremely similar in wild-type and Hira $-/-$ ES cells (**Figure 3.6**). What accounts for the strikingly 'normal' phenotype of Hira $-/-$ ES cell growth and gene expression, but for the significant defects seen upon Hira $-/-$ ES cell differentiation?

Intriguingly, the phenotype of Hira $-/-$ ES cells is similar to that of CHD1 shRNA ES cells (Gaspar-Maia et al., 2009). Like Hira $-/-$ ES cells, CHD1 shRNA ES cells maintain global patterns of ES cell gene expression, but display significant defects upon differentiation (Gaspar-Maia et al., 2009). CHD1 was recently shown to play a role in H3.3 deposition during *Drosophila* pronuclear decondensation, and was suggested to work together with Hira (Konev et al., 2007). It is therefore possible that the defects upon differentiation of Hira $-/-$ and CHD1 shRNA ES cells could result from impaired replication-independent (RI) H3.3 deposition in more slowly dividing or post-mitotic cells. Rapidly dividing Hira $-/-$ and CHD1 shRNA ES cells could be partially 'rescued' by replication-dependent (RD) histone deposition, enabling global maintenance of the ES cell transcriptome, and I will discuss this hypothesis in greater detail in the following chapter.

Notably, I cannot at this time rule out more subtle defects in transcription that were not detectable by genome-wide microarray studies of Hira $-/-$ ES cells. For example, Hira $-/-$ ES cells could still manifest low levels of spurious transcripts, as the *S. pombe* HIRA protein homologs were recently shown to be required for

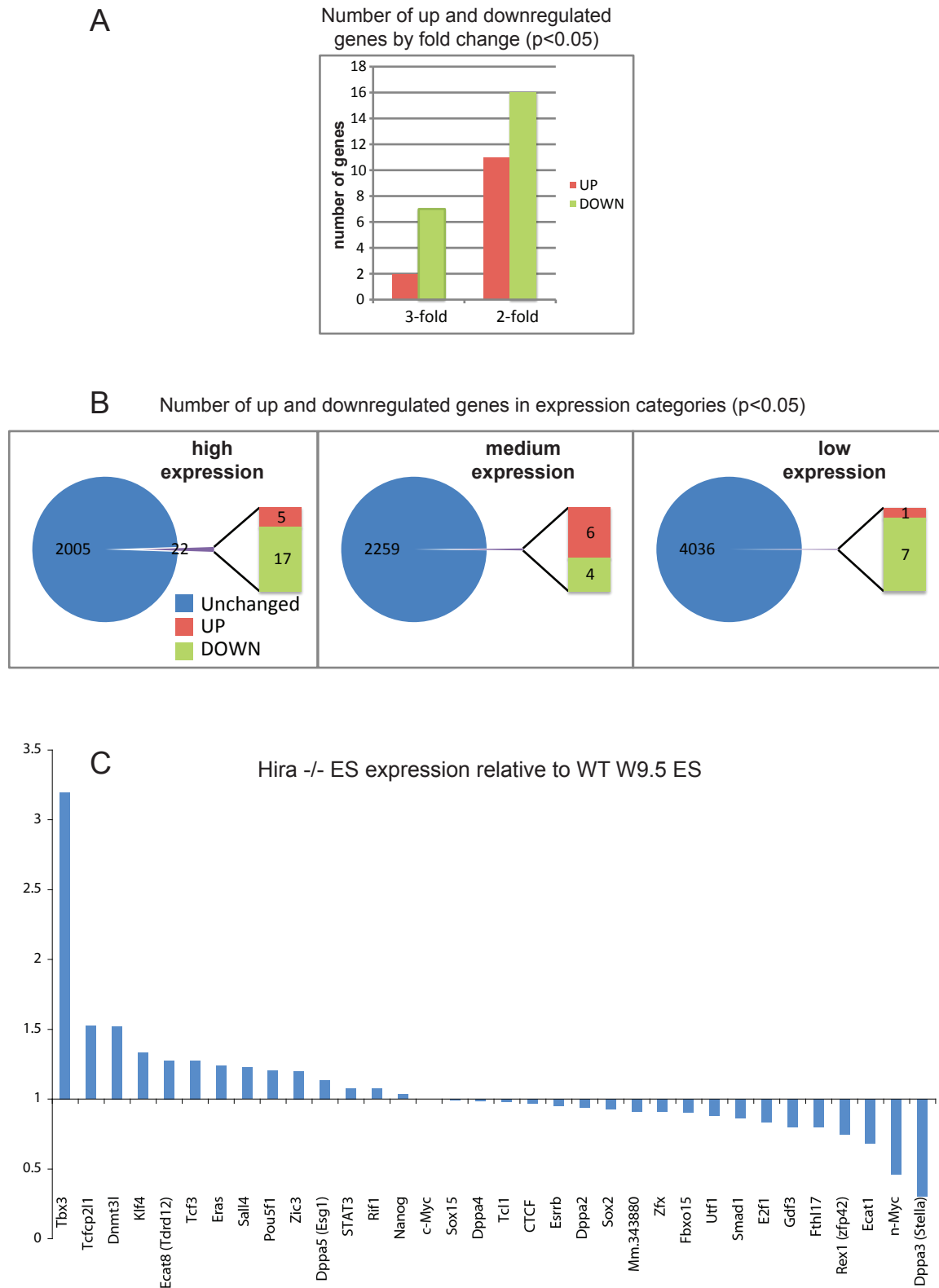
Figure 3.5: Microarray analysis of gene expression in wild-type and Hira ^{-/-} ES cells

A-C. Triplicate GeneChip® Mouse Gene 1.0 ST Arrays were hybridized with cDNA from three independent biological replicates of HIRA^{-/-} ES (cl34) and WT ES (W9.5) cell lines. The data was analyzed using the LIMMA package of Bioconductor v2.0 and R v.2.9.2, applying GC-RMA normalization and correction for multiple testing (Benjamini-Hochberg) with a significance threshold of $p < 0.05$.

A. Significant differentially expressed genes were binned according to their fold-change.

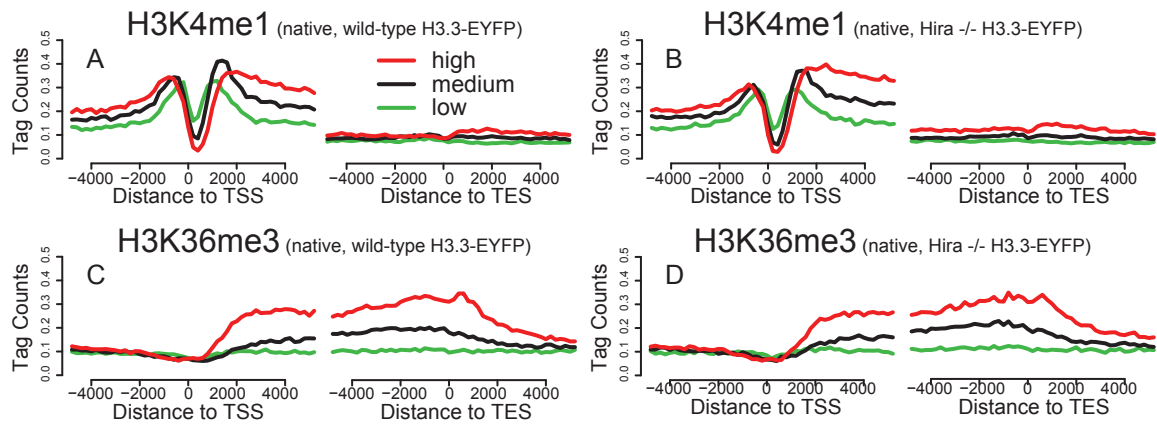
B. The number of differentially expressed genes was determined for each low, medium, high expression gene sets derived from WT ES cells. High, medium, and low expression genes were determined from a previous microarray of WT ES cells (Mikkelsen et al., 2007) as detailed in Materials and Methods. Overall, only a small number of genes in these gene sets were affected in Hira ^{-/-} ES cells.

C. Differential expression of ES cell related genes in wild-type (W9.5) vs Hira ^{-/-} ES cells.



Data in A-C generated by Simon Elsaesser, Ariane Chappier, and Peter Scambler, and used with permission.

Figure 3.5



Data in A-D produced in collaboration with Deyou Zheng.

Figure 3.6: H3K4me1 and H3K36me3 are maintained at genes in the presence and absence of Hira

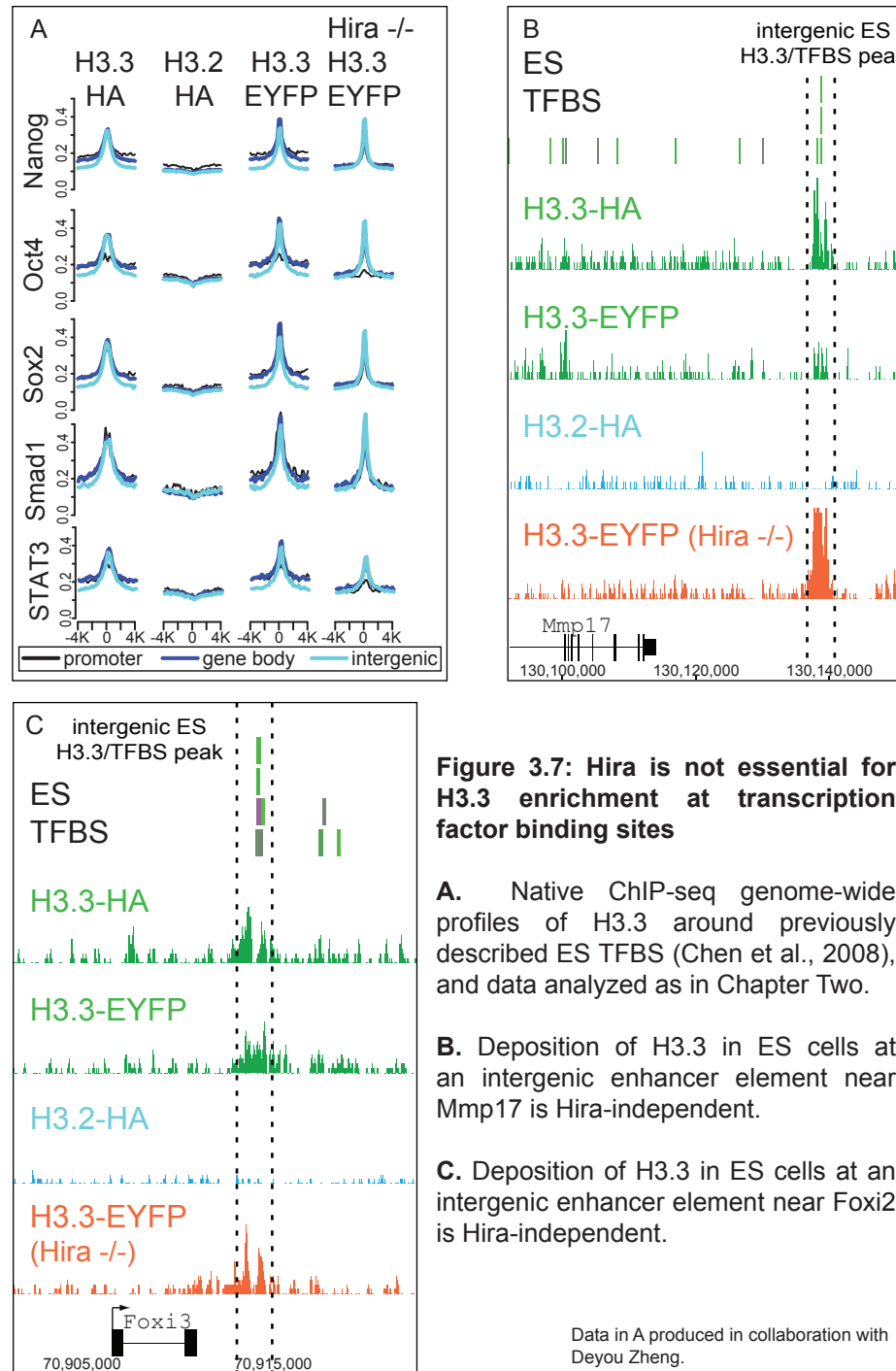
A-D. Profiles of H3K4me1 (**A-B**) and H3K36me3 (**C-D**) in wild-type (**A,C**) and Hira -/- (**B,D**) H3.3-EYFP ES cells across TSS and TES for highly active, medium, or low expressed CpG rich genes. Y-axis represents the average number of tags per gene per 200 bp per 1 million mapped reads.

promoter silencing and global repression of cryptic antisense transcripts (Anderson et al., 2009). Overall, my data show that genome-wide enrichment of H3.3 at active and repressed genes in ES cells is Hira-dependent. However, Hira-dependent genic deposition of H3.3 is not required for maintenance of the undifferentiated ES cell transcriptome.

Hira-independent enrichment of H3.3 at transcription factor binding sites

Due to the significant enrichment of H3.3 at genic and intergenic TFBS, (**Figure 2.28, 2.29**), I also examined whether these patterns were dependent on Hira. As the ES cell transcriptome is globally maintained in Hira *-/-* ES cells, I reasoned that the overall profile of TFBS would be similar in wild-type and Hira *-/-* ES cells, enabling us to use the same existing dataset of 13 different ES cell TFBS (Chen et al., 2008b). Surprisingly, despite the observation that Hira is required for genome-wide H3.3 enrichment at active and repressed genes, my collaborators and I find that Hira is not essential for genome-wide H3.3 enrichment at TFBS.

Global profiles of H3.3 are largely maintained at ES cell TFBS in Hira *-/-* ES cells (**Figure 3.7, 3.8**), demonstrating that H3.3 enrichment at most regulatory elements is Hira-independent. For example, H3.3 enrichment at an intergenic MTL approximately 30 KB from the *Mmp17* gene is maintained in Hira *-/-* ES cells (**Figure 3.7B**), as is H3.3 enrichment at an intergenic MTL near the repressed *Foxi3* gene (**Figure 3.7C**). Indeed, levels of H3.3 are increased at some TFBS in Hira *-/-* ES cells (**Figure 3.9**). However, we note that levels of H3.3 are also reduced at other TFBS in Hira *-/-* ES cells (**Table 3.1, Figure 3.8**). Of all previously identified ES TFBS (Chen et al., 2008b), 34% show greater than 2-fold more H3.3 tags in Hira *+/+* than in Hira *-/-*, while 12% show greater than 2-fold more H3.3 tags in Hira *-/-*, indicating that targeting of H3.3 to the majority (54%) of known ES TFBS is Hira-independent (**Table 3.1, Figure 3.8**). Global profiles of



H3K4me1 are very similar in wild-type and Hira $-/-$ ES cells (**Figure 3.6A-B, 3.8C**), indicating that Hira is also not required to maintain the localization of H3K4me1. Overall, these data demonstrate that mammalian ES cells utilize Hira-independent mechanisms to deposit and enrich H3.3 at many non-genic elements.

Hira-independent association of Atrx and Daxx with histone H3.3

To identify candidates that might mediate Hira-independent deposition of H3.3, I began a collaboration with Assistant Professor Ileana Cristea of Princeton University, formerly a postdoctoral fellow with Dr. Brian Chait here at Rockefeller University. Dr. Cristea and I used an immunoaffinity purification and mass spectrometry approach optimized to help maintain protein complexes in their original state (Cristea et al., 2005). Using an anti-GFP antibody, proteins interacting with EYFP-tagged H3 variants in wild-type and Hira $-/-$ ES cells were immunoprecipitated, resolved using gel electrophoresis and identified using mass spectrometry (**Figure 3.10A, 3.11**). We also performed immunoblots of immunoprecipitated proteins to confirm specific candidates (**Figure 3.10B**).

We found many interacting proteins common to all H3 variants in wild-type and Hira $-/-$ ES cells, including core histones and previously described members common to both RC and RI chromatin assembly complexes, such as Nasp, Asf1a, Asf1b, and Rbap48 (a full list of identified proteins is presented in **Table 3.2**). Notably, the previously described H3.3 chaperone Hira (Tagami et al., 2004) was identified specifically in proteins isolated with H3.3.

However, in addition to Hira, we also identified Atrx and Daxx as proteins that specifically associate with H3.3 (**Figure 3.10, 3.11, 3.12**). Atrx is a member of the SNF2 family of chromatin remodeling factors (Picketts et al., 1996). Mutations of human ATRX give rise to the ATR-X syndrome, a disorder characterized by a form of X-linked mental retardation that is frequently associated with multiple

Figure 3.8 H3.3 enrichment at ES TFBS is partially perturbed in Hira ^{-/-} ES cells, but is not significantly affected in Atrxnull ES cells

A-C. H3.3 and H3K4me1 read densities at ES TFBS (Chen et al., 2008) were analyzed and plotted from anti-GFP and anti-H3K4me1 native ChIP-seq experiments. TFBS were first separated into different groups based on their H3.3 (or H3K4me1) ChIP-seq read densities in wild-type ES cells (X-axis). Corresponding ChIP-seq read densities from Hira ^{-/-} or Atrx null samples are shown on the Y-axis as boxplots. Upper schematics depict possible results. H3.3 levels at ES TFBS in wild-type and Hira ^{-/-} ES cells remain generally correlated, indicating that H3.3 enrichment is not completely abolished at ES TFBS in Hira ^{-/-} ES cells (**A**). However, H3.3 levels at ES TFBS are reduced at many ES TFBS in Hira ^{-/-} ES cells, and increased at other ES TFBS, particularly at ES TFBS with a high density of reads, suggesting that Hira affects H3.3 deposition in some but not all TFBS. H3.3 enrichment at ES TFBS remains highly correlated between wild-type and Atrxnull ES cells (**B**), and H3K4me1 enrichment at ES TFBS remains highly correlated between wild-type and Hira ^{-/-} ES cells (**C**).

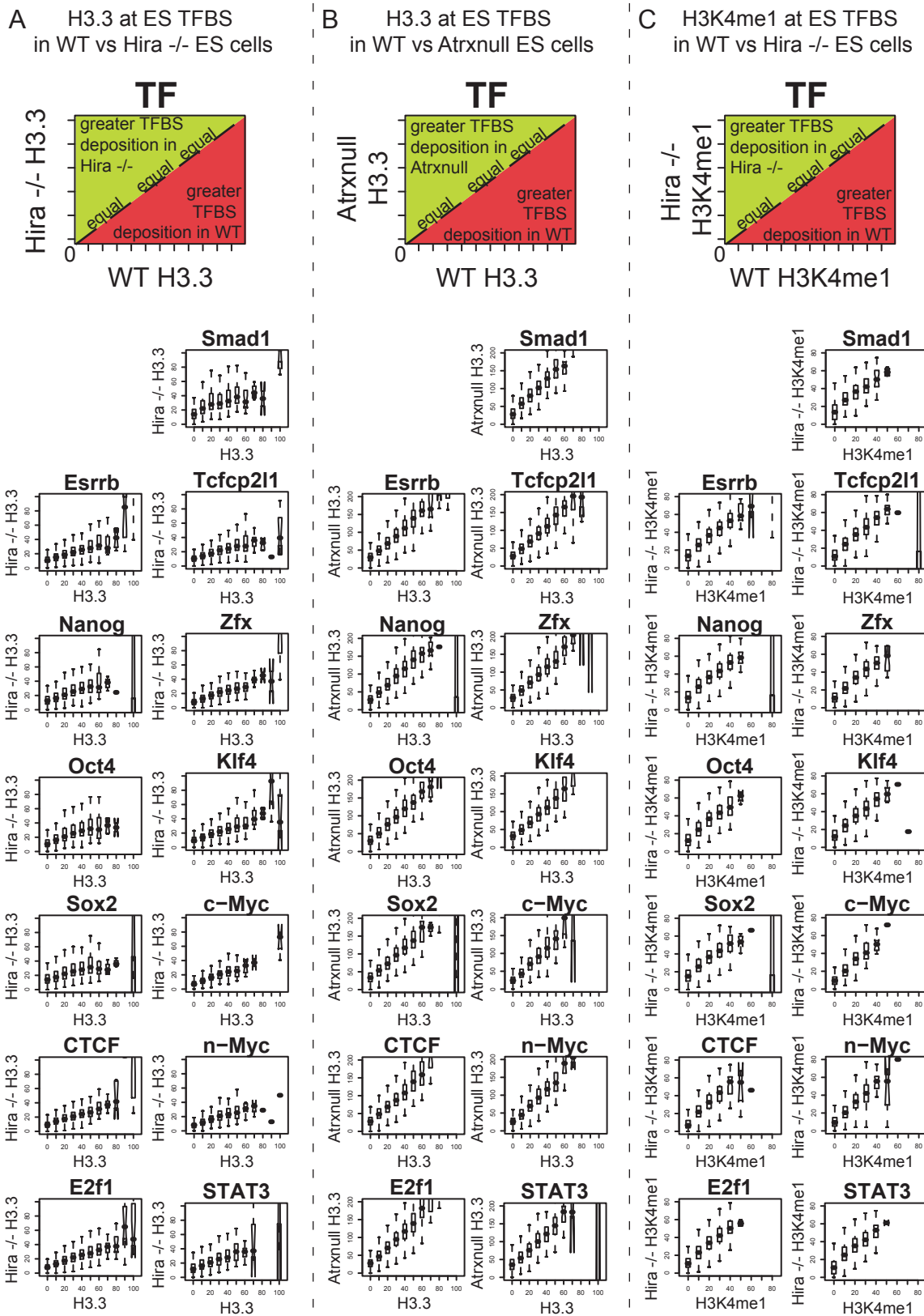


Figure 3.8

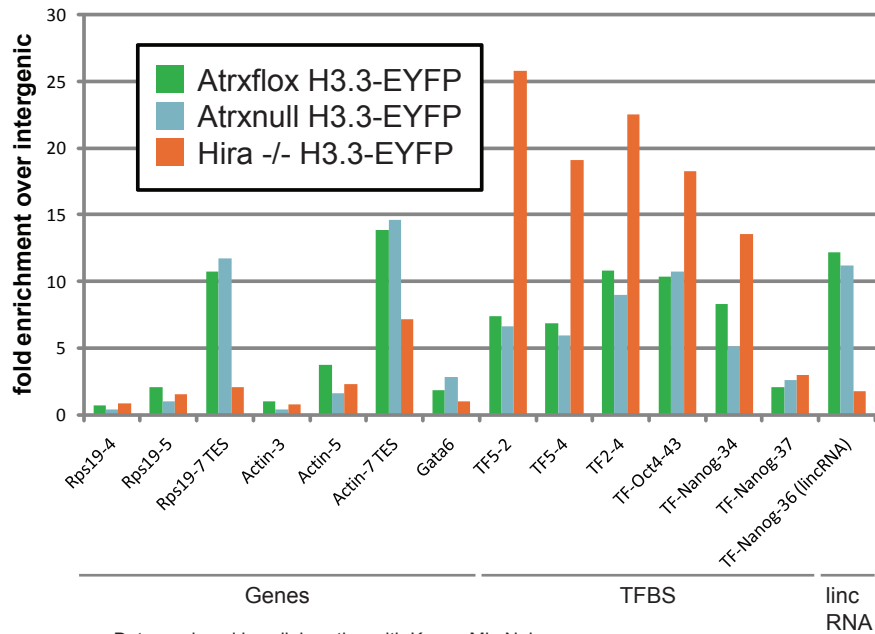


Figure 3.9: Real-time PCR confirmation of ChIP-seq results

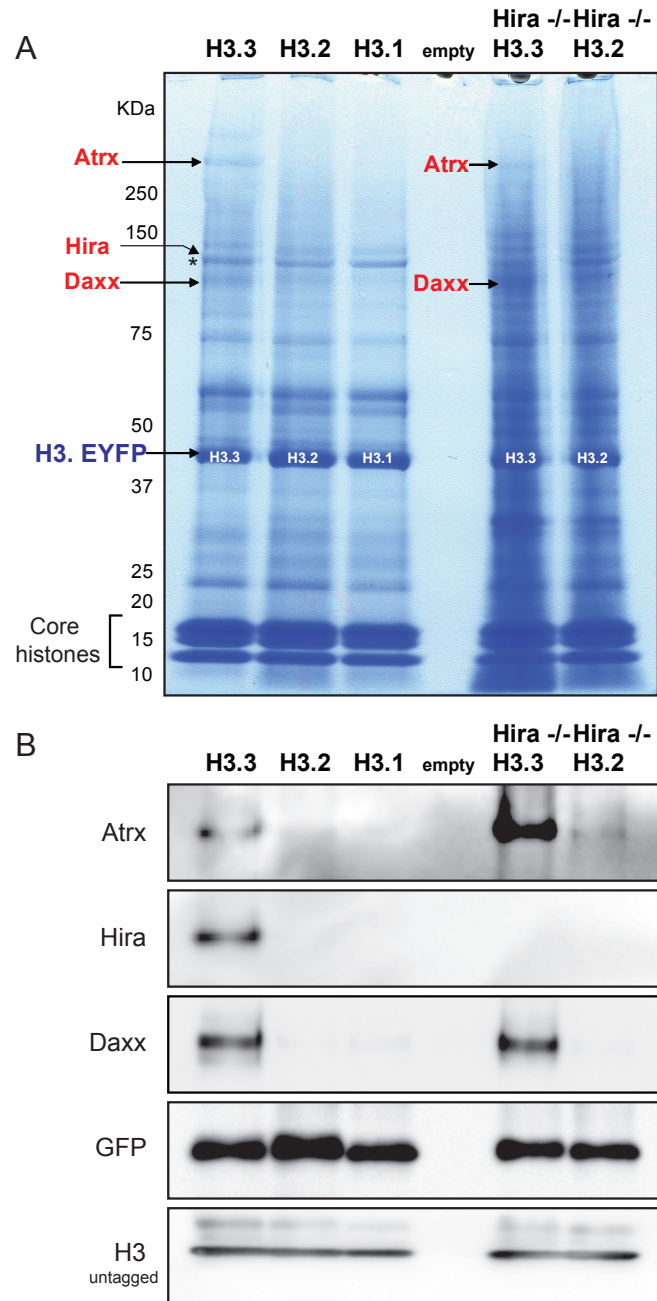
Real-time PCR of ChIP DNA confirms ChIP-seq results. Note similar profiles of H3.3 at genes and TFBS in Atrxflx (green) and Atrxnul (blue) ES cells. At genes, particularly active gene bodies and TES, H3.3 enrichment is decreased in Hira -/- ES cells (orange), while at TFBS, H3.3 enrichment is Hira-independent. Primer sequences are listed in Appendix. Data are plotted as fold enrichment over intergenic (%input at location / %input at intergenic).

TF	2-fold decrease of H3.3 in Hira -/-	2-fold increase of H3.3 in Hira -/-	No change
Tcfcp2l1	0.373	0.105	0.522
CTCF	0.277	0.125	0.598
Klf4	0.387	0.096	0.517
Sox2	0.304	0.201	0.495
Pou5f1	0.312	0.188	0.5
Nanog	0.27	0.216	0.514
Zfx	0.421	0.068	0.511
Smad1	0.265	0.22	0.515
STAT3	0.346	0.146	0.508
E2f1	0.404	0.079	0.517
n-Myc	0.389	0.073	0.538
c-Myc	0.361	0.07	0.569
Esrrb	0.358	0.111	0.531
all	0.344	0.12	0.536

Data analysis performed by Deyou Zheng.

Table 3.1: Global fraction of ES TFBS with differential H3.3 deposition between wild-type and Hira -/- ES cells

Significant difference is defined here as 2-fold greater H3.3 tag densities from either wild-type or Hira -/- ChIP-Seq analyses.



Data in A-B produced in collaboration with Ileana Cristea (A)

Figure 3.10: Atrx and Daxx association with H3.3 is Hira-independent

A. H3.3-EYFP, H3.2-EYFP, H3.1-EYFP, and Hira ^{-/-} H3.3-EYFP associated proteins were immunopurified from heterozygous H3.3B ES cells via the EYFP tag, resolved by SDS-PAGE, and visualized by Coomassie Blue. Proteins were identified by mass spectrometry, and those listed in red are specific to H3.3. The arrow indicating Hira points to a polypeptide in the H3.3 lane running immediately above the polypeptide Nasp (*), which is common to all purified H3 variants. A full list of identified proteins is presented in **Table 3.2**.

B. Immunoblots of immunopurified H3 variant associated proteins from ES cells with antibodies as indicated.

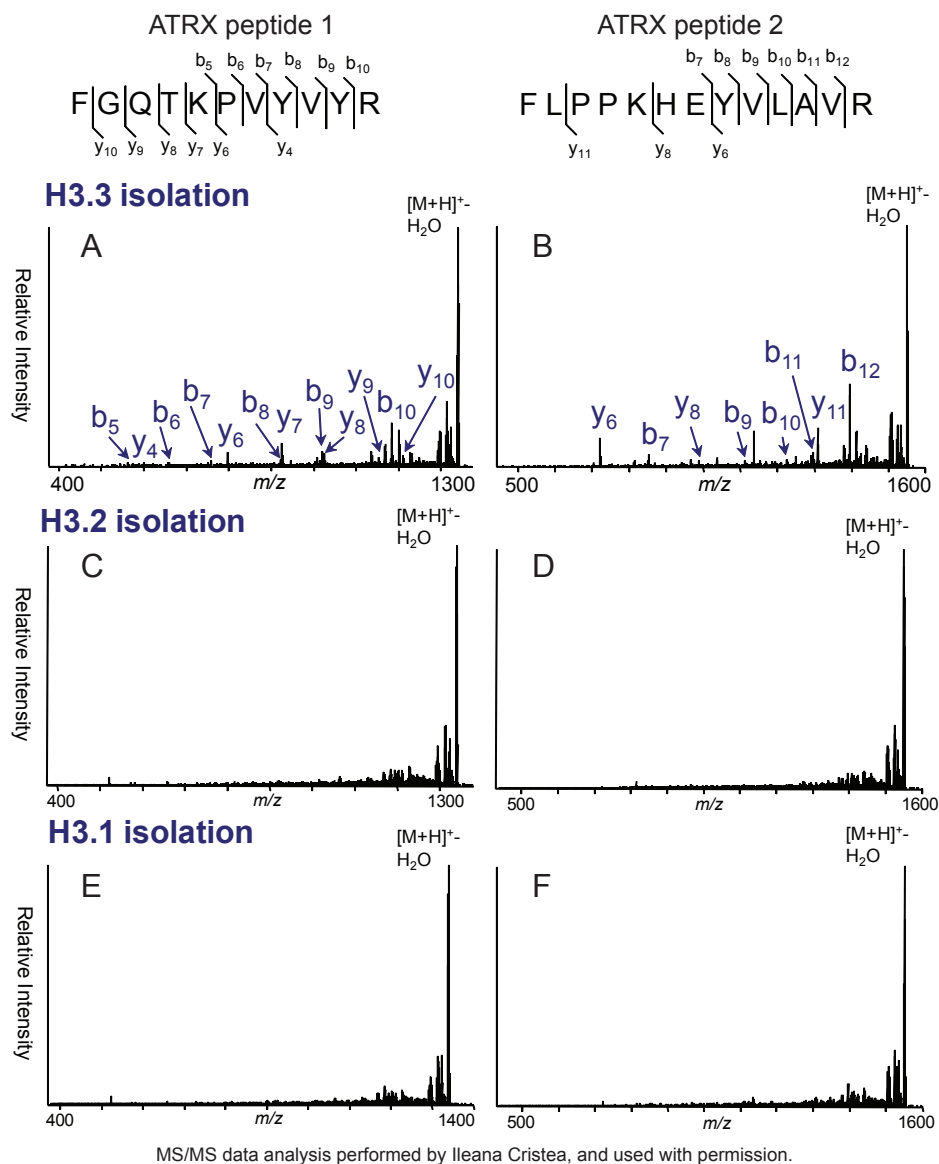


Figure 3.11: Probing the specificity of Atrx association with H3.3 using hypothesis-driven MS/MS analysis

Hypothesis-driven multistage mass spectrometry (Kalkum et al., 2003) was utilized to obtain a highly sensitive detection of specific peptides from proteins of interest, even if these peptide species could not be discerned in the primary matrix-assisted laser desorption/ionization MALDI MS analysis because of insufficient signal-to-noise. Even if Atrx peptides were not observed in the H3.2 and H3.1 samples at the MS level, their corresponding m/z values ($[M+H]^+$) were selected, subjected to collision induced dissociation (CID) fragmentation, and analyzed using MALDI IT MS/MS analyses. The expected fragmentation pattern for the Atrx peptides was confirmed only in the sample isolated with H3.3 (A-B) as assessed by b and y ion fragments and the preferential cleavages C-terminus of Glu and N-terminus of Pro residues. See also top row of **Table 3.2**.

Table 3.2: List of H3 variant associated proteins identified by mass spectrometry

The presence of the interacting partners in the isolations of the different H3 variants (H3.3, H3.2, H3.1, Hira -/- H3.3 and Hira -/- H3.2) is indicated with + (detected) and - (not detected). Only the proteins identified as main hits by database searching and confirmed by manual data interpretation are illustrated. The specificity of Hira, Atrx, and Daxx association with H3.3 were confirmed by western blots and further experiments (see main text for details). The specific association of other proteins with particular H3 variants will require further investigation. The number of peptides, protein sequence coverage and XProteo d' (discriminability) score obtained following the analysis of the MS data are indicated for each protein isolated with the H3 variants. The XProteo score is based on an improved version of the ProFound Bayesian algorithm (Zhang and Chait, 2000) and measured as d'—the normalized distance between the score distribution (of the candidate protein) and the distribution of randomly matched proteins (in units of standard deviation). A d'=4 correspond to a true positive rate of 0.99 and a false positive rate of 0.05. Candidate proteins were confirmed by CID MS/MS analyses of the indicated peptides that were acquired and interpreted manually.

*See **Figure 3.11** for an example of MS/MS analysis of ATRX peptides.

Table 3.2

Protein name	Protein ID	H3.3	H3.2	H3.1	Hira -/- H3.3	Hira -/- H3.2	peptides selected for MS/MS
Atrx*	gi 6678609 ref NP_033556.1	+	-	-	+	-	FGQTKPVVYR*
							FLPPKHEYVLAVR*
							WAEEFNDETNVR
							EEISDHENNVILLESDLR
							TTSTSNPSSPAPDWYKDFVTDDAEV- LEHSGK
Daxx	gi 24636788 sp O35613	+	-	-	+	-	LEQLLALYVAEIR
							TETSDHPEVVPFLHK
							ENRTLAMNR
							NSEPAEGLR
							ELDLSELDDPDSSYLQEAR
hnRNP G-T	gi 94381180 ref XP_914797.2	-	-	-	+	-	RDPLYGPR
							VAQATKPVFESGR
Macf1	gi 94373701 ref XP_996894.1	-	-	-	+	-	LQAALASTQQFQQMFDEL
							SLNQPTPPMPILSQSEAKNPR
Hira	gi 52426778 ref NP_034565.2	+	-	-	-	-	FPEILATLR
							YLVNEGFEYR
Dnmt3a	gi 13938621 gb AAH07466.1	+	-	-	-	-	LFFEFYR
							AIYEVQLQVASSR
Ssrp1	gi 110283006 sp Q08943	+	-	-	-	-	FDEISFVNFR
							NEVTLEFHQNDDAEVSLMEVR
RGS9	gi 2739458 gb AAC99481.1	+	-	-	-	-	SPPGMNVLVDYGLDR
							EAANTVDITQVMSK
FACTp140	gi 110287968 sp Q920B9	+	-	-	-	-	NEGNIFNPPEATFVK
							APGEQTVPALNLQNAFR
							VEALTKEELEFEVPFR
mHEL-5	gi 71153505 sp O70133	+	+	+	+	+	DFVNYLVR
							LAHFEPSSQR
							EGERVEPYK
							DINTDFLLVLR
							TTQVPQYILDDFIQNR
SNF2H	gi 14028669 gb AAK52454.1	+	+	+	+	+	YLVIDEAGR
							QPNVQDFQFFPPR
							TPEEVIEYSAVFWER
NASP	gi 13384598 ref NP_058057.2	+	+	+	+	+	EIEELKELLPEIR
							SLQENEEEEIGNLEAWDMLDLAK
573041019Rik	gi 51261128 gb AAH79531.1	+	+	+	+	+	REDIQQFEEFQSK
							DAATYWPLNWR
RBBP-4	gi 90101524 sp Q60972	+	+	+	+	+	TVALWDLR
							TPSSDVLVFDYTK

Dnajc9	gi 23956266 ref NP_598842.1	+	+	+	+	+	EIPYSAFVK
							ISLEDIAFEK
Parp1	gi 74144748 dbj BAE27352.1	+	+	+	+	+	TLGDFLAEYAK
							QPDVEVDGFSELR
							DELGFRPEYSASQLK
Ubiquitin	gi 51701919 sp P63049	+	+	+	+	+	ESTLHLVLR
							EGIPPDQQR
ASF1A	gi 13384964 ref NP_079817.1	+	+	+	+	+	ENPPVKPDFSK
							VGYVNNNEYTETELR
ASF1B	gi 12859888 dbj BAB31809.1	+	+	+	+	+	VGYVNNNEYPDPCLR
							ENPPPKPDFSQLQR
Atp5b	gi 23272966 gb AAH37127.1	+	+	+	+	+	VALTGLTVAEYFR
							AIAELGIYPAVDPLDSTR
H2A	gi 28316756 ref NP_783589.1	+	+	+	+	+	AGLQFPVGR
							VTIAQGGVLPNIQAVLLPK
H2B	gi 30061385 ref NP_835507.1	+	+	+	+	+	LLPGELAK
							AMGIMNSFVNDIFER
H2A.Z	gi 7949045 ref NP_058030.1	+	+	+	+	+	GDEELDSLIK
							AGLQFPVGR
							ATIAGGGVIPHIHK
							VGATAAVYSAAILEYLTAEVLELAGNASK
YFP	gi 3891568 pdb 1YFP	+	+	+	+	+	FEGDTLVNR
							SAMPEGYVQER
H3	gi 53734121 gb AAH83353.1	+	+	+	+	+	YRPGTVALR
							VTIMPKDTQLAGR
							EIAQDFK
							STELLIR
H4	gi 85057093 gb AAI11814.1	+	+	+	+	+	VFLENVIR
							ISGLIYEETR
MCM2	gi 37359742 dbj BAC97849.1	+	+	+	+	+	ISHLPLVEELR
							IFASIAPSIYGHEDIKR
MCM6	gi 6678832 ref NP_032593.1	+	+	+	+	+	DFYVAFQDLPTR
							EIPFAKDFYVAFQDLPTR
MCM7	gi 74210112 dbj BAE21331.1	+	+	+	+	+	LMMEQRSR
							GFTPAQFQAALDEYEELNVVQVNTSR
Hsp1	gi 40556608 ref NP_032328.2	+	+	+	+	+	HLEINPDHPIVETLR
							HNDDEQYAWESSAGGSFTVR
HAT1	gi 28076885 ref NP_080391.2	+	+	+	+	+	FQELVEDYRR
							FPEDLENDIR
							LLVTDMSDAEQYR
							YYISFPTVLDITAEDPSR
HSP70	gi 74211667	+	+	+	+	+	ARFEELNADLFR
							TVTNAVTVPAYFNDQR

Mass spectrometry analysis performed in collaboration with Ileana Cristea, and used with permission.

Table 3.2 continued

congenital abnormalities and alpha thalassemia (Gibbons et al., 2008). ATRX has been described to co-exist in a chromatin-remodeling complex with the death domain-associated protein Daxx, and these proteins have been localized to heterochromatin and promyelocytic leukemia (PML) nuclear bodies in human and mouse cells (Ishov et al., 2004; Tang et al., 2004; Xue et al., 2003). As Atrx and Daxx specifically associate with H3.3 both in wild-type and Hira ^{-/-} ES cells, we conclude that this association is Hira-independent (**Figure 3.10A-B**).

To determine if the association between H3.3, Atrx, and Daxx was also conserved in differentiated human cells, I collaborated with Dr. Peter Lewis, a

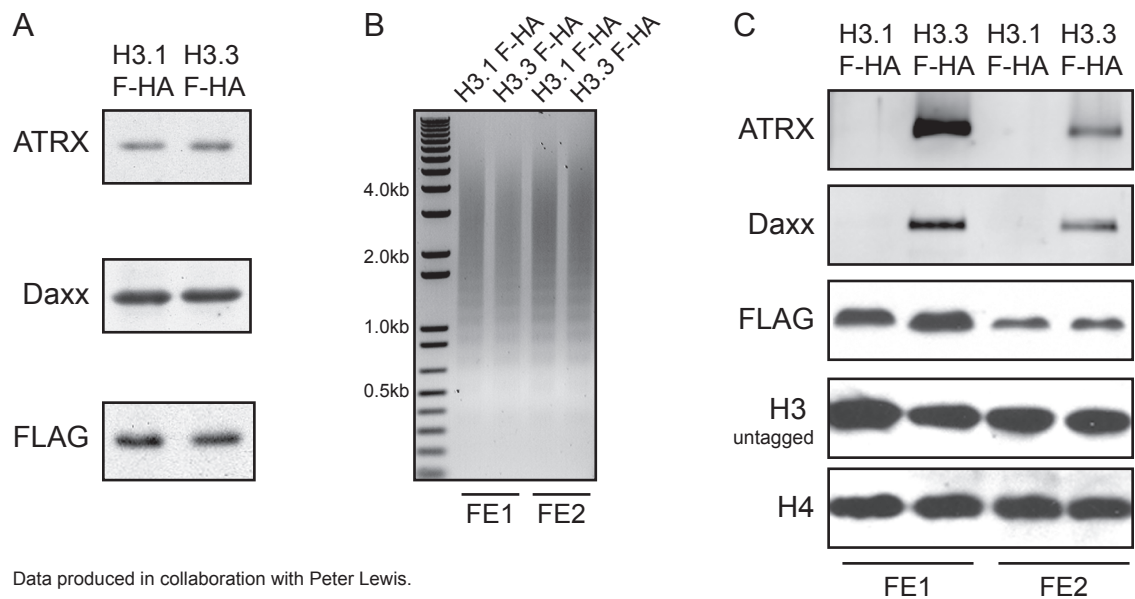


Figure 3.12: H3.3 specific association with Atrx and Daxx is conserved in differentiated human cells

A. Immunoblots of Atrx, Daxx and FLAG-HA-H3 from HeLa cells stably expressing H3.3-FLAG-HA (FHA) or H3.1-FHA (Tagami et al., 2004).

B. Oligonucleosomes purified from H3.3-FHA and H3.1-FHA HeLa cells. Micrococcal nuclease digested chromatin was incubated with M2 agarose beads in order to purify H3.3-FHA and H3.1-FHA oligonucleosomes and associated proteins. Two adjacent FLAG elution fractions (FE1 and FE2) from each H3.3-FHA and H3.1-FHA purification are shown in **B-C**. Oligonucleosomal DNA was purified and run on a 1% agarose gel, with fragment size indicated.

C. Immunoblots of eluate from two adjacent FLAG elution fractions (FE1 and FE2) following FLAG affinity purification from H3.3-FHA and H3.1-FHA oligonucleosomes (**B**).

postdoctoral fellow in the Allis lab. We isolated oligonucleosomes and chromatin-associated proteins from human cervical cancer (HeLa) cells carrying stably expressed FLAG-HA tagged H3.3 or H3.1 (Tagami et al., 2004) (**Figure 3.12**). Following FLAG affinity purification and FLAG peptide elution, immunoblots of H3.3 and H3.1 associated proteins reveal that both Daxx and Atrx are also specifically associated with H3.3 but not H3.1 oligonucleosomes in human cells (**Figure 3.12C**).

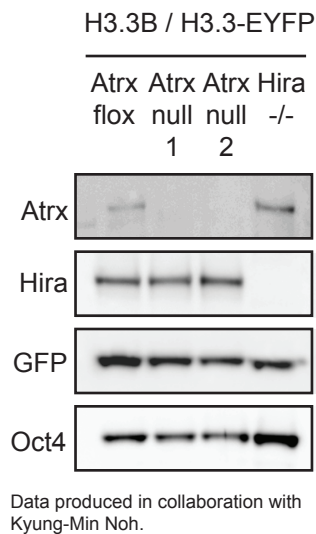


Figure 3.13: Expression of H3.3-EYFP in *Atrx*^{flox}, *Atrx*^{null}, and *Hira*^{-/-} ES cells

Immunoblots showing expression of H3.3-EYFP in *Atrx*^{flox}, *Atrx*^{null}, and *Hira*^{-/-} ES cells, with antibodies as indicated.

Atrx is required to maintain H3.3 deposition at telomeric DNA and for repression of telomeric (TERRA) in ES cells

To determine if Atrx is required for H3.3 enrichment in genes, regulatory regions, or telomeres, I again used ZFNs to knock-in an epitope tag into the endogenous allele of H3.3B, generating heterozygous H3.3B / H3.3B-EYFP in the background of *Atrx*^{flox} and *Atrx*^{null} mouse ES cells (**Figure 3.13**). *Atrx*^{flox} and *Atrx*^{null} ES cells were a

generous gift from the laboratory of Dr. Douglas Higgs and Dr. Richard Gibbons (Garrick et al., 2006). Although *Atrx*^{null} mice die by E9.5 due to

a defect in trophoblast development, Cre-mediated deletion of Atrx from *Atrx*^{flox} ES cells results in viable *Atrx*^{null} ES cells (Garrick et al., 2006). Immunoblots demonstrate comparable expression of H3.3B-EYFP in *Atrx*^{flox} and *Atrx*^{null} ES cells (**Figure 3.13**).

I used native ChIP-seq to generate genome-wide profiles of H3.3 in the presence and absence of Atrx. My collaborators and I find that Atrx is not

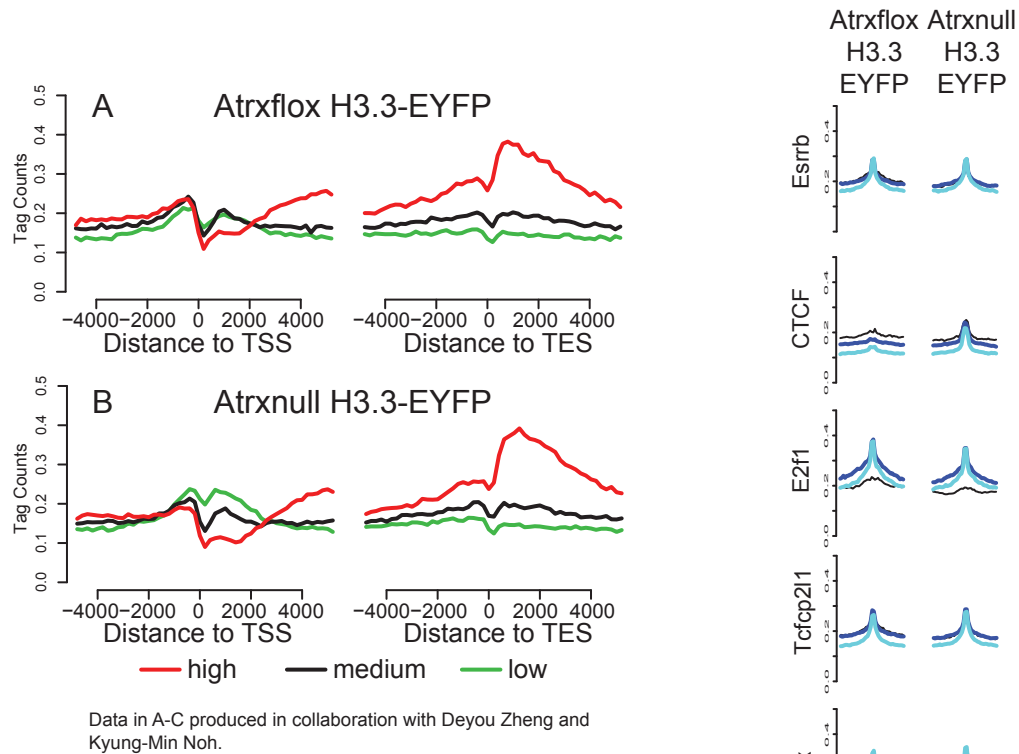


Figure 3.14: Similar genome-wide profiles of H3.3 in Atrxflox and Atrxnull ES cells at genes and TFBS

A-B. Native ChIP-seq profiles of H3.3 in Atrxflox (B) and Atrxnull ES cells (C) as indicated across the TSS and TES for highly active (red), medium expressing (black), or low expressing (green) CpG rich genes.

C. Genome-wide profiles of H3.3 around previously described ES TFBS (Chen et al., 2008), and data analyzed as in 3.7.

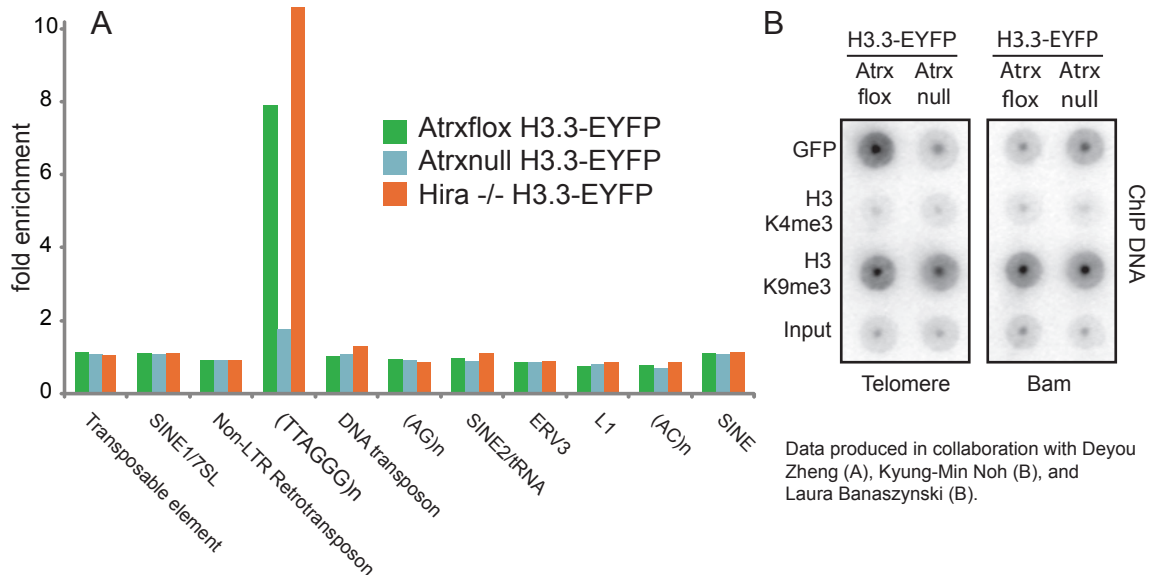


Figure 3.15: Atrx is required for Hira-independent enrichment of H3.3 at telomeres

A. Analysis of H3.3 enrichment at repetitive elements in Atrx^{flox}, Atrx^{null}, and Hira^{-/-} ES cells. Fold enrichment of repetitive elements in ChIP-seq data are plotted over input.

B. Dot blot with Sty11 telomeric probe (left) or control Bam repeat probe (right) of ChIP-seq DNA from Atrx^{flox} H3.3-EYFP or Atrx^{null} H3.3-EYFP IPed with indicated antibody (GFP, H3K4me3, or H3K9me3) or Input DNA.

required for H3.3 incorporation at active or repressed genes (**Figure 3.14A-B**), or at regulatory elements (**Figure 3.8, 3.14C**), as genome-wide profiles of H3.3 are similar at genes and TFBS in Atrx^{flox} and Atrx^{null} ES cells. Strikingly, analysis of repeats demonstrates that Atrx, but not Hira, is specifically required for H3.3 enrichment at telomeres (**Figure 3.15**).

To determine whether there were any functional consequences of Atrx deletion and the loss of Atrx-dependent telomeric enrichment of H3.3, we examined the chromatin state and transcriptional output of ES cell telomeres. The chromatin of ES and induced pluripotent cell telomeres has previously been shown to have lower levels of the heterochromatin marker H3K9me3, increased telomere length, and increased transcription of telomeric repeat-containing RNA (TERRA) in comparison to differentiated cells (Marion et al., 2009) (see **Figure 1.20**). In particular, TERRA

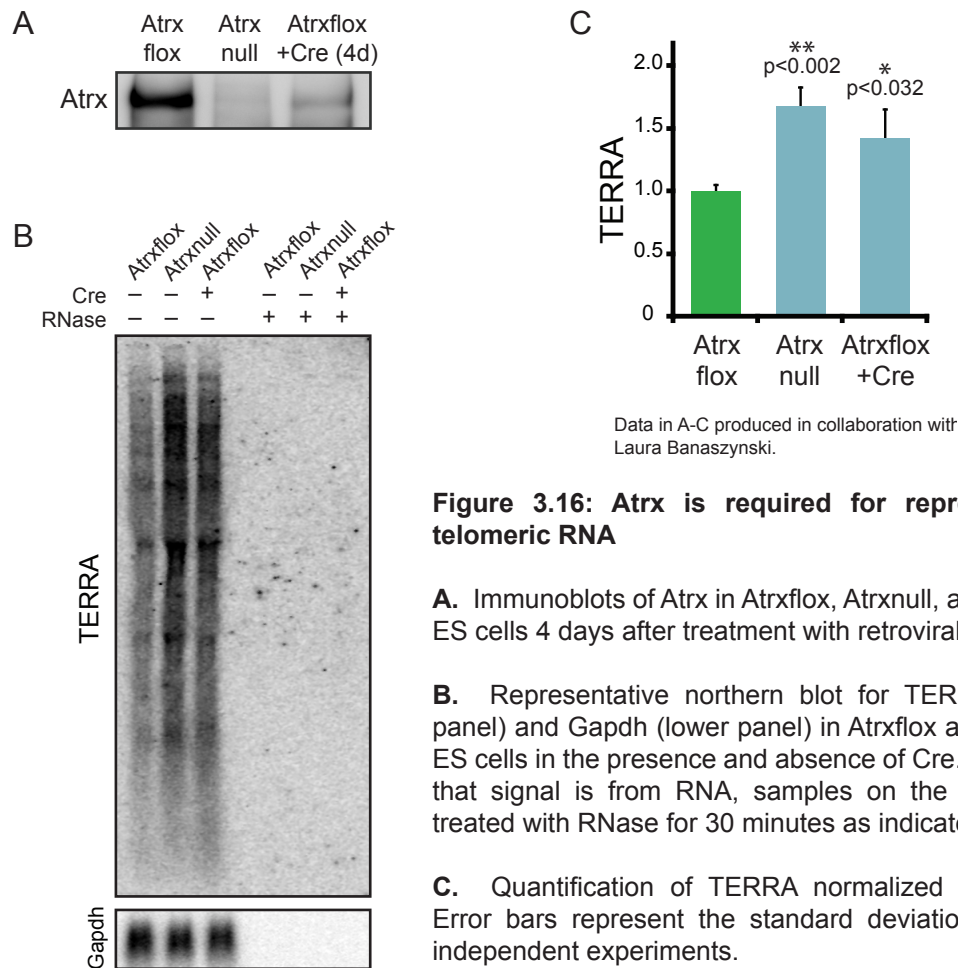


Figure 3.16: Atrx is required for repression of telomeric RNA

A. Immunoblots of Atrx in Atrxflox, Atrxnull, and Atrxflox ES cells 4 days after treatment with retroviral Cre.

B. Representative northern blot for TERRA (upper panel) and Gapdh (lower panel) in Atrxflox and Atrxnull ES cells in the presence and absence of Cre. To ensure that signal is from RNA, samples on the right were treated with RNase for 30 minutes as indicated.

C. Quantification of TERRA normalized to Gapdh. Error bars represent the standard deviation of three independent experiments.

has recently been identified as a component of telomeric heterochromatin and an inhibitor of telomerase, and levels of TERRA have been shown to be regulated by chromatin modifying enzymes (Azzalin and Lingner, 2008; Azzalin et al., 2007; Luke and Lingner, 2009; Schoeftner and Blasco, 2008; Yehezkel et al., 2008). Human ATR-X patients have also previously been shown to have altered DNA methylation in subtelomeric repeats (Gibbons et al., 2000). ChIP of H3K4me3 and H3K9me3 does not show a significant difference in telomeric enrichment between *Atrx*^{flox} and *Atrx*^{null} ES cells (**Figure 3.15B**). However, northern blots from *Atrx*^{flox}, *Atrx*^{null}, and *Atrx*^{flox} ES cells 4 days after treatment with Cre reveal reproducible (~1.7 fold) upregulation of TERRA in the absence of Atrx (**Figure 3.16A-C**).

Atrx is required to maintain H3.3 deposition at ribosomal DNA and for repression of ribosomal RNA in ES cells

In addition to subtelomeric repeats, ATRX patients also show alterations in DNA methylation at rDNA arrays (Gibbons et al., 2000). Genome-wide and immunofluorescence studies in differentiated human cells have previously demonstrated that ATRX is localized to rDNA as well as heterochromatin (Law, 2006; McDowell et al., 1999). To determine whether enrichment of H3.3 at rDNA

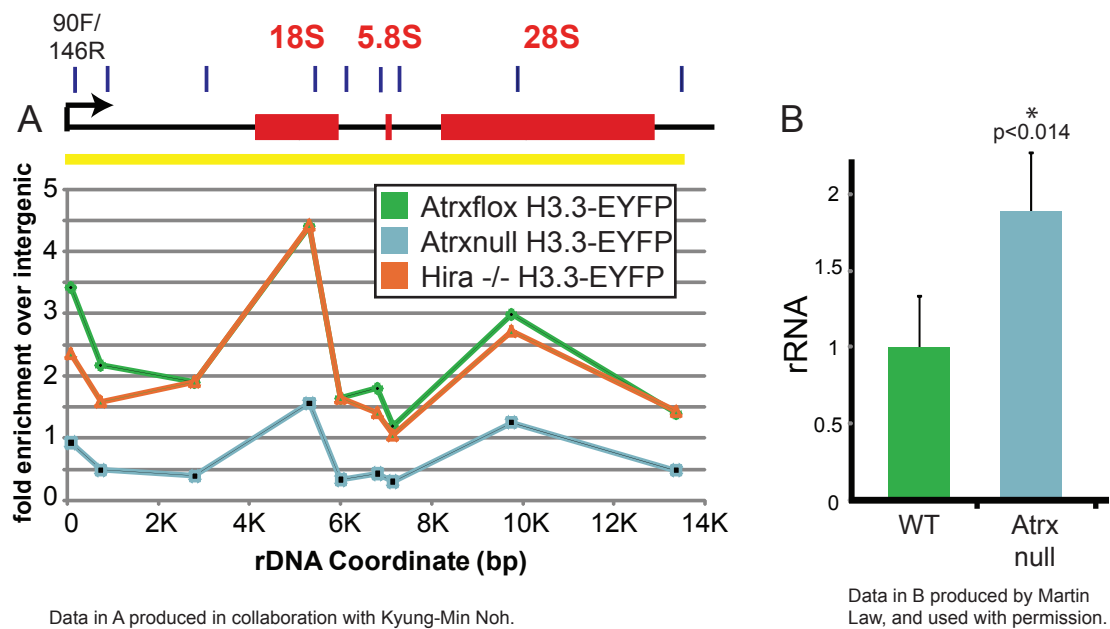


Figure 3.17: Atrx is required for Hira-independent enrichment of H3.3 at ribosomal DNA, and for repression of ribosomal RNA

A. (Top) Scale diagram showing positions of real-time PCR amplicons (blue boxes) across mouse rDNA. Numbering is kilobases from the start of transcription. Coding regions are shown in red. Yellow bar indicates transcribed region. Location of the real-time PCR amplicon used for rDNA primary transcript analysis in H (90F/146R) is indicated. (Bottom) Analysis of H3.3 enrichment at mouse rDNA sequences using real-time PCR of H3.3-EYFP ChIP DNA from Atrxflox (green), Atrxnull (blue) and Hira -/- (orange) ES cells, plotted as fold enrichment over intergenic region.

B. Total RNA extracted from wild-type and Atrxnull ES cells was reverse transcribed and real-time PCR was used to determine levels of ribosomal RNA primary transcript relative to Gapdh mRNA. Error bars represent standard deviation from four ES cell subclones.

might also be dependent on Hira or Atrx, I analyzed ChIP samples from H3.3-EYFP *Atrx^{flox}*, *Atrx^{null}*, and Hira -/- ES cells using real-time PCR primers specific for mouse rDNA. Intriguingly, H3.3 enrichment at the transcribed region of rDNA is also Hira-independent and Atrx-dependent (**Figure 3.17A**). While H3.3 enrichment at rDNA transcribed regions is reduced in *Atrx^{null}* ES cells, levels of rDNA transcripts are also increased in *Atrx^{null}* ES cells (**Figure 3.17B**). Therefore, in both telomeres and rDNA, Atrx is required for H3.3 enrichment in these transcribed repetitive elements, and the transcription of telomeres and rDNA is increased in the absence of Atrx.

ATRX ADD domain interacts with the histone H3 tail

ATRX associates specifically with H3.3 in the presence and absence of Hira (**Figure 3.10-3.12**). In addition, I have shown that ATRX is required for H3.3 deposition at telomeres and rDNA, and for the repression of TERRA and rRNA (**Figure 3.15-3.17**). However, the mechanism by which ATRX directs H3.3 deposition remains unknown. ATRX is a very large protein of approximately 280 kDa, with two highly conserved domains (Gibbons et al., 2008). At its N-terminus, ATRX contains an ATRX-DNMT3-DNMT3L (ADD) domain, a PHD finger-like domain, and at its C-terminus, ATRX contains an ATPase domain with seven highly conserved helicase motifs (Gibbons et al., 2008) (see **Chapter Four** for further discussion of ATRX functional domains). The ADD domain of DNMT3L was recently shown to bind preferentially to the unmodified histone H3 tail (Ooi et al., 2007).

To determine if the ADD domain of ATRX might mediate a direct association between ATRX and histone H3.3, I performed peptide pull-down studies with the recombinant ATRX ADD domain (residues 159–296) (Argentaro et al., 2007) and biotinylated histone peptides (**Figure 3.18**). As controls, I used recombinant DNMT3L (Ooi et al., 2007), as well as WDR5, which has been shown to preferentially bind to H3K4me2 (Ruthenburg et al., 2006; Wysocka et al., 2005). I found that the

ATRX ADD domain interacts specifically with the histone H3 tail, amino acids 1-20, similar to the described interaction of H3 1-20 with DNMT3L (Ooi et al., 2007), (**Figure 3.18A**). As DNMT3L shows a strong preference for H3 1-20 unmethylated at H3K4, I used histone H3 tail peptides methylated at K4 and K9 to determine if the ATRX ADD domain displays a similar preference. I found that the ATRX ADD domain shows a slight preference for H3K9 methylation, with increasing binding with

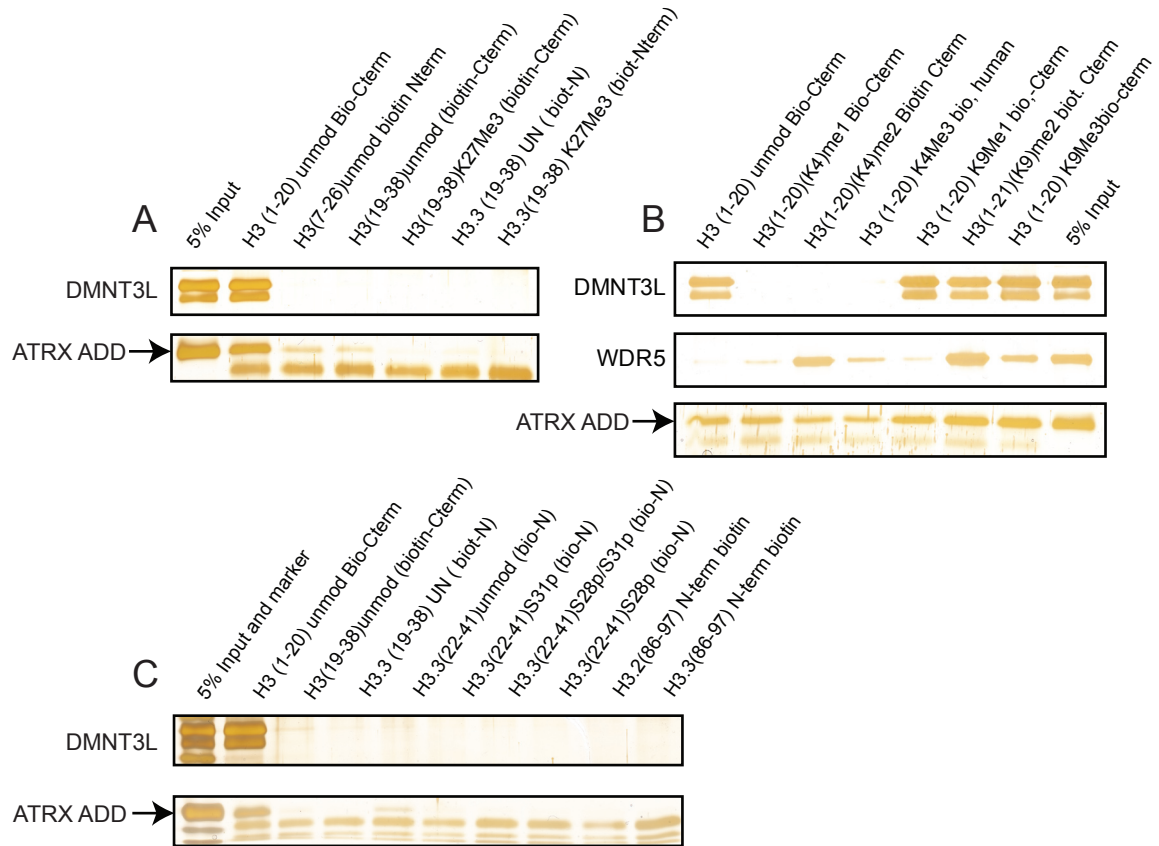


Figure 3.18: ATRX ADD domain interacts with H3 1-20, and does not show a preference for H3.3 specific sequences

A-C. Peptide interaction assays with modified or unmodified biotinylated histone H3 peptides as indicated, incubated with the human ATRX ADD domain, or full-length DNMT3L and WDR5 as controls.

A. The ATRX ADD domain prefers the N-terminal 1-20 amino acids of H3.

B. The ATRX ADD domain interacts with the N-terminus of H3, is relatively unaffected by H3K4 methylation, and displays a slight preference for H3K9 methylation.

C. The ATRX ADD domain does not prefer H3.3-specific peptides, but interacts with the N-terminal 1-20 amino acids of H3.

higher H3K9 methylation states (**Figure 3.18B**). DNMT3L binding was completely abolished by H3K4 methylation, as previously reported (Ooi et al., 2007), while a fraction of the ATRX ADD domain still bound to H3 1-20 peptides methylated at H3K4 (**Figure 3.18B**). Recent unpublished studies from other laboratories confirm that the ATRX ADD domain binds to unmodified H3 tails, and this binding is enhanced by di and tri-methylation of H3K9 (D. Neuhaus, D. Rhodes, Y. Shi, personal communication). However, as described previously, the histone H3 tail residues 1-20 are identical in all non-centromeric H3 variants (**Figure 1.16**), and this interaction therefore does not demonstrate why I have seen such a selective association of ATRX with H3.3.

The amino acids that distinguish H3.3 from H3.1/H3.2 are located in the N-terminal tail at position 31, as well as in a core region at 87, 90, 90, and 96 (**Figure 1.16**). Therefore, I also used peptide sequences spanning these regions to determine if the ATRX ADD domain might show any H3 variant-specific preference. Disappointingly, the ATRX ADD domain did not interact significantly with H3 19-38 or 22-41 from H3.3 or H3.2/H3.1, or with 86-97 from H3.3 or H3.2 (**Figure 3.18B**). Therefore, the mechanism by which ATRX associates with and specifically deposits H3.3 remains to be established.

A new perspective on H3.3 deposition pathways

The studies described in this chapter make clear that H3.3 deposition in mammalian cells is significantly more complicated than current models suggested by the field (De Koning et al., 2007; Loyola and Almouzni, 2007) (**Figure 1.22**). Although Hira plays an essential role for H3.3 deposition at genes and multiple TFBS, it is only responsible for a subset of genome-wide H3.3 deposition patterns in mammalian cells. Deposition of H3.3 at specific TFBS, telomeres, and rDNA is Hira-independent. In addition, my work has identified two novel players in H3.3 deposition: ATRX

and Daxx. We have seen that ATRX is required for H3.3 deposition at telomeres and rDNA, and for repression of TERRA and rRNA transcription. Although many questions remain, the data in this chapter provide a new perspective on the complexity of genomic region-specific H3.3 deposition in mammalian cells. In the following chapter, I will discuss the implications of my work, describe ongoing projects and future studies, and discuss the potential functions of H3.3 deposition in mammalian development and cellular differentiation.

CHAPTER 4

GENERAL DISCUSSION, WORK IN PROGRESS, AND FUTURE DIRECTIONS

The data presented in the previous chapters represent the first comprehensive evaluation of non-centromeric H3 variant localization in undifferentiated mammalian embryonic stem (ES) cells and differentiated neuronal precursor cells (NPCs).

Chromatin immunoprecipitation (ChIP) grade antibodies that distinguish H3.3 from the other H3 variants are not available. Therefore, in Chapter Two, I introduced a system for the rapid knock-in of epitope tag sequences into the endogenous histone H3.3B gene in mouse ES cells. Using these genetically engineered ES cells in combination with point mutagenesis of H3.3B, I have found compelling evidence that genome-wide patterns of H3.3 are dependent on H3.3-specific amino acid sequence. H3.3 displays specific patterns of incorporation in the mammalian genome, while H3.2 and H3.1 are much more widely distributed. Following the differentiation of ES cells into NPCs, I demonstrated that patterns of H3.3 enrichment at cell-type specific genes and transcription factor binding sites (TFBS) are dependent on cellular state. Overall, the data in this thesis identify three general categories of H3.3 enrichment in mammalian cells: 1) genes and other transcribed non-repetitive sequences, 2) TFBS, and 3) ribosomal DNA (rDNA) and telomeres. Remarkably, as shown in Chapter Three, each of these general categories of H3.3 enrichment in ES cells are mediated by distinct mechanisms of deposition.

In Chapter Three, I examined genome-wide patterns of H3.3 deposition in the presence and absence of the histone chaperone Hira. As expected, the H3.3 chaperone Hira is required for enrichment of H3.3 at active and repressed

genes. Unexpectedly, deposition of H3.3 at specific TFBS, rDNA, and telomeres is Hira-independent. Immunoaffinity purification of H3.3 associated proteins in the presence and absence of Hira revealed a specific, Hira-independent association of H3.3 with the chromatin remodeling protein Atrx (alpha thalassemia and X-linked mental retardation) and the death domain associated protein Daxx. Using Atrxflox and Atrxnull ES cells, my collaborators and I found that Atrx is specifically required

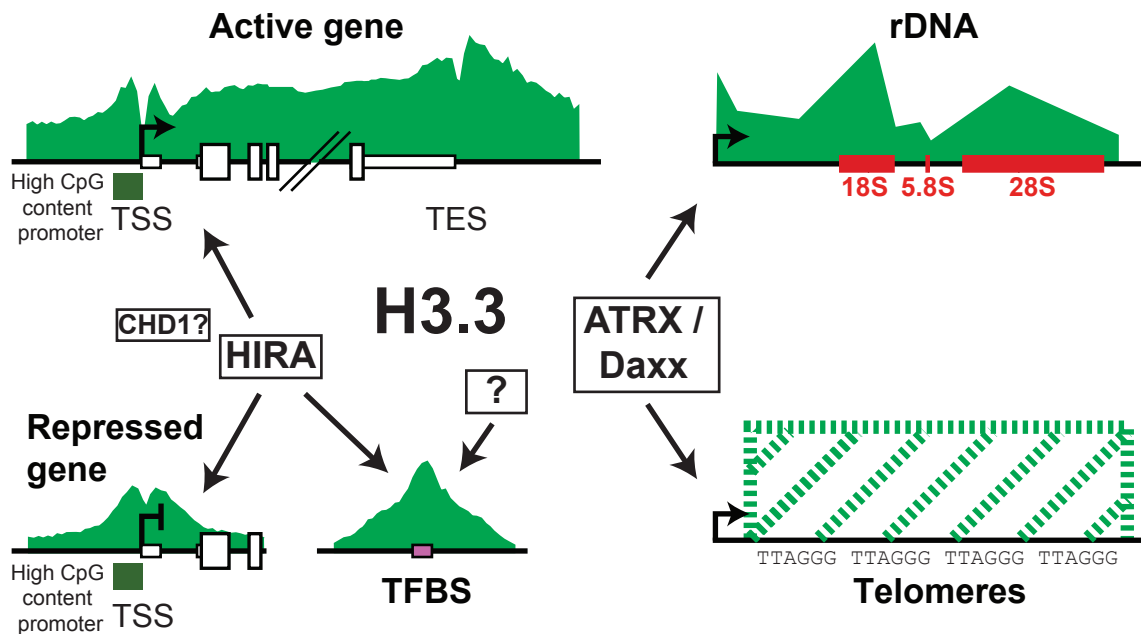


Figure 4.1: Hira and Atrx are required for deposition of H3.3 at distinct genomic regions

The patterns and levels of H3.3 enrichment at specific elements are represented by green profiles. At highly repetitive elements with short repeat length such as telomeres, the precise pattern of H3.3 across the element cannot be determined, and the relative enrichment is indicated by a green hashed box.

H3.3 is enriched in specific patterns at active and repressed genes with high CpG content promoters (HCP), genic and intergenic transcription factor binding sites (TFBS), ribosomal DNA (rDNA), and telomeres. TSS = transcription start site, TES = transcription end site. The H3.3 chaperone HIRA is required for H3.3 enrichment at active genes, repressed genes, and specific TFBS. Based on the described localization of CHD1 to transcription start sites in embryonic stem cells (Gaspar-Maia et al., 2009), and the cooperative role of Hira and CHD1 in H3.3 nucleosome assembly in *Drosophila* (Konev et al., 2007), it is possible that CHD1 may also play a role in H3.3 deposition at these locations, but this possibility has not been examined. Surprisingly, H3.3 enrichment at specific TFBS, rDNA, and telomeres is Hira-independent. The boxed question mark denotes as yet unidentified factors that are required for Hira-independent and Atrx-independent H3.3 deposition at specific TFBS. ATRX is required for H3.3 enrichment at rDNA and telomeres. Although Daxx specifically associates with both H3.3 and Atrx, the role of Daxx in H3.3 deposition has not yet been established. Note that the profiles shown in this chapter are generalized summary depictions. For actual ChIP-seq data, see Chapters Two and Three.

for H3.3 deposition at telomeres and rDNA.

Overall, these data demonstrate that mammalian cells utilize distinct mechanisms to deposit histone H3.3 at different genomic locations (**Figure 4.1**). As presented in Chapter One, the chromatin field generally accepts a model in which

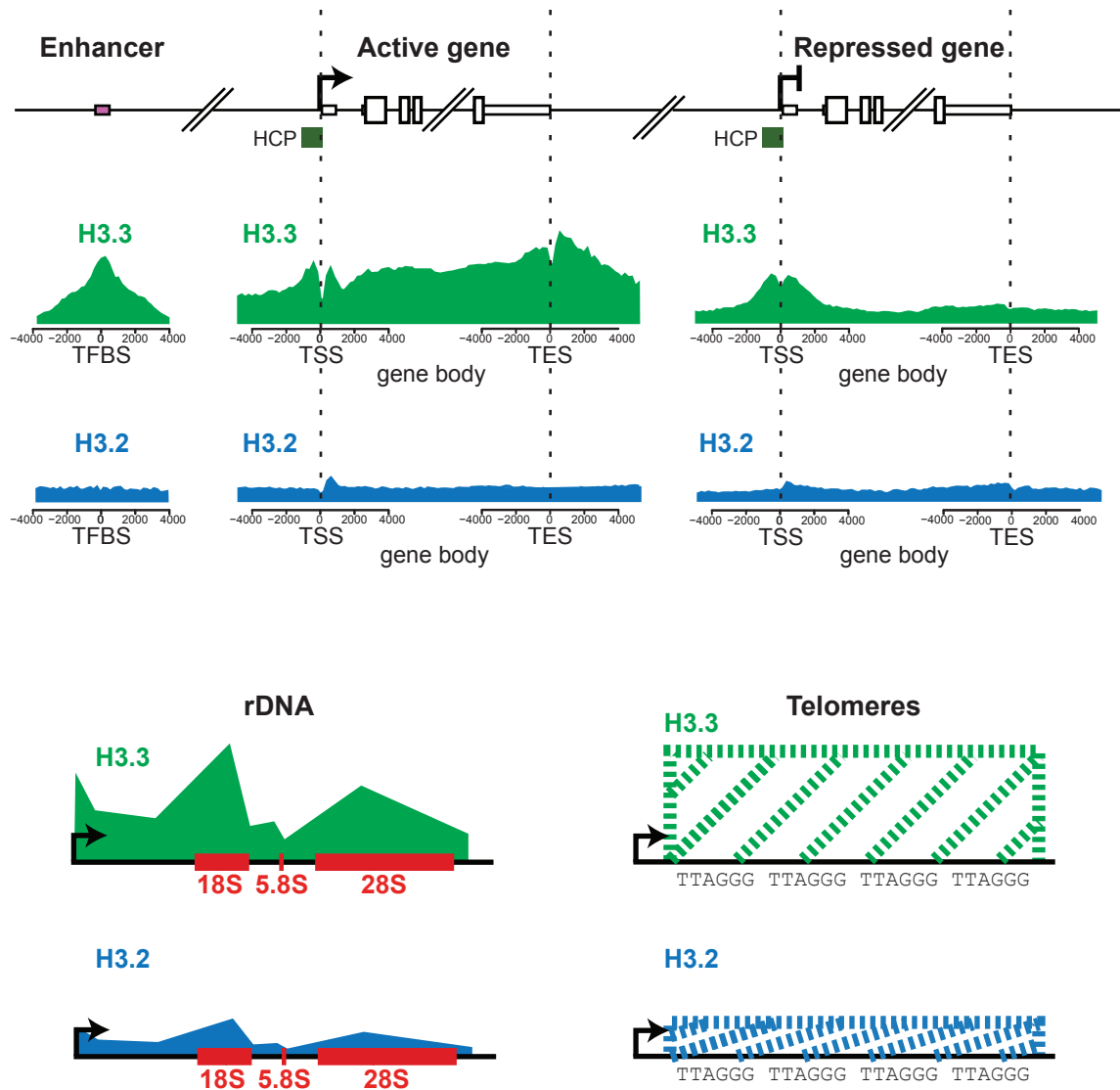


Figure 4.2: Genome-wide localization of H3.3 requires H3.3-specific amino acid sequence

Mutation of endogenous H3.3B to H3.2 or towards H3.1 abolishes H3.3 enrichment at TFBS, active and repressed genes, rDNA, and telomeres. Data is summarized in a manner similar to the previous figure.

Hira is the sole H3.3-specific chaperone required for H3.3 deposition. Therefore, the results presented in this thesis provide a significant conceptual advance to this overly simplistic paradigm. The observation that Atrx is required for H3.3 deposition at telomeres and rDNA opens up a new area of investigation regarding the mechanisms, recruitment, and function of genomic region-specific histone deposition machinery. More speculatively, these results may suggest potential connections between H3.3 deposition and human genetic disease. Below, I will discuss some of the implications of our findings, describe some current work in progress, and suggest some possibilities for future studies.

Genomic localization of H3.3 is specific and requires H3.3 amino acid sequence

The data presented in Chapter Two demonstrate that H3.3 is enriched in specific patterns at active and repressed coding and non-coding genes, including lincRNA (Guttman et al., 2009), as well as TFBS, ribosomal DNA, and telomeres (**Figure 4.1**). Using point mutagenesis of the endogenous H3.3B gene, I provided strong evidence that this localization pattern requires H3.3-specific amino acid sequence. Expression of a non-H3.3 variant from the H3.3 promoter is insufficient for proper targeting and localization of H3.3, as genome-wide H3.3 patterns are abolished upon the mutation of H3.3B to H3.2 or H3.1 (**Figure 4.2**). Notably, I have not performed the reciprocal experiment that places H3.3 under a replication-dependent H3.1 or H3.2 promoter. Therefore, it remains possible that the timing of H3 variant expression may still play an important role in H3 variant localization if the H3 variant is only expressed during the defined window of S-phase. However, the results presented in this thesis clearly show that the sequence of the H3 variant dictates genome-wide localization when expressed from the endogenous H3.3B promoter (**Figure 4.2**).

Constitutive incorporation of H3.3 at active and repressed genes with CpG rich promoters

In *Drosophila*, H3.3 is constitutively enriched around poised HSP70 promoters, even in the absence of heat shock (Mito et al., 2007). However, in previous studies of vertebrates, the profile of H3.3 at repressed genes differed depending on the system being studied (see **Chapter One**) (Jin and Felsenfeld, 2006; Jin et al., 2009; Sutcliffe et al., 2009).

Our data show that genome-wide patterns of H3.3 around TSS in mammalian ES cells and NPCs vary significantly based on the CpG content of gene promoters. H3.3 is constitutively enriched around the TSS of both active and repressed high CpG content promoter (HCP) genes in ES cells and NPCs, including the TSS of bivalent genes in ES cells (**Figure 4.2**). In contrast to HCP genes, we do not observe any significant pattern of H3 variant enrichment at low CpG content promoter (LCP) genes in ES cells and NPCs (data not shown).

LCP and HCP genes have been described to display distinct modes of regulation (Mikkelsen et al., 2007; Ramirez-Carrozzi et al., 2009; Saxonov et al., 2006). Most HCP genes in both undifferentiated and differentiated cells show evidence of transcriptional initiation, assemble unstable nucleosomes, and do not require SWI/SNF nucleosome remodeling complexes for gene induction, while LCP genes assemble stable nucleosomes and require SWI/SNF (Guenther et al., 2007; Ramirez-Carrozzi et al., 2009). Indeed, nearly all (99%) of HCP genes are marked by H3K4me3 in mouse ES cells, whether they are transcriptionally active or repressed, while only 6.5% of LCP genes in ES cells carry H3K4me3 (Mikkelsen et al., 2007). Our results are therefore consistent with a model in which H3K4 methylation and H3.3 deposition at HCP TSS are coupled to transcriptional initiation.

Gene body deposition and transcription end site peaks of H3.3 in mammalian cells are proportional to transcriptional activity

In addition to constitutive TSS deposition of H3.3 at HCP genes, we find enrichment of H3.3 in the gene body and after the transcription end site (TES) in proportion to transcriptional activity (**Figure 4.2**). These gene body and TES patterns of H3.3 are distinct from other histone modifications. While H3.3 and H3K36me3 are strongly correlated in gene bodies, H3K36me3 begins to decline prior to the TES at highly expressed genes (see **Figure 1.10**), while levels of H3.3 and Ser-5 phosphorylated RNA polymerase II (RNAPII) increase and are enriched in a specific peak beyond the TES of highly expressed genes (**Figure 4.2**).

Significant peaks of H3.3 were not initially observed after the 3' ends of genes in *Drosophila* (Mito et al., 2005, 2007). However, a more recent report found a peak of H3.3 at genic 3' ends of both active and inactive *Drosophila* genes, particularly from chromatin fractions extracted in low-salt (Henikoff et al., 2009). Interestingly, a dramatic enrichment of trypanosome-specific histone variants H3V and H4V was recently observed at the probable RNAPII transcriptional termination sites of polycistronic transcription units (Siegel et al., 2009), leading to the proposal that “chromatin-based transcriptional punctuation” in trypanosomes may represent an evolutionary adaptation to regulate transcription in eukaryotes that lack sophisticated regulatory sequences (Talbert and Henikoff, 2009). In addition, TES peaks of H3.3 were also observed in a recent genome-wide study of HeLa cells (Jin et al., 2009). Here, I have shown that “transcriptional punctuation” by histone variant H3.3 and phosphorylated RNA polymerase also marks the boundaries of highly expressed genes in both undifferentiated and differentiated mammalian cells, calling attention to a potentially more universal mechanism for histone variant utilization as a genomic “boundary marker.”

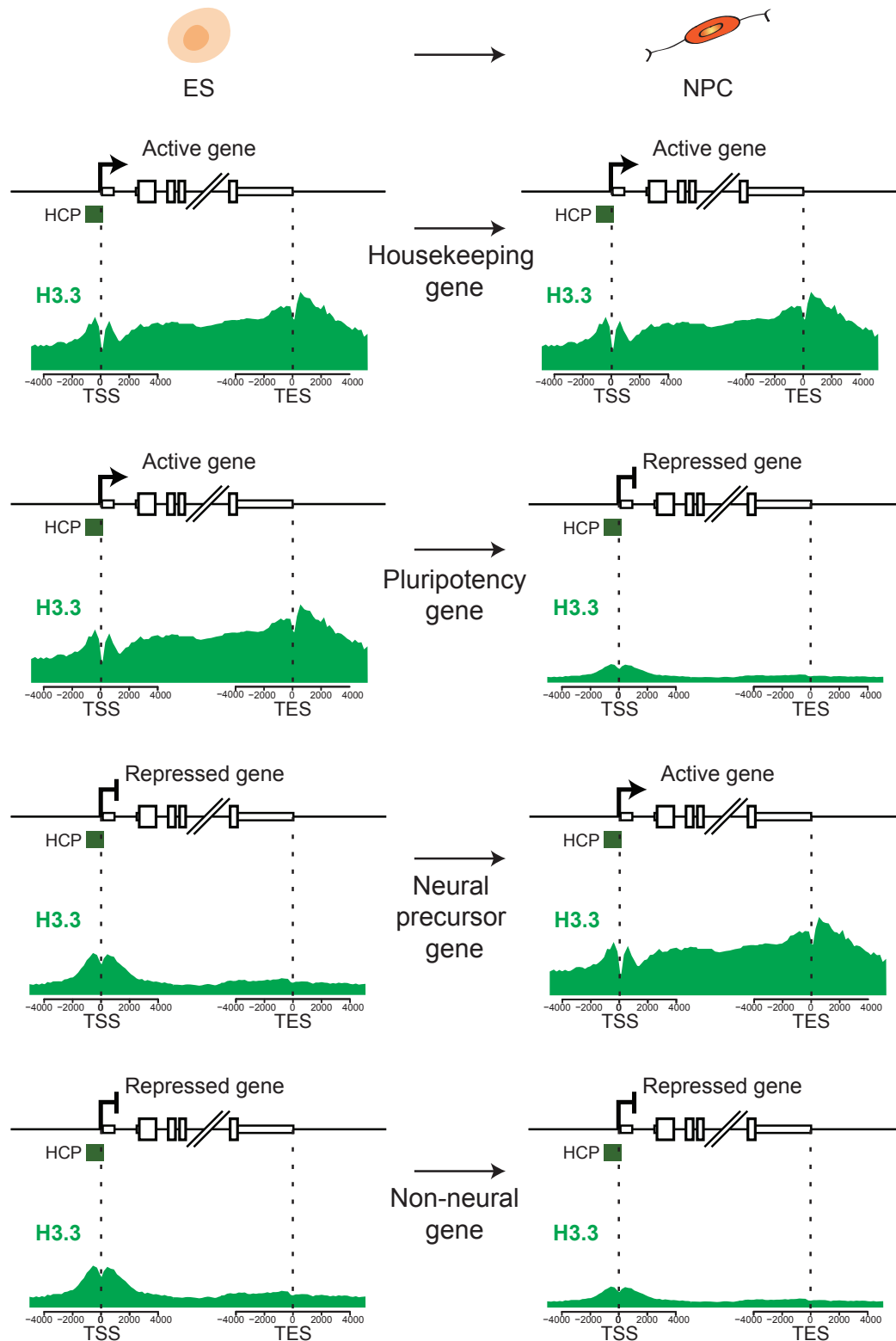


Figure 4.3: H3.3 localization at cell-type specific genes changes with cell differentiation

Representative ES and NPC profiles for H3.3 are depicted for housekeeping genes, pluripotency genes, neural precursor genes, and non-neural genes with HCP promoters.

What is the function of H3.3 deposition after transcriptional end sites? Does genic and post-TES deposition of H3.3 simply demarcate active genes, or does H3.3 play a more active role? Previous studies suggest that H3.3 deposition in actively transcribed gene bodies may be coupled to transcription, potentially mediated by factors associated with elongating polymerase (Daury et al., 2006; Janicki et al., 2004; Schwartz and Ahmad, 2005; Tagami et al., 2004; Wirbelauer et al., 2005). Recent studies in human cells documented accumulation of RNAPII beyond the site of RNA transcript cleavage and downstream of H3K36me3, suggesting that paused RNAPII may accumulate prior to release at the end of genes (Brodsky et al., 2005; Lian et al., 2008). One can speculate that the colocalized post-TES

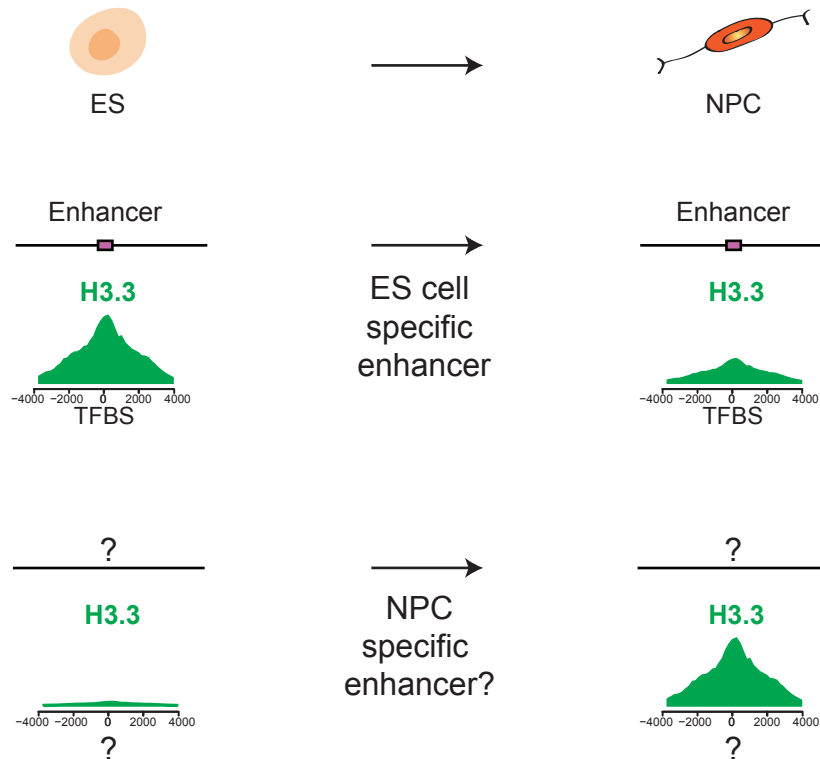


Figure 4.4: H3.3 localization at cell-type regulatory elements changes with cell differentiation

H3.3 enrichment at ES cell specific enhancers is decreased in NPCs (top). In addition, numerous unannotated cell-type specific peaks were found, and these represent candidates for possible cell-type specific regulatory elements.

peaks of phosphorylated RNAPII and H3.3 at highly active genes in mammalian cells represent a site of significant chromatin disruption, as transcription-coupled deposition of H3.3 “builds up” with paused RNAPII prior to polymerase release. It is also possible that peaks of significant H3.3 replacement may play an active role in slowing the extension of RNAPII to facilitate transcriptional termination, as nucleosomes have been shown to serve as a barrier to transcriptional elongation (Armstrong, 2007).

H3.3 localization changes at specific loci following embryonic stem cell differentiation

I chose ES cells as a model system because they would enable me to study how patterns of H3 variants change during mammalian cellular differentiation. I have found that the overall pattern of H3.3 enrichment remains consistent in ES cells and NPCs, with H3.3 enriched at the TSS of active and inactive genes, within active gene bodies and after active TES, as well as TFBS, rDNA, and telomeres (**Figure 4.3**). Notably, my collaborators and I found remarkable genome-wide changes of H3.3 localization at cell-type specific genes and TFBS, including pluripotency genes, differentiation-specific genes, and known regulatory regions (**Figure 4.3, 4.4**). In addition, we also found multiple cell-type specific sites of H3.3 and H3K4me1 enrichment that do not correspond to annotated genes or TFBS (**Figure 4.4**). When combined with future datasets, it is possible that our data will yield additional insights, perhaps as a means of identifying novel ES or NPC regulatory and transcribed regions. A similar approach recently used genome-wide patterns of H3K4me3 and H3K36me3 to identify novel mammalian lincRNAs (Guttman et al., 2009; Mikkelsen et al., 2007), and genome-wide patterns of p300 to identify tissue-specific enhancers (Visel et al., 2009).

Comparison of H3.3 and H2A.Z localization in embryonic stem cells and neural precursor cells

As described in Chapter One, the histone variants H3.3 and H2A.Z have been found to co-occupy nucleosomes at specific regions of the genome, and such H3.3/H2A.Z nucleosomes were found to be relatively unstable *in vitro* (Jin and Felsenfeld, 2006, 2007; Jin et al., 2009). H3.3 was previously found to be co-enriched with H2A.Z at active promoters and enhancers of differentiated chicken and human cells (Jin and Felsenfeld, 2007; Jin et al., 2009). However, the

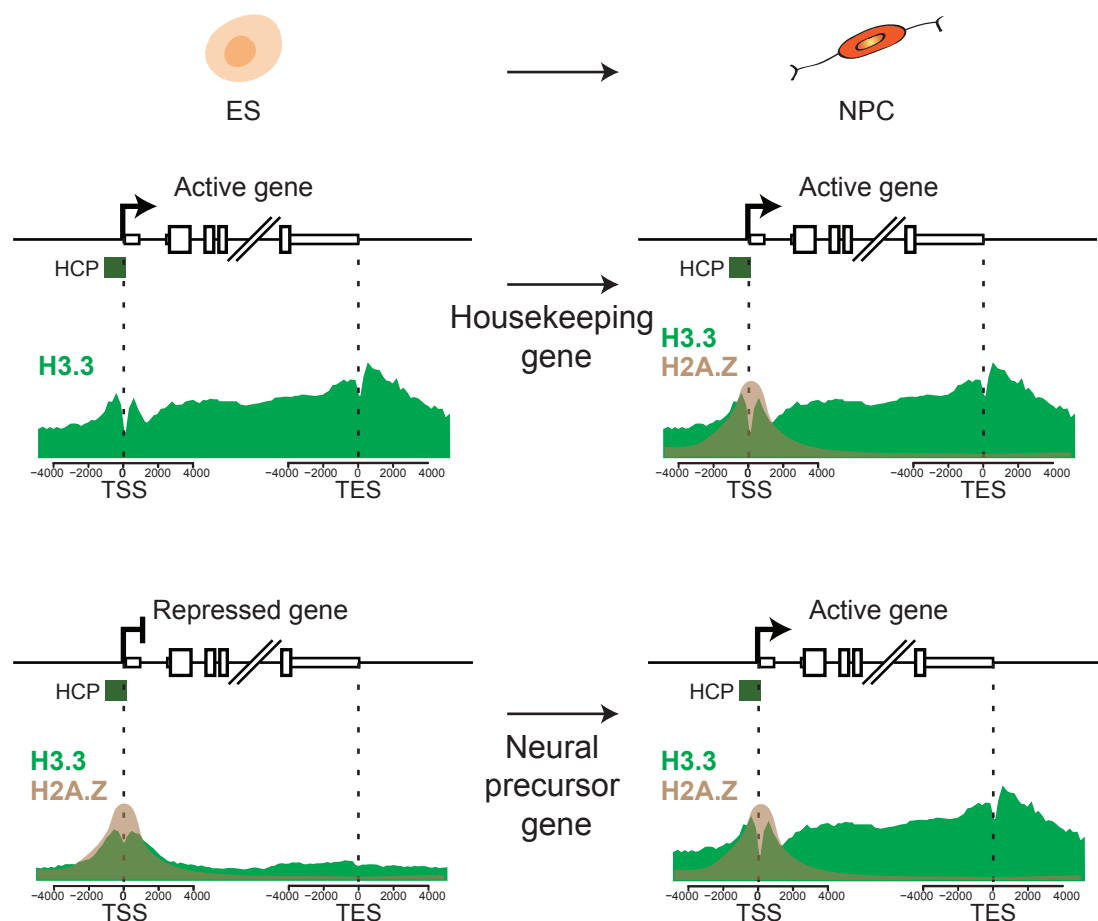


Figure 4.5: Predicted genic locations of H3.3 / H2A.Z nucleosomes in ES cells and NPCs based on current study and reported literature

Genome-scale patterns of H2A.Z in ES cells and NPCs have been previously reported (Creyghton et al., 2008). Based on this published data and on the data presented in this thesis, H3.3 / H2A.Z nucleosomes can be predicted to co-exist at the TSS of repressed genes in ES cells, but active genes in NPCs.

localization of H3.3/H2A.Z nucleosomes has not been previously reported in ES cells. Interestingly, the data I have presented in Chapter Two shows that genome-wide patterns of H3.3 in ES cells differ significantly from the described patterns of H2A.Z (Creyghton et al., 2008).

In ES cells, H2A.Z is enriched in the promoters of repressed polycomb target genes, but upon differentiation to NPCs, H2A.Z becomes enriched at the promoters of highly expressed genes (Creyghton et al., 2008). I have shown that H3.3 is enriched more widely than the reported patterns of H2A.Z in both ES cells and NPCs, as H3.3 is enriched around the TSS of both active and repressed genes, in addition to expressed gene bodies, TFBS, and telomeres. Moreover, while H3.3 enrichment changes at specific locations upon cell differentiation, the overall relationship between H3.3 enrichment and gene expression remains similar in undifferentiated and differentiated cells. In both ES cells and NPCs, the greatest enrichment of H3.3 is found at highly expressed metabolic and housekeeping genes (**Figure 2.27**). If we integrate the previous results of Boyer, Creyghton, and colleagues with the data presented here, one potential implication is that destabilizing H3.3/H2A.Z nucleosomes in ES cells would be present specifically at the TSS of repressed, poised genes, rather than at active housekeeping genes (**Figure 4.5**). It is possible that such cell-type specific localization of destabilizing nucleosomes may be important to prepare developmental genes in ES cells for future expression. However, no one has yet demonstrated the genome-wide localization of H3.3/H2A.Z nucleosomes in any mammalian cell type other than HeLa cells (Jin et al., 2009). Therefore, the first step would be to confirm the existence and localization of H3.3/H2A.Z nucleosomes in ES cells and NPCs, an experiment which would now be possible by performing a double anti-HA and anti-H2A.Z ChIP-seq from the H3.3B / H3.3B-HA ES cells described in this thesis.

Hira-dependent enrichment of H3.3 at active and repressed genes

The data described in Chapter Three represent the first genome-wide study to compare chromatin in the presence and absence of a mammalian histone chaperone. Previous biochemical experiments showed that Hira specifically associates with H3.3 in mammalian cells, and *in vivo* experiments determined that Hira is required for H3.3 deposition and chromatin assembly in the *Drosophila* male pronucleus (Bonney et al., 2007; Loppin et al., 2005; Tagami et al., 2004). In addition, biochemical and fluorescence recovery after photobleaching (FRAP) studies in wild-type and Hira $-/-$ ES cells showed global increases in unbound H3 in the absence of Hira (Meshorer et al., 2006). From these studies, one would expect that Hira would play an important role in H3.3 deposition. However, although these experiments were suggestive, the data in this thesis are the first to directly show by ChIP that Hira is required for genome-wide H3.3 enrichment at active and

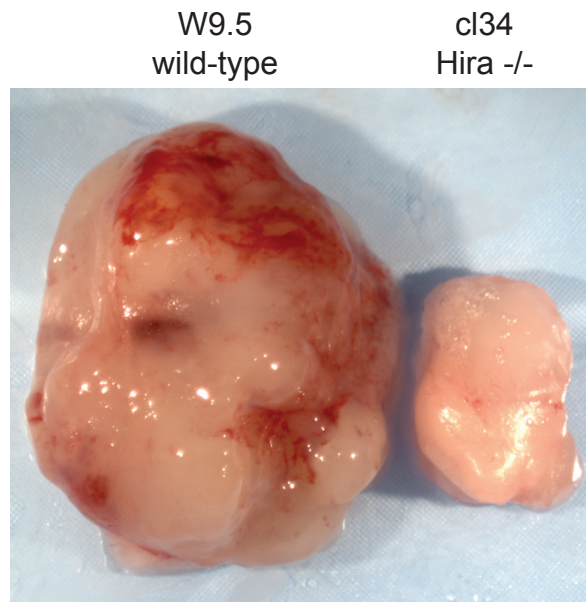


Figure 4.6: Hira $-/-$ ES cells are defective in teratoma formation

Teratomas from Hira $-/-$ ES cells are significantly smaller than teratomas from wild-type ES cells. 5×10^6 W9.5 or Hira $-/-$ ES cells were injected into NOD-SCID mice, and teratomas were harvested at day 21. Representative teratomas are shown. Experiments were performed in collaboration with Jiyeon Kim, Dr. Marco Seandel, and Dr. Shahin Rafii.

repressed genes. How is Hira targeted to these specific genomic regions?

Previous studies hypothesize that H3.3 deposition in actively transcribed gene bodies may be coupled to transcription, potentially mediated by factors associated with elongating polymerase (Daury et al., 2006; Janicki et al., 2004;

Schwartz and Ahmad, 2005; Tagami et al., 2004; Wirbelauer et al., 2005). Our data are consistent with Hira-dependent transcription-coupled deposition of H3.3 at transcribed non-repetitive sequences. However, Hira has not been found in purifications of the general transcriptional machinery, and the likely association between Hira and RNAPII would therefore be mediated through factors that have yet to be determined. Much further work is needed on the recruitment of Hira to active and inactive genes.

Hira-mediated deposition of H3.3 and embryonic stem cell differentiation

Remarkably, although Hira is required for early embryonic development (Roberts et al., 2002), my collaborators and I do not observe a significant phenotype in Hira $-/-$ ES cells, despite a global lack of H3.3 enrichment at active and repressed genes. However, published experiments (Meshorer et al., 2006; Roberts et al., 2002) as well as recent collaborative data indicates that despite their maintenance of the ES cell phenotype in culture, Hira $-/-$ ES cells are significantly defective in cell differentiation.

One test of pluripotency involves the injection of cells into immunocompromised mice. Following injection into non-obese diabetic severe combined immunodeficiency (NOD-SCID) mice, pluripotent cells such as ES and induced pluripotent stem (iPS) cells will differentiate into multi-lineage tumors known as teratomas (Takahashi and Yamanaka, 2006). In collaborative experiments with Jiyeon Kim, Dr. Marco Seandel, and Dr. Shahin Rafii, wild-type and Hira $-/-$ ES cells were injected into NOD-SCID mice. Examination of the resulting teratomas revealed that Hira $-/-$ teratomas were remarkably smaller than wild-type teratomas (**Figure 4.6**). It is possible that this defect in teratoma formation could simply be due to slower growth of Hira $-/-$ cells. However, at least in culture, Hira $-/-$ ES cells divide as rapidly as wild-type ES cells, and show a similar preponderance of cells

in S phase (Ariane Chappier and Peter J. Scambler, unpublished data).

Intriguingly, preliminary histological analysis indicated that unlike wild-type teratomas, Hira $-/-$ teratomas failed to form fully differentiated tissue such as cartilage (data not shown). This data is also consistent with the reported inability of Hira $-/-$ ES cells to fully differentiate into post-mitotic neurons (Meshorer et al., 2006). However, it must be noted that these experiments are preliminary, and much work will be required to explore the function of Hira in mammalian cell differentiation.

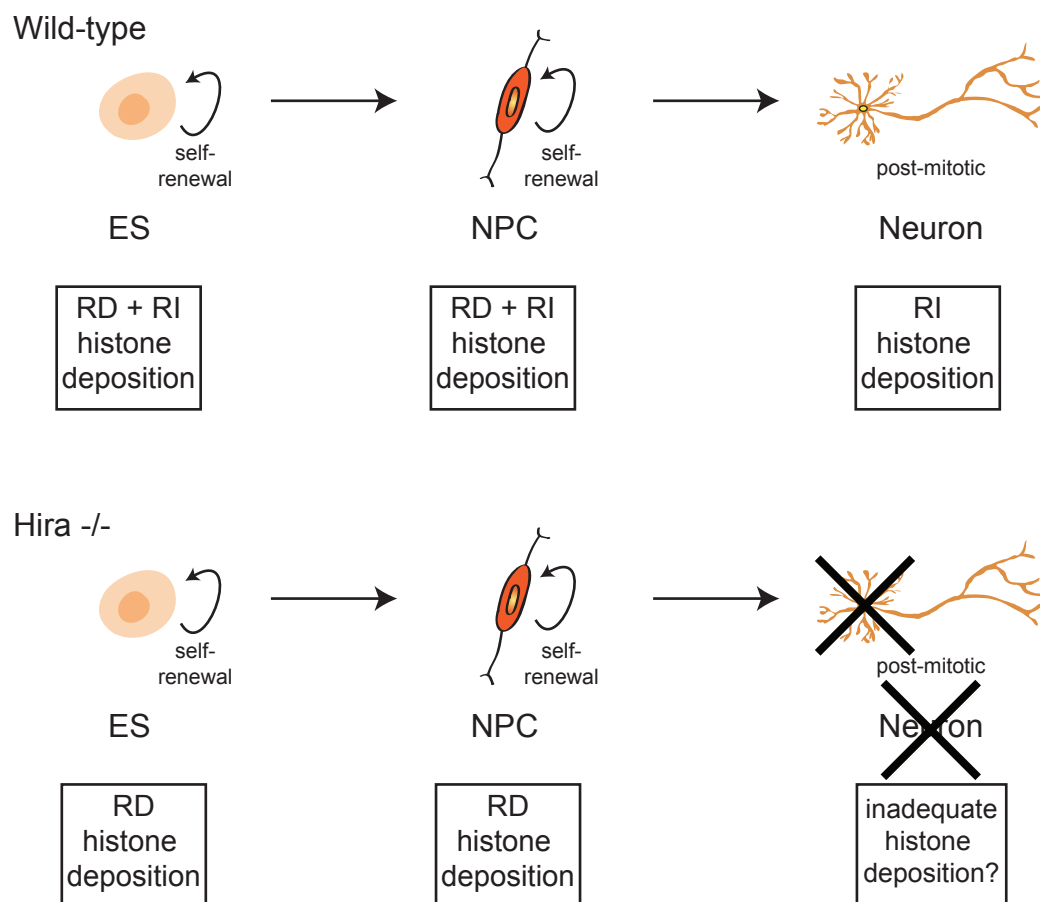


Figure 4.7: Defects in Hira $-/-$ ES cells upon cell differentiation may be due to impaired replication-independent histone deposition

Rapidly cycling cells such as ES cells and NPCs are competent for both replication-dependent (RD) and replication-independent (RI) histone deposition. However, post-mitotic neurons are dependent upon RI histone deposition. In Hira $-/-$ cells (bottom), rapidly cycling ES cells and NPCs may be rescued by RD histone deposition during frequent S-phases. However, Hira $-/-$ ES cells do not differentiate into post-mitotic neurons (Meshorer et al., 2006), and it is possible that such post-mitotic cells may be unable to compensate for the loss of Hira-mediated RI histone deposition.

What might account for the differentiation-specific phenotype of Hira $-/-$ cells?

As both Hira $-/-$ ES cells and Hira $-/-$ embryos display significant patterning abnormalities following differentiation (Meshorer et al., 2006; Roberts et al., 2002), I speculate that rapidly dividing Hira $-/-$ ES cells may be rescued by the replication-coupled deposition of histones during the frequent S-phases of ES cells (**Figure 4.7**). ES cells cycle very rapidly, have a short G1 phase, and lack a G1/S checkpoint (Burdon et al., 2002). It is possible that in the absence of Hira-dependent incorporation of genic H3.3, differentiating cells with slower cell cycles may be deficient in replacing histones at active or initiating gene loci. Without H3.3 or sufficient localized replication-dependent H3.1 or H3.2 as substrates for specific post-translational modifications (K4 and K27 methylation) to mark particular loci, more slowly-dividing or terminally differentiated cells may become progressively incapable of epigenetic memory.

This hypothesis regarding the developmental importance of RI dependent H3.3 deposition by Hira fits with the patterning abnormalities and early lethality of Hira $-/-$ embryos, the inability of Hira $-/-$ ES cells to undergo terminal neuronal differentiation (Meshorer et al., 2006), the neuromuscular defects seen in H3.3A hypomorphs (Couldrey et al., 1999), and with earlier descriptions of H3.3 in the literature, in which H3.3 RNA and protein levels were observed to increase during models of cell differentiation (Krimer et al., 1993; Lord et al., 1990; Wunsch and Lough, 1987). High levels of H3.3 protein have been found in differentiating rat neurons, and as noted in Chapter One, H3.3 has been shown to become the dominant histone H3 variant in rat brain cortical neurons (Pina and Suau, 1987; Scaturro et al., 1995). In further support of this hypothesis, recent studies in *Drosophila* have determined that surviving H3.3 null flies are 'rescued' by upregulation of replication-dependent histones (see Chapter One) (Sakai et al., 2009).

Future experiments by my colleagues in the Allis laboratory and in Dr. Peter

Scambler's laboratory will further explore the functional role of Hira-mediated deposition of H3.3 in ES cell differentiation. In particular, my colleagues Simon Elsaesser and Dr. Ariane Chapgier will use the wild-type and Hira $-/-$ H3.3B-HA ES cells I have generated to examine the specific requirements of Hira for H3.3 deposition in differentiated cells, as my experiments in Hira $-/-$ cells have only examined H3.3 incorporation in ES cells. Although the precise function of Hira in cellular differentiation remains to be established, Hira is clearly required for early embryonic development (Roberts et al., 2002), and the data presented in this thesis are the first to demonstrate that Hira is required for genome-wide genic incorporation of H3.3.

CHD1 in mammalian H3.3 deposition

CHD1 is important for H3.3 deposition in the *Drosophila* male pronucleus, and ChIP-chip studies show that CHD1 localizes to gene TSS in ES cells (Gaspar-Maia et al., 2009; Konev et al., 2007). However, the role of CHD1 in H3.3 incorporation has not yet been explored in mammalian cells. While both CHD1 and Hira are likely to play important roles beyond H3.3 incorporation, it is notable that the phenotype of CHD1 shRNA and Hira $-/-$ ES cells is similar, with global maintenance of the ES cell transcriptome and specific defects upon cellular differentiation (Gaspar-Maia et al., 2009; Meshorer et al., 2006; Roberts et al., 2002). For example, CHD1 shRNA ES cells appear to be unable to differentiate into post-mitotic cardiomyocytes (Gaspar-Maia et al., 2009). Future collaborative studies could use the tagged H3.3B mouse ES cells that I have developed in combination with CHD1 shRNA (Gaspar-Maia et al., 2009) to examine the potential requirements for CHD1 in mammalian H3.3 deposition.

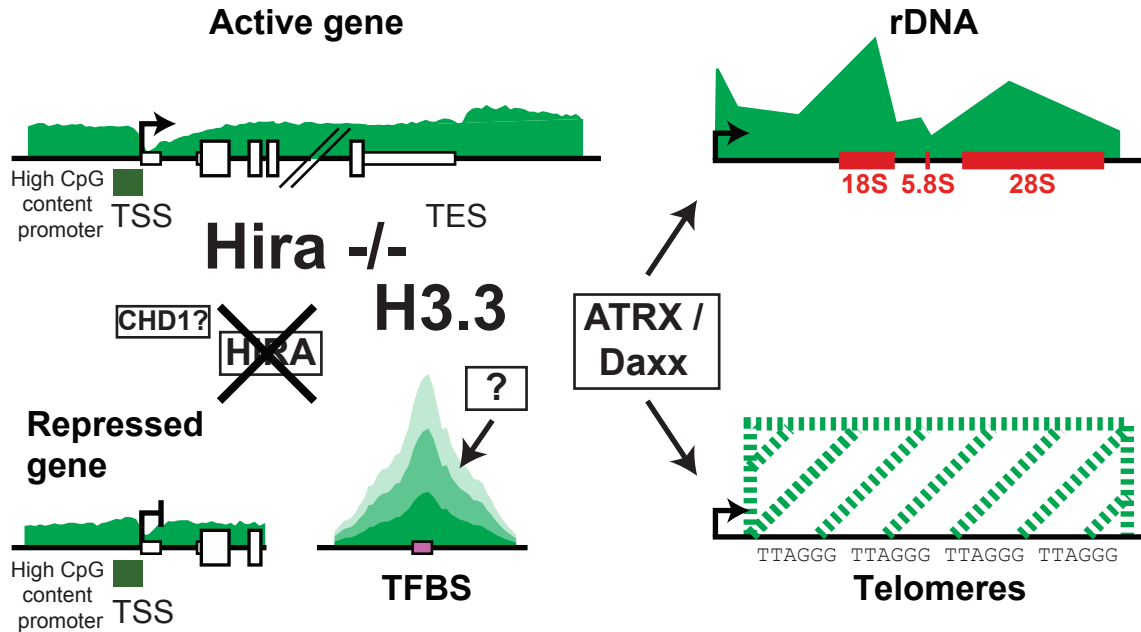


Figure 4.8: Hira is required for deposition of H3.3 at active genes, repressed genes, and specific transcription factor binding sites

In Hira^{-/-} ES cells, H3.3 enrichment at active and repressed genes, as well as at specific TFBS, is lost. Some TFBS show increased H3.3 in the absence of Hira, while other TFBS show decreased H3.3, and at most TFBS, H3.3 deposition is unaffected. Therefore, H3.3 enrichment at many TFBS is maintained by Hira-independent factors (boxed question mark). H3.3 deposition at rDNA and telomeres is mediated by ATRX and possibly Daxx, and is Hira-independent.

Hira-independent H3.3 deposition at regulatory elements

Recent studies found specific peaks of H3.3 deposition at TFBS in *Drosophila* and human cells (Jin et al., 2009; Mito et al., 2007). Deposition of H3.3 at TFBS has been proposed to serve as a mechanism for maintaining regulatory elements in a more accessible chromatin conformation (Henikoff, 2008). Thus, while I anticipated an evolutionarily conserved role for H3.3 deposition at regulatory elements, the extent of H3.3-specific deposition at mammalian ES TFBS was unexpected. Close comparison of our data to a recent dataset of 13 different TFs in mouse ES cells (Chen et al., 2008b) shows H3.3 enriched in ES cells at all known types of TFBS genome-wide, whether in gene bodies, promoters, or intergenic regions. My collaborators and I also found a strong positive correlation between previously

identified loci bound my multiple transcription factors (MTL) and H3.3 deposition, indicating particular enrichment of H3.3 at enhancer elements. In all cases, patterns of H3.3 enrichment at TFBS were abolished upon mutation of H3.3 to H3.2 or H3.1S31.

Our data demonstrate that Hira is involved in H3.3 deposition at some genic and intergenic TFBS. However, we also find that genome-wide H3.3 enrichment at many regulatory elements is Hira-independent and Atrx-independent. Collectively, these data therefore suggest that H3.3 deposition at TFBS may be mediated by multiple and distinct factors, including Hira, with different factors mediating H3.3 deposition at specific regulatory elements (**Figure 4.8**). Notably, genome-wide patterns of H3.3 correspond closely with H3K4me1, and this correlation between H3.3 and H3K4me1 is more apparent at regulatory elements than at genes or other genomic regions. Perhaps H3K4me1, potentially in combination with other enhancer-enriched histone modifications (Wang et al., 2008), may serve to recruit or stabilize as yet unidentified H3.3 deposition machinery at specific regulatory elements.

ATRX and Daxx: new histone variant deposition machinery?

Using ChIP-seq in both undifferentiated ES cells and NPCs, the data presented in Chapter Two demonstrate that H3.3 is specifically enriched in the canonical (TTAGGG)_n repeat that is the hallmark of telomeres in vertebrates (Meyne et al., 1989). Cell imaging studies also show that H3.3 is present in telomeric foci on the largely heterochromatic Y-chromosome in ES cells, and H3.3S31 phosphorylation is also present on the telomeres of ES cell chromosomes. These results agree with a recent report that described the enrichment of H3.3S31 phosphorylation and overexpressed MYC-tagged H3.3 on telomeres in ES cells (Wong et al., 2009).

Surprisingly, the data presented in Chapter Three demonstrate that telomeric

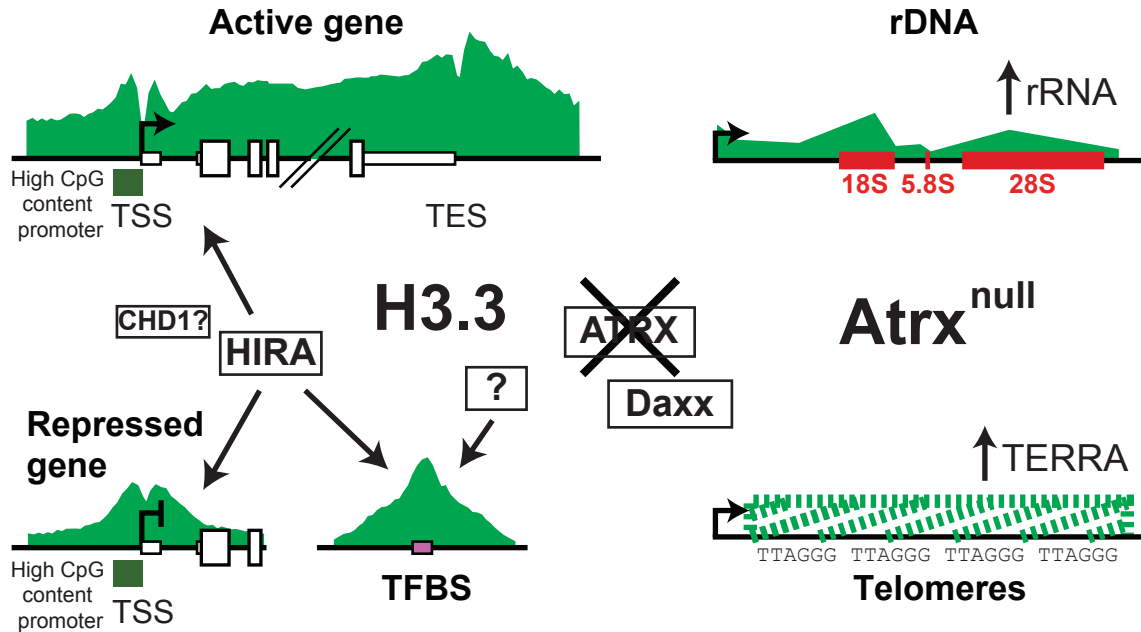


Figure 4.9: Atrx is required for H3.3 deposition at ribosomal DNA and telomeres, and for the repression of ribosomal RNA and telomeric repeat-containing RNA

In *Atrx*^{null} ES cells, H3.3 enrichment at active genes, repressed genes, and TFBS is unaffected. However, H3.3 enrichment at rDNA and telomeres is lost in *Atrx*^{null} ES cells. In addition, rRNA and telomeric repeat-containing RNA (TERRA) transcripts are increased in the absence of Atrx.

H3.3 deposition, rDNA H3.3 deposition, specific TFBS H3.3 deposition, and mitotic H3.3 serine 31 phosphorylation are all Hira-independent. These observations led to a search for factors that might be responsible for Hira-independent deposition of H3.3. In collaboration with Drs. Ileana Cristea and Peter Lewis, I have identified Atrx and Daxx as proteins that associate with H3.3 oligonucleosomes in the presence and absence of Hira. Using ChIP-seq, my colleagues and I have demonstrated that Atrx is required for H3.3 enrichment at telomeres and rDNA.

Our results suggest that Atrx is required for the maintenance of telomere and rDNA chromatin state in ES cells. Recent studies have shown that the *Drosophila* homolog of Atrx, XNP, co-localizes with H3.3 at sites of nucleosome replacement on polytene chromosomes, but is not required for H3.3 localization at these sites (Schneiderman et al., 2009). The data presented in this thesis show that Atrx is required for enrichment of H3.3 at mammalian ES cell telomeres and

the transcribed regions of rDNA, potentially defining a new category of specifically targeted histone variant deposition machinery. Moreover, in the absence of Atrx, ES cells show a similar upregulation of both telomeric repeat-containing RNA (TERRA) and ribosomal RNA (rRNA) (**Figure 4.9**). This result is similar to a recent finding in human patients with ICF (immunodeficiency, centromeric region instability, facial anomalies) syndrome, where mutations in DNMT3B and hypomethylation of subtelomeric DNA are also associated with increased levels of TERRA (Yehezkel et al., 2008).

rDNA and telomeres: specialized, transcribed repetitive regions of heterochromatin

Could ATRX and Daxx serve as specific H3.3 variant deposition machinery for specialized heterochromatin and repetitive regions such as telomeres and rDNA? What might telomeres and rDNA have in common? Transcription of both of these repetitive elements is subject to epigenetic regulation by DNA methyltransferases and histone modifying enzymes (Luke and Lingner, 2009; McStay and Grummt, 2008). Although ribosomal RNA synthesis represents the majority of transcriptional activity in dividing cells, even in highly proliferative cells, a significant portion of rDNA is silenced (McStay and Grummt, 2008). The ratio of active and silent rDNA clusters is regulated by chromatin modifying complexes, and these complexes promote distinct patterns of DNA methylation, histone modifications, and nucleosome positions at active and silent rDNA repeats (McStay and Grummt, 2008). Similarly, subtelomeric regions and transcribed telomeric repeats share characteristics of heterochromatin, including DNA methylation of subtelomeric regions, H3K9me3 and H4K20me3, and histone deacetylase activity (Blasco, 2007; Luke and Lingner, 2009; Marion et al., 2009; Schoeftner and Blasco, 2009).

As originally established by Roeder and colleagues and described in Chapter

Figure 4.10: Hypothetical model for Atrx/Daxx-mediated deposition of H3.3 at telomeres and ribosomal DNA

A. In wild-type ES cells, Atrx and Daxx mediate H3.3-containing nucleosome replacement at telomeres and ribosomal DNA (rDNA). Telomeric chromatin is depicted here, though a similar model would apply for rDNA. The ATRX-DMNT3L-DNMT3A (ADD) domain of ATRX interacts with an unmodified or H3K9me3 histone tail to stabilize ATRX on chromatin. Daxx interacts specifically with H3.3 as well as ATRX, and Daxx presents H3.3-H4 dimers to ATRX for deposition and nucleosome assembly. rDNA and possibly telomeres are transcribed by RNA polymerase I (RNAPI). The growing pink strand with UUAGG represents telomeric repeat-containing RNA (TERRA). Transcription evicts nucleosomes, which are replaced by H3.3-containing nucleosomes deposited by the actions of Atrx and Daxx. It is possible that Atrx/Daxx-mediated deposition of H3.3 at telomeres and rDNA may be coupled to RNAPI transcription, though this hypothesis has not been tested.

B. In Atrxnull ES cells, nucleosomes are depleted at rDNA and telomeres by transcription. In the absence of Atrx-Daxx-mediated nucleosome replacement, nucleosome density at telomeres and rDNA decreases, and aberrant spurious TERRA and rRNA transcription takes place.

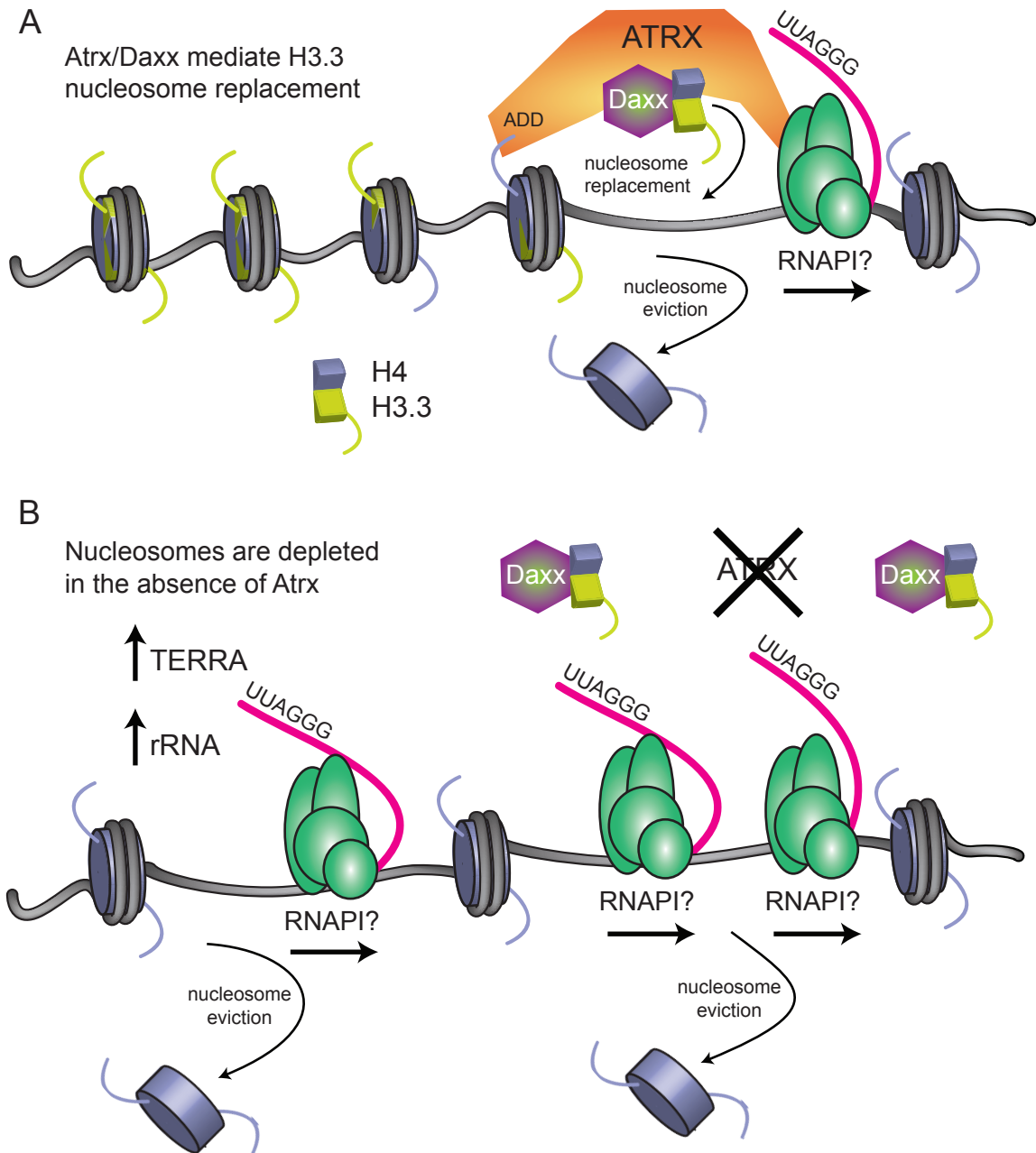


Figure 4.10

One, rDNA is transcribed by RNAPI (Grummt, 2003; McStay and Grummt, 2008; Roeder, 2003; Roeder and Rutter, 1969, 1970). While RNAPII is believed to play a major role in TERRA transcription, various data also suggest that that another RNA polymerase (RNAPI or RNAPIII) could function in mammalian telomeric transcription (Luke and Lingner, 2009). As RNAPI is responsible for transcription of rDNA, it is therefore possible that Atrx-dependent H3.3 deposition at telomeres and rDNA may both be coupled to RNAPI-mediated transcription. Like the RNA polymerases discovered by Roeder, distinct histone deposition complexes may be specialized to perform the same activity (H3.3 deposition) at distinct genomic regions. How might ATRX-mediated deposition of H3.3 serve to repress telomeric and ribosomal RNA transcription?

In the absence of Atrx-mediated deposition of H3.3, transcribed telomeric and rDNA chromatin might acquire an inappropriately low nucleosome density, enabling aberrant upregulation of TERRA and rRNA (**Figure 4.10**). Just as H3K36me3 recruitment of histone deacetylases (HDACs) prevents spurious intragenic transcription from initiating within gene bodies (Carrozza et al., 2005; Joshi and Struhl, 2005; Keogh et al., 2005; Li et al., 2007b; Li et al., 2007c), Atrx-mediated replacement of H3.3 containing nucleosomes may be required to prevent aberrant upregulated transcription of TERRA and rRNA. Indeed, recent experiments show a reduction in total histone H3 occupancy at the human rDNA promoter and transcribed region in alpha thalassemia and X-linked mental retardation (ATR-X) syndrome patient lymphoblastoid cells (Martin Law, Richard Gibbons, Douglas R. Higgs, personal communication).

Daxx as a potential H3.3 binding partner

While this thesis does not explore the role of Daxx, recent data from my laboratory colleagues Dr. Peter Lewis and Simon Elsaesser suggest that Daxx is likely to

play an important role in H3.3 deposition. Like Atrx and Hira, Daxx is required for embryonic development, and loss of Daxx results in embryonic lethality by E9.5 (Michaelson et al., 1999; Salomoni and Khelifi, 2006). Daxx is a ubiquitously expressed ~120 kDa protein with two predicted paired amphipathic helices, an acid-rich domain and a Ser/Pro/Thr (SPT)-rich domain (Hollenbach et al., 2002). Daxx is known to interact with many proteins in mammalian cells, including various transcription factors, and has been shown to function as a transcriptional repressor at specific TFBS (Chang et al., 2005; Croxton et al., 2006; Hollenbach et al., 2002; Huang and Shih, 2009; Mizuta and Kuroda, 2004; Muromoto et al., 2006; Puto and Reed, 2008; Wethkamp and Klemphauer, 2009). As we have seen in Chapter Three, the association between Daxx/ATRX and H3.3 is conserved in both undifferentiated ES cells and human cervical cancer (HeLa) cells.

The localization of Daxx has been shown to be developmentally regulated. Daxx is localized to promyelocytic leukemia (PML) bodies in differentiated human cells, but in undifferentiated human ES cells, Daxx localizes to specific foci that do not overlap with PML bodies (Butler et al., 2009). Intriguingly, Daxx has been specifically localized to the transcriptionally silenced XY body in mid to late stage pachytene spermatocytes (Rogers et al., 2004), as has H3.3 (van der Heijden et al., 2007). Notably, the timing of Daxx co-localization with the XY body actually coincides better with H3.3 (peaking in mid to late pachytene) than the timing of Hira localization with the XY body (early pachytene) (Rogers et al., 2004; van der Heijden et al., 2007). Therefore, I speculate that Daxx may be required to mediate H3.3 deposition in the XY body.

My colleagues in the Allis laboratory are currently engaged in further studies of Daxx and H3.3. Like Atrx^{null} ES cells, Daxx ^{-/-} ES cells are viable (Michaelson et al., 1999). Therefore, Dr. Kyung-Min Noh and I have recently used the zinc finger nuclease method that I described in Chapter Two to generate Daxx ^{-/-} H3.3B / H3.3-

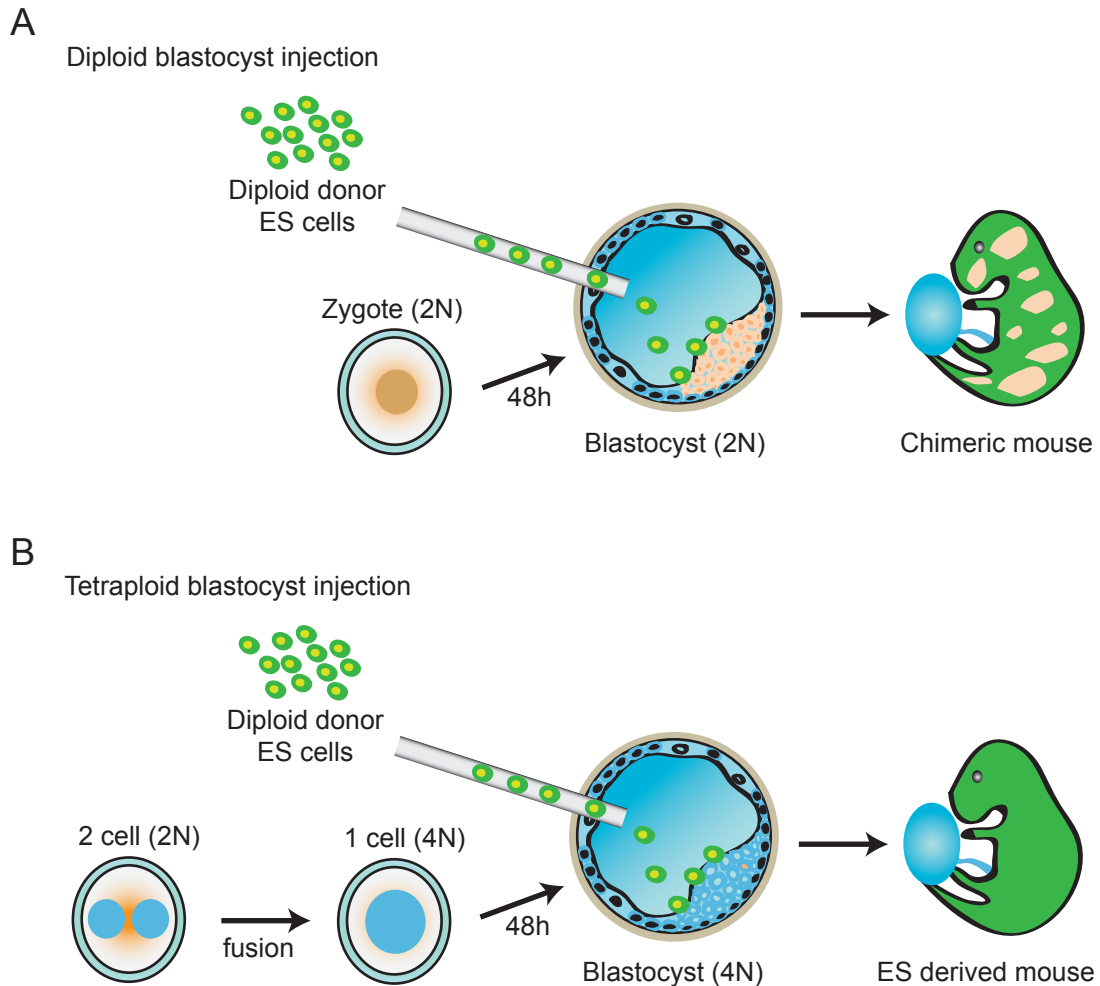
HA-IRES-EYFP ES cells (data not shown). Future experiments will involve ChIP-seq of these targeted Daxx ^{-/-} ES cells to examine the genome-wide localization of H3.3 in the presence and absence of Daxx. Remarkably, recent biochemical experiments suggest that Daxx may serve as a direct H3.3 binding protein (S. Elsaesser and P. Lewis, unpublished data), and additional future studies will explore the specific role of Daxx and Atrx in nucleosome assembly using defined factors *in vitro*.

Potential implications for ATR-X syndrome pathogenesis

More speculatively, the results in this thesis raise the possibility that the pathogenesis of human ATR-X syndrome might involve an early developmental alteration in telomeric chromatin state.

ATRX has been previously localized to heterochromatin, and the ATRX/Daxx complex has been shown to have chromatin remodeling activity (Tang et al., 2004; Xue et al., 2003). However, despite significant advances in the characterization of the ATRX protein and the specific mutations that occur in ATR-X syndrome, the molecular mechanisms underlying the pathogenesis of this syndrome remain a mystery (Gibbons, 2006; Gibbons et al., 2008; Higgs et al., 2005). Intriguingly, patients with ATR-X syndrome show a downregulation in alpha-globin gene expression, but not beta-globin gene expression, and human alpha globin genes are located near the telomere of human chromosome 16 (Higgs et al., 2005) (**Figure 4.11**). While it has been speculated that this chromosomal environment might be important for the pathogenesis of ATR-X syndrome (Higgs et al., 2005), this thesis is the first report to document that Atrx plays a role at telomeres.

One hypothesis is that Atrx-mediated telomeric deposition of H3.3 is required to prevent the spreading of heterochromatin and silencing of susceptible genes adjacent to telomeres in later development (**Figure 4.12**). Better described in lower



Adapted from Meissner et al. 2009.

Figure 4.13: Pluripotent embryonic stem cells give rise to chimeras and are competent for tetraploid complementation

A. In conventional generation of mice from ES cells, diploid donor ES cells (green) are injected into diploid blastocysts. The resulting mice are chimeras of the diploid donor ES cells and the pluripotent cells from the recipient blastocyst inner cell mass (pink).

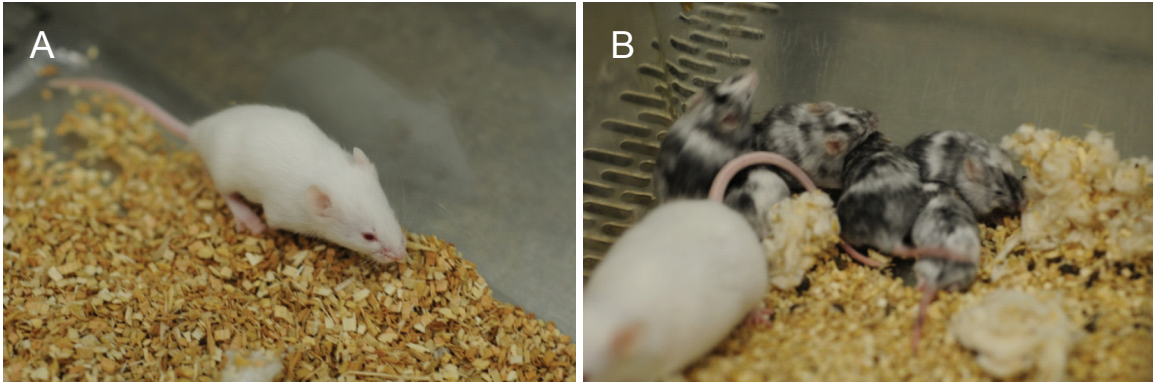
B. Tetraploid complementation provides a rapid method to generate fully ES-derived mice, but this procedure is only effective with healthy, pluripotent cells. Two cell embryos are fused to generate a tetraploid cell (4N), and the resulting tetraploid blastocyst is injected with diploid donor ES cells (green). Tetraploid cells of the blastocysts will only give rise to extra-embryonic tissues such as the placenta (blue), while injected diploid ES cells will give rise to the entire embryo (green). See (Eggan et al., 2001; Meissner et al., 2009).

eukaryotes, this phenomenon of silencing of telomere proximal genes is known as telomere position effect (Ottaviani et al., 2008) (**Figure 4.12**). In support of this hypothesis, in addition to the telomere proximal alpha globin genes downregulated

in ATR-X human patients, the top two genes downregulated in a microarray analysis of *Atrx*^{null} mouse forebrain at E13.5 and P0.5 (*Dhrsxy* and *Csf2ra*) are located near telomeres (Levy et al., 2008). In the absence of *Atrx*, TERRA is upregulated, and TERRA has been associated with maintenance of the telomeric heterochromatin state (Luke and Lingner, 2009). Other histone variants (H2A.Z) in yeast have been shown to antagonize spreading of telomeric heterochromatin and the silencing of telomere adjacent euchromatin (Meneghini et al., 2003). To test the hypothesis that spreading of telomeric heterochromatin may be important for ATR-X pathogenesis, future collaborative experiments will examine the chromatin state of telomeres and proximal telomeric regions in wild-type and *Atrx*^{null} ES cells, as well as patients with ATR-X syndrome.

Localization and function of H3.3B in mammalian development – recent progress towards a H3.3B-HA-IRES-EYFP mouse line

Given the nature of the genetic modifications that I have made to ES cells in order to track the localization of histone H3.3, it is important to determine whether or not these cells remain pluripotent. Moreover, for future experiments, it is desirable to examine the genomic localization of H3.3 in differentiated mammalian tissues, as well as to establish the endogenous pattern of H3.3B expression. There are a variety of tests to determine the developmental potential of ES cells, including embryoid body formation, teratoma formation in immunocompromised mice, and the generation of chimeras following injection of ES cells into diploid blastocyst embryos (Meissner et al., 2009). The most rigorous test of pluripotency involves the injection of diploid ES cells into tetraploid blastocysts (Eggan et al., 2001; Meissner et al., 2009) (**Figure 4.13**). As tetraploid cells can maintain extraembryonic tissues such as the placenta, but are largely incapable of giving rise to the embryo itself, the injected diploid ES cells will generate the entire embryo (Eggan et al., 2001;



Blastocyst injections and in vivo experiments performed by Duancheng Wen and Shahin Rafii, and data used with permission.

Figure 4.14: Generation of H3.3B-HA-IRES-EYFP mice

H3.3B / H3.3-HA-IRES-EYFP ES cells injected into diploid B6 blastocysts contribute to mice with high chimerism. Most ES cells described in this study are B6 but maintained on a tyrosinase $-/-$ background, and their coat color is therefore albino.

A. H3.3B / H3.3-HA-IRES-EYFP ES cells gave rise to a healthy mouse that is $\sim 100\%$ donor ES-derived as judged by white coat color and genotyping. In this case, the chimera is surprisingly from diploid blastocyst injection, but healthy H3.3B / H3.3-HA-IRES-EYFP mice were also obtained from tetraploid complementation.

B. Other mice from the same litter show significant chimerism.

Meissner et al., 2009) (**Figure 4.13**).

In collaboration with Dr. Duancheng Wen and Dr. Shahin Rafii, heterozygous H3.3B / H3.3-EYFP and H3.3B / H3.3-HA-IRES-EYFP ES cells were injected into both diploid and tetraploid blastocysts. Despite many attempts with several clones, H3.3B-EYFP ES cells generated pups with generally poor ($<45\%$) chimerism (data not shown). In addition, none of the of H3.3B / H3.3-EYFP chimeras was found to be capable of H3.3-EYFP germline transmission.

It is possible that the poor chimerism and lack of germline transmission of H3.3B-EYFP ES cells could be due to a hypomorphic effect of the EYFP tag. The EYFP tag (238aa) is large relative to the H3.3 histone (135aa). Although the data presented in Chapter Two indicate that genome-wide patterns of H3.3-EYFP are extremely similar to H3.3-HA, the bulky fluorescent tag could have resulted in a mild hypomorphic effect on nucleosome assembly or nucleosome interactions that

is manifested during development. H2B-GFP fusions have been used in various lines of fertile mice (Hadjantonakis; Hadjantonakis and Papaioannou, 2004; Tumber et al., 2004). However, there is precedent for hypomorphic effects of a H3 family GFP fusion *in vivo*, following knock-in of a GFP tag fused to the H3 centromeric variant CENP-A (Kalitsis et al., 2003). Although *Cenpa*-GFP heterozygous mice were completely healthy and fertile, *Cenpa*-GFP homozygotes died at E10.5 of embryonic development, with significant chromosome missegregation, aneuploidy and apoptosis (Kalitsis et al., 2003).

Despite the lack of success with H3.3B-EYFP ES cells, recent injections of H3.3B / H3.3-HA-IRES-EYFP ES cells into diploid blastocysts yielded viable pups with extremely high chimerism from several different clones, with almost 100% ES contribution in one particular chimera (**Figure 4.14**). In addition, several clones of H3.3B / H3.3-HA-IRES-EYFP ES cells also gave rise to healthy pups following complementation of tetraploid blastocysts, with 100% contribution from the tagged ES cells (D. Wen, data not shown). These results are strong evidence for pluripotency of H3.3B / H3.3-HA-IRES-EYFP ES cells. Beyond tetraploid complementation, the only further test of pluripotency involves germline transmission. Once old enough for breeding, these mice will be bred to establish a permanent H3.3B / H3.3-HA-IRES-EYFP mouse line.

The establishment of H3.3B / H3.3-HA-IRES-EYFP mice opens up a wide array of experimental possibilities. The localization of H3.3 during and following X chromosome inactivation (Chow and Heard, 2009) can now be assessed using female H3.3B-HA mouse embryos, female H3.3B-HA mouse ES cells, and somatic cells from adult female H3.3B-HA mice. H3.3B / H3.3-HA-IRES-EYFP mice will also enable analysis of the developmental expression of H3.3B *in vivo*, a subject that has previously only been addressed through northern blots and *in situ* hybridization of select tissues (Bramlage et al., 1997). Although I have profiled the genome-wide

localization of H3.3 in ES cells and NPCs, the establishment of H3.3B-HA mice will also allow us to profile the genome-wide localization of H3.3 in other embryonic and adult mammalian cell types, including derived cell lines as well as organs such as liver and brain (Meissner et al., 2008; Visel et al., 2009; Wederell et al., 2008). Multiple H3.3B-HA cell lines can be derived for ChIP-seq and immunofluorescence studies, including mouse embryonic fibroblasts, and spermatogonial stem cells (SSCs) (Seandel et al., 2007).

Provided that H3.3B is expressed in primordial germ cells (PGCs), the role of H3.3 in germ cell specification during embryonic development can also be addressed, as previous studies of PGCs were hampered by the lack of a H3.3 antibody (Hajkova et al., 2008). Moreover, the localization of endogenously expressed H3.3B in spermatogenesis can also be determined. Could H3.3 localization to the repressive XY body be dependent on Daxx, and not Hira (Rogers et al., 2004; van der Heijden et al., 2007)? In addition, although most sperm chromatin is packaged in protamines, specific histone H3 modifications have been shown to mark particular developmental genes in sperm, and both H3.3 and H3.1 have been found at lower levels in human sperm chromatin (Gatewood et al., 1990; Hammoud et al., 2009). Are these sperm H3 histone modifications carried on H3.3? ChIP-seq from sperm isolated from H3.3B / H3.3-HA-IRES-EYFP mice could be used to address these questions. Finally, the localization of endogenously expressed H3.3B in paternal pronuclear decondensation can also be addressed in mammals for the first time (Bonney et al., 2007; Konev et al., 2007; Loppin et al., 2005; Torres-Padilla et al., 2006).

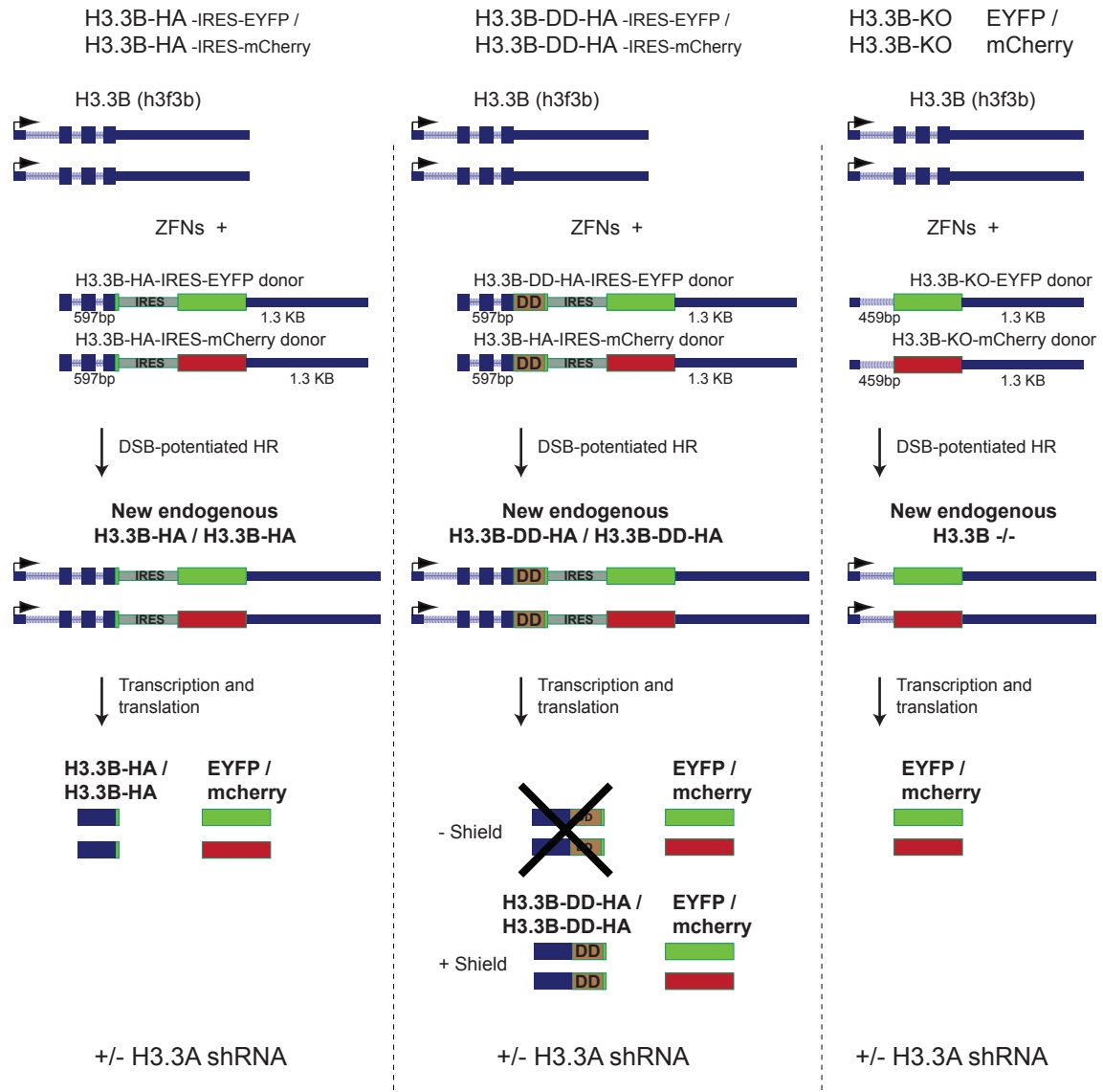


Figure 4.15: Strategy for homozygous targeting of H3.3B

Left. Zinc finger nucleases (ZFNs) in combination with H3.3B-HA-IRES-EYFP and H3.3B-HA-IRES-mCherry donors enable homozygous targeting of H3.3B, via double-strand break (DSB) potentiated homologous recombination. Exon coding sequences are thick rectangles, while non-coding untranslated sequences are thin rectangles. The resulting EYFP and mCherry double-positive cells have both H3.3B alleles fused to a C-terminal HA epitope tag, with an IRES and separately translated EYFP or mCherry reporter.

Center. A similar strategy can be used to enable homozygous addition of a destabilizing domain (DD) sequence (brown) as well as an HA epitope tag fused to the C-terminus of H3.3B. In the absence of shield ligand, DD-tagged H3.3B protein is degraded, but in the presence of shield ligand, H3.3B-DD-HA protein expression is maintained. For more details regarding destabilizing domains, see: (Banaszynski et al., 2006).

Right. A similar strategy can be used to generate a homozygous knockout of H3.3B, with full replacement of coding exons by EYFP and mCherry. All of these cells can be used with and without the addition of lentiviral shRNA vectors for H3.3A knockdown.

Ongoing experiments to probe the function of H3.3 through H3.3B knockout, regulated protein expression, and H3.3A shRNA

Although H3.3B-HA-IRES-EYFP mice unlock a wide range of experimental possibilities, fundamentally these mice will only enable the tracking of epitope-tagged H3.3B through development and differentiation. To determine the functional significance of H3.3, it is necessary to generate cellular models that reduce, eliminate, or regulate H3.3 protein expression. Recent collaborative experiments with Dr. Laura Banaszynski in our laboratory provide significant progress towards this goal.

Mammalian cells contain four alleles that code for H3.3, including two alleles of H3.3B and two alleles of H3.3A (Bramlage et al., 1997; Marzluff et al., 2002). To target both copies of H3.3B, Dr. Banaszynski and I devised a strategy to modify my

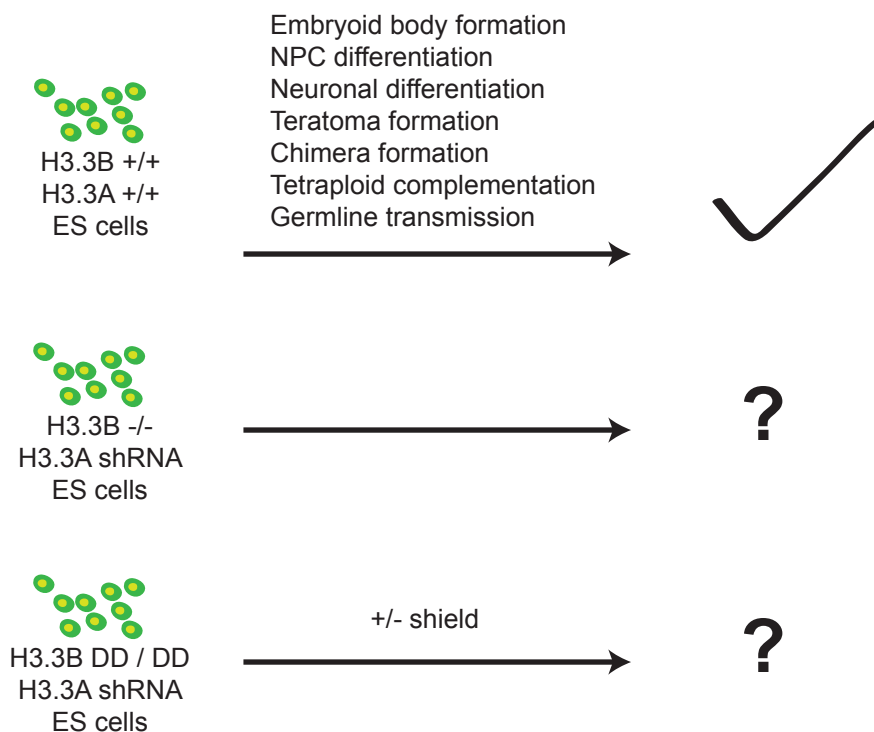


Figure 4.16: Experiments to assess the function of H3.3 in mammalian development and cellular differentiation

Wild-type H3.3B $+/+$ H3.3A $+/+$ ES cells are pluripotent, and are competent in all of the assays listed. The phenotypes of H3.3 'null' (middle) or H3.3B-DD 'rescue' ES cells in these assays are unknown, and these experiments are currently underway.

donor construct to include the red fluorescent protein mCherry in place of EYFP. Following transfection of cells with zinc finger nucleases and both EYFP and mCherry H3.3B donor constructs, cells that become stably EYFP and mCherry double positive are likely to be doubly targeted at both H3.3B alleles (**Figure 4.15**). This strategy can be used for homozygous tagging of H3.3B with epitope tags (**Figure 4.15, left**). In addition, Dr. Banaszynski modified my donor constructs to fuse a destabilizing domain to the C-terminus of H3.3B (Banaszynski et al., 2006). Such destabilizing domains (DD) were developed by Dr. Banaszynski during her graduate work, and they enable rapid, reversible, and tunable regulation of protein levels through the use of a small molecule (Banaszynski et al., 2006). Targeting of both H3.3B alleles with H3.3B-DD-HA-EYFP and H3.3B-DD-HA-mCherry donor constructs will allow total control over H3.3B protein expression (**Figure 4.15, center**). To generate a full H3.3B knockout, I also modified my donor constructs to fully replace the coding exons of H3.3B with the open reading frame of EYFP and mCherry (**Figure 4.15, right**). To knock-down expression of H3.3A, we have validated several lentiviral shRNA vectors that enable greater than 90% H3.3A mRNA knock-down (data not shown).

H3.3 'null' (H3.3B $-/-$ + H3.3A shRNA, or H3.3B-DD-HA + H3.3A shRNA) ES cells have recently been established, and they are being used for a variety of experiments. Preliminary experiments suggest that H3.3 null ES cells are viable, but display a mild growth defect in culture (data not shown). Most immediately, Dr. Banaszynski is examining the expression of rRNA and TERRA in the absence of H3.3. If the function of ATRX-mediated H3.3 nucleosome replacement at rDNA and telomeres is to repress transcription of these heterochromatic repeats, then loss of H3.3 may also lead to aberrant upregulation of rRNA and TERRA (**Figure 4.10**). In addition, my colleagues will use whole transcriptome profiling (RNA-seq) (Wang et al., 2009a) to examine whether loss of H3.3 leads to any global alterations in

transcriptional regulation.

To assess the function of H3.3 in mammalian development and cellular differentiation, wild-type and H3.3 null ES cells will be used in several experiments (**Figure 4.16**). *In vitro* cell differentiation assays will include embryoid body formation, retinoic acid differentiation, NPC formation, and differentiation into post-mitotic neurons (Conti et al., 2005). In collaboration with Dr. Shahin Rafii's laboratory, H3.3 null ES cells will also be injected into immunocompromised mice in teratoma formation experiments, as well into diploid and tetraploid blastocysts to assess the requirements for H3.3 *in vivo* (**Figure 4.16**). In addition, Dr. Duancheng Wen has already generated chimeras from heterozygous H3.3B / H3.3B-KO-EYFP ES cells into diploid blastocysts (data not shown), and these chimeras will be bred with wild-type mice in an attempt to establish a heterozygous H3.3B +/-EYFP line.

Perspective, predictions, future questions, and challenges

This thesis provides the first genome-wide profiles of histone H3 variants in mammalian pluripotent cells. My collaborators and I have shown that Hira is required for H3.3 deposition at genes, and we identify Atrx as required for Hira-independent H3.3 deposition at telomeres and rDNA. Our data also suggest the existence of Hira and Atrx-independent specific machinery responsible for H3.3 deposition at specific regulatory elements. Overall, the work presented in this thesis raises the prospect that distinct, region-specific chaperone and remodeling complexes mediate the targeting of a single histone variant (H3.3) to particular genomic regions.

These findings raise multiple questions. What is the function of H3.3 at these different genomic sites? Does H3.3 play a similar role at genes, TFBS, rDNA and telomeres? I predict that H3.3 has largely evolved to function in nucleosome replacement and 'chromatin repair' in a variety of genomic, cellular,

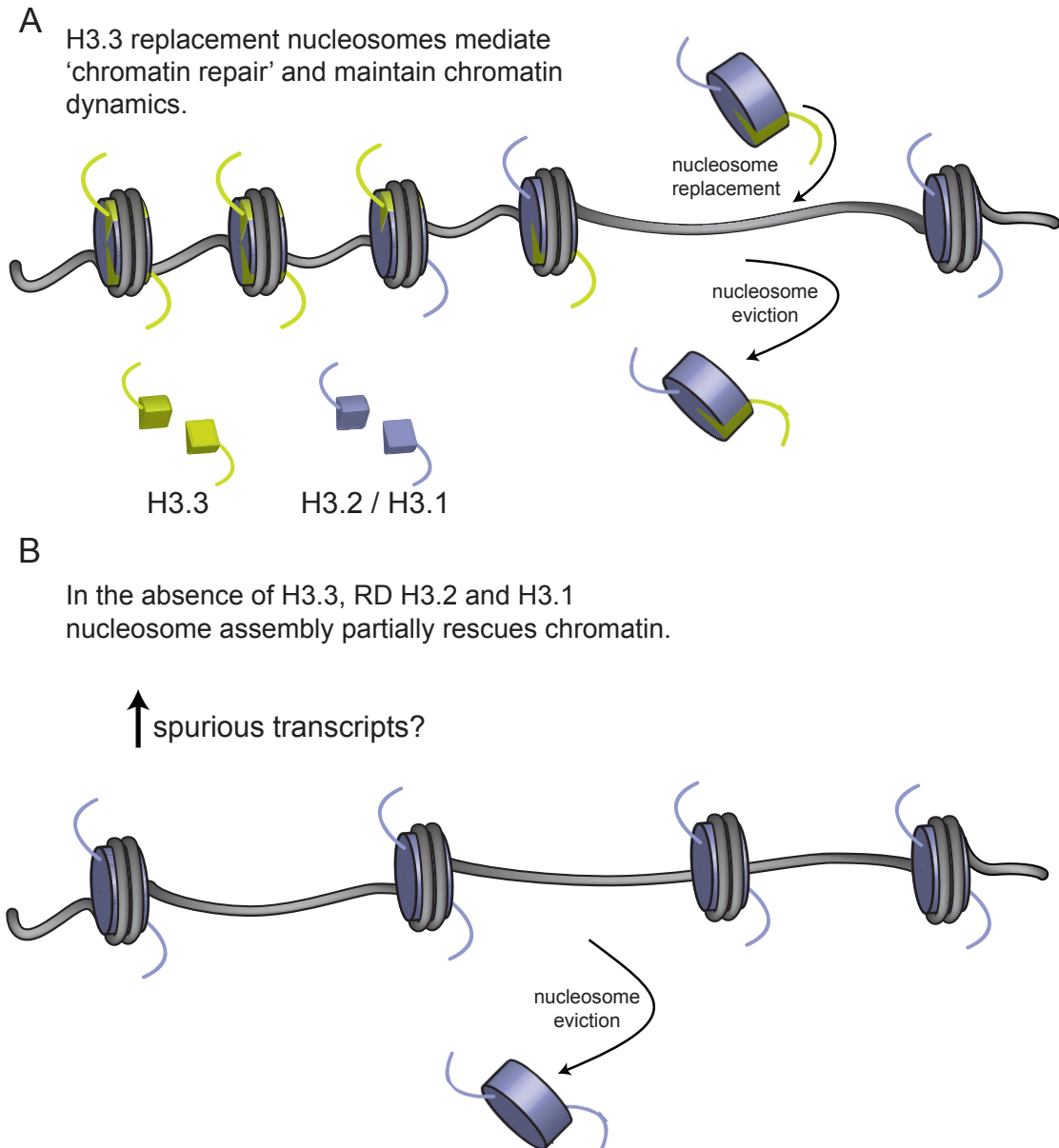


Figure 4.17: Speculative model: the primary non-reproductive functions of H3.3 are nucleosome replacement and chromatin repair at regions of chromatin disruption

A. In wild-type cells, nucleosome replacement with H3.3-containing nucleosomes enables 'chromatin repair' at regions of chromatin disruption, while the intrinsic instability of H3.3-containing nucleosomes also promotes dynamic chromatin. At regions of chromatin disruption such as genes and regulatory elements, H3.3 containing nucleosomes strike an ideal balance between overly exposed naked DNA and overly stable H3.2 or H3.1 nucleosomes.

B. In the absence of H3.3, nucleosome replacement at regions of chromatin disruption is impaired. Nucleosome replacement and chromatin 'rescue' can be partially mediated by H3.2 or H3.1 containing nucleosomes, most likely through replication-dependent (RD) nucleosome assembly (Sakai et al., 2009). However, nucleosome replacement in mammalian cells may be inadequate in the absence of H3.3. Loss of H3.3 could potentially result in progressive defects such as spurious transcripts, particularly in post-mitotic cells that lack the capacity for RD nucleosome assembly.

and developmental contexts, and that this nucleosome replacement function is critically important in the large-scale chromatin remodeling that takes place during specific developmental transitions in multicellular organisms (Orsi et al., 2009). In addition, recent data shows that H3.3-containing nucleosomes are more sensitive to salt-sensitive disruption than H3.1-containing nucleosomes, suggesting intrinsic differences in stability that could reduce the energy required for nucleosome mobilization (Jin and Felsenfeld, 2007). Therefore, it is likely that H3.3 nucleosomes do not only serve a replacement function, but also contribute to chromatin dynamics (**Figure 4.17**).

What is the function of H3.3 in mammalian development? Recent studies have shown that H3.3 plays a role in *Drosophila* transcription regulation, and is essential for fly fertility (Hodl and Basler, 2009; Ng and Gurdon, 2008; Sakai et al., 2009). As lower levels of H3.3A already lead to partial neonatal lethality, fertility defects, and neuromuscular deficits in mice (Couldrey et al., 1999), it is likely that H3.3 may play an even more important role in mammals. Do cellular requirements for H3.3 differ in dividing versus post-mitotic cells, where replication-independent deposition might play a larger role?

Recent studies from yeast may be informative regarding the possible phenotypes of H3.3 deletion. The yeast homologs of HIRA also play a role in replication-independent histone deposition, and localize to transcribed regions in *S. cerevisiae* (Green et al., 2005; Nourani et al., 2006). Intriguingly, the *S. pombe* HIRA homologs were recently shown to be required for genome-wide promoter silencing and for the repression of cryptic antisense transcripts (Anderson et al., 2009). The HIRA complex has therefore been proposed to restore chromatin structure in the wake of RNAPII (Anderson et al., 2009). Intriguingly, H2A.Z has also recently been observed to suppress spurious antisense transcripts in *S. pombe* (Zofall et al., 2009). Therefore, I predict that H3.3 containing nucleosomes may also serve to

replace nucleosomes disrupted by a variety of processes, including transcription and transcription factor binding, and that this 'chromatin repair' function of H3.3 may be critical for the repression of spurious transcripts (**Figure 4.17**). It is likely that such a 'chromatin repair' function of H3.3 would be particularly important for post-mitotic cells that are incapable of RD histone deposition during S-phase.

The surprising finding that distinct factors are required to deposit H3.3 at specific genomic regions also raises many questions. How are Hira and Atrx targeted to specific genomic regions? Neither Hira nor Atrx have been shown to interact directly with transcription machinery, and significant work remains regarding the recruitment of these complexes to distinct regions. It is likely that like many chromatin-mediated processes, the targeted recruitment of Hira and Atrx may depend significantly on an interaction with site-specific DNA binding factors (Farnham, 2009). As Daxx interacts with multiple transcription factors (Chang et al., 2005; Croxton et al., 2006; Hollenbach et al., 2002; Huang and Shih, 2009; Mizuta and Kuroda, 2004; Muromoto et al., 2006; Puto and Reed, 2008; Salomoni and Khelifi, 2006; Wethkamp and Klempnauer, 2009), it is possible that Daxx-mediated interaction with both H3.3 and cell-type specific TFs may be important for targeting of H3.3 to specific regions.

Finally, what is the mechanism by which H3.3 is deposited? Despite the biochemical requirement of Hira for replication-independent nucleosome assembly, we have very little insight into how Hira, much less Atrx and Daxx, mediate H3.3 deposition into chromatin. Is Atrx-mediated deposition of H3.3 also replication-independent, like Hira, or does it occur during replication? Why are there site-specific mechanisms that deposit H3.3 at transcribed sequences, TFBS, rDNA and telomeres? Many of these questions regarding the mechanism of H3.3 deposition are the subject of active work by my laboratory colleagues, and I am confident that we can look forward to significant advances in the near future.

Overall, the data presented in this thesis make clear that the existing paradigm of Hira-dependent deposition of H3.3 is too simplistic for mammalian cells. While many questions remain, these findings lay the groundwork for a more nuanced understanding of H3 variants in mammalian cells, and provide a foundation for future studies of H3.3 in mammalian development.

CHAPTER 5

MATERIALS AND METHODS

Mouse embryonic stem cells

The parental “C6” embryonic stem (ES) cell line was derived from tyrosinase $-/-$ (white, albino) homozygous mouse embryos of B6(Cg)-Tyrc-2J/J C57BL/6J background. <http://jaxmice.jax.org/strain/000058.html> The parental “965” ES line is an F1 hybrid 129SVJae x *M. m. castaneus* (Eggan et al., 2001; Mikkelsen et al., 2007), and was a gift from Dr. Rudolf Jaenisch. Parental W9.5 and Hira $-/-$ (cl34) ES cells were a gift from Dr. Peter Scambler, and have been described previously (Meshorer et al., 2006). Parental *Atrx*^{flox} and *Atrx*^{null} ES cells were a gift from Drs. David Garrick, Richard Gibbons, and Douglas Higgs, and have been described previously (Garrick et al., 2006).

ES cell culture and neural precursor cell differentiation

Undifferentiated mouse ES cells were cultured under standard conditions in the presence of leukemia inhibitory factor (LIF) (Meissner et al., 2009). For early passages, undifferentiated ES cells were maintained on a feeder layer of mitomycin C treated or irradiated mouse embryonic fibroblasts (MEFs) in standard ES cell media, KO-DMEM (Invitrogen 10829-018), 2mM L-glutamine (Sigma G7513) and pen-strep, 15% ES grade fetal bovine serum (Gibco 10439-024), 10⁻⁴ mM 2-mercaptoethanol, and leukemia inhibitory factor (LIF) (Meissner et al., 2009). To remove feeders for ES cell whole cell extracts, chromosome spreads, chromatin immunoprecipitation, and NPC differentiation, ES cells were passaged at least two passages off of feeders on to gelatin-coated plates. Neural precursor cell (NPC) differentiation was performed as described (Conti et al., 2005). NPCs

were cultured and expanded on gelatin-coated plates in NeuroCult® NSC Basal Medium (Mouse, Stem Cell Technologies, cat#05700) supplemented with 2mM final L-glutamine and pen-strep (Omega Scientific PG-30), modified N2 supplement freshly prepared in house as described (Conti et al., 2005), and 10ng/ml of both mouse EGF (Peprotech 315-09) and human FGF-2 (Peprotech 100-18B).

Zinc finger nuclease design, targeting, and verification

Zinc finger nucleases (ZFNs) directed against the mouse H3.3B (h3f3b) gene were designed using an archive of validated two-finger modules (Doyon et al., 2008; Urnov et al., 2005). Zinc finger nucleases (ZFNs) were designed against a region immediately upstream of the H3.3 stop codon (see **Figure 2.2**), since gene conversion tracts from DSBs induced by I-SceI (Elliott et al., 1998) and by ZFNs in somatic mammalian cells are rather narrow. My goal was to both tag the native H3.3B open-reading frame (ORF) at the C-terminus (Figure 1B) as well as co-convert the H3.3B coding region to a H3.2 or H3.1 allelic form (Figure S1D-E). ZFNs were designed using an archive of validated two-finger modules (Doyon et al., 2008; Shukla et al., 2009; Urnov et al., 2005), assembled using a PCR-based method, cloned into expression vectors and validated as described (Urnov et al., 2005). The ZFNs used carried obligate heterodimer forms of the FokI endonuclease (Miller et al., 2007). ZFN target sequences and the recognition helices are provided below.

ZFN target site and designed zinc finger helices sequences

Locus	ZFN Binding Sequence (underlined)	ZFN	Finger 1	Finger 2	Finger 3	Finger 4
h3f3b	GACATCCAGTTGGCTCGCCGGATACGGGG	#25000	RSDHLE	RNDTRKT	QSSNLAR	RSDDRKT
	CTGTAGGTCAAACGAGCGGCCTATGCCCC	#25001	DRSALSR	TSANLSR	RSDVLSE	QRNHRTT

H3.3B EYFP and HA-IRES-EYFP donor plasmid construction

Briefly, a PCR fragment of genomic DNA from mouse H3.3B was cloned out of a genomic bacterial artificial chromosome (BAC) from C57BL/6J mouse chromosome 11 using Phusion polymerase (NEB F-530L) and into a pCR2.1 vector (pCR2.1-H3.3B) using TA-TOPO cloning (Invitrogen K4500-02). To generate the H3.3B-EYFP donor construct (pCR2.1-H3.3B-EYFP), a 6 amino acid (SRPVAT) linker (Ahmad and Henikoff, 2002) followed by the open-reading frame of EYFP (Clontech) was inserted in-frame into the last coding exon of H3.3B, using ligation-independent cloning (LIC) (Geiser et al., 2001). The H3.3B-EYFP donor included no H3.3B promoter sequence, containing approximately 0.6kb of 5' homologous genomic sequence starting at the second H3.3B codon, including introns, until the last H3.3B coding amino acid, followed by the linker and EYFP, a stop codon, and approximately 1.3kb homologous to the H3.3B 3'UTR (see **Figure 2.4**). To generate the H3.3B-HA-IRES-EYFP construct (pCR2.1-H3.3B-HA-IRES-EYFP), a C-terminal HA-IRES sequence was cloned out of the vector pMSCV-F-del Casp9. IRES.GFP (Straathof et al., 2005), Addgene Plasmid 15567, and inserted in-frame using LIC into the pCR2.1-H3.3B-EYFP vector. To generate H3.2-EYFP, H3.1-EYFP, H3.2-HA-IRES-EYFP, and H3.1-HA-IRES-EYFP donors, point mutations were made in pCR2.1-H3.3B-EYFP and pCR2.1-H3.3B-HA-IRES-EYFP using the Quick-Change Mutagenesis II kit (Stratagene).

ZFN targeting and verification

To deliver ZFNs and donor constructs, mouse ES cells were transfected by Amaxa nucleofection. In brief, immediately prior to transfection, ES cells were feeder depleted by harvesting the ES cells, plating on a feeder-free dish for 30min, and then collecting the ES-enriched non-adherent cells for transfection. 2.5×10^6

cells ES cells were resuspended in 90µl solution, mixed with two non-linearized plasmids (1 µg of ZFN plasmid with both ZFNs separated by a 2A peptide sequence + 10µg of donor plasmid) in 10µl nucleofection solution, and transfected using program A-013 as described in the Amaxa manufacturer's protocol for mouse ES cells. Following transfection, we used sterile plastic pipettes to transfer the cells to warm ES media in tissue culture dishes that were already prepared with feeders. After transfer, ES cells were cultured in standard conditions on treated feeders for 3-5 days prior to fluorescent activated cell sorting (FACS) (**Figure 2.3, 2.5**) or fluorescent colony picking. Following colony picking, clonal isolation and expansion, we prepared genomic DNA using the Qiagen DNeasy Blood & Tissue Kit (Qiagen 69504). Individual clones were screened by PCR using a forward H3.3B primer (TTGGTGGAGTATCTGCCCGTTCTG) outside of the donor construct and reverse primer (GTTCTTCTGCTTGTCGGCCATGATA) within EYFP (see **Figure 2.6**). To verify that the wild-type H3.3B allele remained intact, we also amplified both H3.3B alleles by using the same forward primer above, and a reverse primer (TGAAGCCAACCTGCTGCTCTACAGT) outside the donor construct, performed gel electrophoresis, and isolated the smaller band corresponding to the wild-type H3.3B allele (data not shown). PCR products from both wild-type and modified H3.3B alleles were sequenced using standard methods (**Figure 2.6**). To confirm heterozygosity, we used the primers described above to sequence both H3.3B alleles from all ES lines used in this study. For further confirmation of heterozygosity, we performed Southern blotting on a subset of the ES lines. To perform Southern blotting, we digested genomic DNA from wild-type and targeted ES cells with BsrBI, and used a labeled 638bp Avall fragment of the H3.3B donor as probe to visualize wild-type H3.3B and integrated H3.3B donors. All targeted C6 ES lines used for manuscript were found to be pure heterozygotes (**Figure 2.6**). The hybrid H3.3-HA ES line and the Hira ^{-/-} H3.3-EYFP ES lines were heterozygous, with one

H3.3B allele targeted, but also had additional donor integrations by Southern blot. Western blot analysis indicates that the hybrid ES H3.3-HA is not overexpressed relative to the pure heterozygote hybrid H3.1S31-HA ES line (data not shown), and global patterns of H3.3 were extremely similar to patterns in the pure heterozygous C6 H3.3-HA and H3.3-EYFP ES lines. H3.3-HA and H3.2-HA ES cells were also fully karyotyped, and found to have normal male mouse karyotypes (data not shown).

RT-PCR

Total RNA from ES cells and NPCs was isolated using TRIzol Reagent (Invitrogen) or RNeasy (Qiagen). 1µg of RNA from each sample was used to generate cDNA using the Superscript First-Strand Synthesis kit (Invitrogen). Real-time PCR was performed with SYBR Green PCR master mix (Applied Biosystems) according to the manufacturer's instructions. Pooled cDNA was used to generate a standard curve from which the relative amount of cDNA amplified in each sample was determined as indicated. Primer sequences for RT-PCR are as follows. *Gapdh*_F: GGTGTCTCCTGCGACTTCAACAGC, *Gapdh*_R: CGAGTTGGGATAGGGCCTCTCTTGC (Chew et al., 2005), *Pou5f1*_F: TTGGGCTAGAGAAGGATGTGGT, *Pou5f1*_R: GGAAAAGGGACTGAGTAGAGTGTGG (Chew et al., 2005), *Nanog*_F: CCTCTTCAAGGCAGCCCTGATTCT, *Nanog*_R: AGAGTTCTTGCACTCTGCTGGAGGCT, *Olig1*_F: GCTCGCCCAGGTGTTTTGT, *Olig1*_R: GCATGGAACGTGGTTGGAAT (Boyer et al., 2006), *h3f3a*_F: ACAAAGCCGCTCGCAAGAG, *h3f3a*_R: ATTTCTCGCACCAGACGCTG, *h3f3b*_F: TGGCTCTGAGAGAGATCCGTCGTT, *h3f3b*_R: GGATGTCTTTGGGCATGATGGTGAC,

Flow cytometry

FACS and analysis was performed on a BD FACSAriaII-2. In brief, ES cells were trypsinized and gently separated to single cells for FACS, and resuspended in PBS + 5% ES quality FBS (Gibco 10439-024). FACS collection tubes were pre-coated with ES quality FBS for at least 30 minutes prior to the sort. To prevent clumping during the sort, cells were gently filtered into a single cell suspension through a cell-strainer mesh (Polystyrene Round-Bottom Tube with Cell-Strainer Cap, Falcon #352235). Following sort, cells were immediately plated on feeders with ES media.

Isolation of protein complexes and mass spectrometric analysis

Immunoaffinity purifications of GFP-tagged H3.3, H3.2, H3.1, H3.3 HIRA^{-/-}, and H3.2 HIRA^{-/-} were performed as previously described (Cristea et al., 2005), and MALDI MS and MS/MS analyses were performed as described (Luo et al., 2009). Briefly, for each isolation, ~2.5g cryogenically ground cells were extracted in 9 ml of 20 mM K-HEPES, pH 7.4, 110 mM K-acetate, 0.1% Tween 20, 0.5% Triton, 300 mM NaCl, and 1/100 (v/v) protease inhibitor cocktail (Sigma-Aldrich). Resulting cell homogenates were incubated with 12 mg M-270 epoxy magnetic beads (Dyna, Oslo, Norway), conjugated with in-house developed rabbit polyclonal anti-GFP antibodies (5µg antibodies/mg beads) for one hour at 4°C. Isolated proteins were eluted in 700 µl 0.5N NH₄OH, 0.5mM EDTA for 20 minutes at room temperature, dried by vacuum centrifugation, suspended in protein electrophoresis sample buffer, resolved by 1-D SDS-PAGE (4–12% NuPAGE Novex Bis-Tris gel, Invitrogen), and stained with Coomassie Blue (GelCode Blue, Pierce). The entire gel lanes were cut into sections (~30 sections per gel lane), and proteins were digested with 12.5ng/µl sequencing grade modified trypsin (Promega, WI, USA). Resulting peptides were extracted on reverse phase resin (Poros 20 R2, PerSeptive Biosystems),

eluted in 2 mg/ml α -cyano-4-hydroxycinnamic acid, 70% (v/v) acetonitrile and 0.1% (v/v) trifluoroacetic acid, and deposited onto an in-house made magnetic MALDI target, as described (Blethrow et al., 2007). MALDI MS and MS/MS analyses were performed using prOTOF (Perkin Elmer), vMALDI LTQ XL (Thermo Electron, Bremen, Germany), and MALDI LTQ Orbitrap XL (Thermo) mass spectrometers, as described (Blethrow et al., 2007; Luo et al., 2009). Protein candidates were identified by database searching against the most recent version of the National Center for Biotechnology Information nonredundant protein database using the XProteo computer algorithm (<http://www.xproteo.com>). The MS/MS CID data was acquired and interpreted manually to confirm the candidate peptides.

Antibodies

Antibodies used for western blotting, immunofluorescence, and ChIP-seq were as follows. H3K4me1 (Abcam ab8895), H3K4me3 (Abcam ab8580), H3K27me3 (Upstate/Millipore #07-449), H3K36me3 (Abcam ab9050), C-terminal H3 (Abcam ab1791), anti-HA (mouse monoclonal 12CA5, western and ChIP-seq), anti-HA (mouse monoclonal HA.11, Covance MMS-101R, immunofluorescence), Oct-3/4 (BD Biosciences 611202), affinity-purified rabbit polyclonal anti-GFP (Cristea et al., 2005), Ldha (Santa Cruz H-160, sc-33781), Atrx (mouse monoclonal 23c) (McDowell et al., 1999). The mouse monoclonal antibody to the RNA polymerase CTD phosphorylated on Ser5 CTD4H8 (Upstate/Millipore #05-623) was raised against a chemically synthesized phospho-Ser 5 peptide sequence from the CTD of the largest RPB1 subunit of RNAPII, and has been extensively characterized previously (Stock et al., 2007). The anti-Nestin hybridoma (Rat-401) developed by Susan Hockfield was obtained from the Developmental Studies Hybridoma bank developed under the auspices of the NICHD and maintained by the University of Iowa, Dept. of Biological Sciences, Iowa City, IA 52242. Mouse monoclonal anti-

HA (12CA5) antibody was a gift from Dr. Christina Hughes, and mouse monoclonal anti-Hira (WC119) antibody was a gift from Dr. Peter Adams.

Cellular extract preparation and metaphase chromosome spreads

Whole cell extracts were prepared by resuspending cell pellets in SDS-Laemmli sample buffer, followed by brief sonication and boiling. Chromosome spreads were performed as described (Perez-Burgos et al., 2004).

Gene expression analysis of wild-type and Hira ^{-/-} ES cells

RNA expression data for W9.5 and Hira ^{-/-} ES cells was generated from polyA RNA and random primers using the GeneChip Mouse Gene 1.0 ST Array kit (Affymetrix).

TERRA analysis

RNA was isolated using RNeasy Mini Kit (Qiagen), and TERRA analysis was performed as described (Azzalin et al., 2007; Sfeir et al., 2009).

Peptide pull-downs and protein expression

Histone peptide pull-downs were performed as described (Ooi et al., 2007). Recombinant ATRX ADD domain was a gift from Dr. Daniela Rhodes (Argentaro et al., 2007), DNMT3L was a gift from Dr. Emily Bernstein (Ooi et al., 2007). WDR5 plasmid was a gift from Dr. Alex Ruthenburg, and recombinant WDR5 protein was expressed as previously described (Ruthenburg et al., 2006).

ChIP and ChIP-seq

Crosslinking and native ChIP were performed as described (Barski et al., 2007; Lee et al., 2006), with minor modifications as described in **Appendix A**. Please

see **Appendix A** for detailed native and crosslinking ChIP and ChIP-seq methods and protocols. Purified ChIP DNA was used for cluster generation on Illumina/Solexa flow cells, and sequencing analysis was performed on an Illumina/Solexa Genome Analyzer 2 following manufacturer protocols. Primer sequences for ChIP validation are listed in **Table 5.1**. **Table 5.2** lists ChIP-seq assays performed and total tag# for each experiment.

Table 5.1: Real-time PCR primers and validation of ChIP-seq

Sequence of forward and reverse primers and mouse genomic position (mm9) is provided. Data is presented as fold enrichment over chr8 intergenic (%input at location / % input at intergenic). Primers with Hira-independent enrichment of H3.3 (TFBS and rDNA) are indicated in bold.

Genomic position	Description	H3.3 enrichment	Forward primer	Reverse primer	Primer sequence previously published
chr7:25668795-25669006	Rps19-1	3.39	TGGGGGATCATGAGT-TCAAAGCC	TGCCGGAGTCTGGT-TGTTCA	
chr7:25669207-25669375	Rps19-2	2.18	GCCACATGTCATAGTGC-CTTTTCCC	CGGGAACAAGGAGGCG-GAAA	
chr7:25669916-25670076	Rps19-4	4.64	TGAGGGAGCGGGGTGTG-GAG	CGTCGCCAGTCCAC-CCCAAA	
chr7:25671099-25671315	Rps19-5	9.68	ATGTCCACGGGATCCA-CTTTGTTC	TGCACGCTTGCTTC-CAAGAG	
chr7:25672343-25672506	Rps19-6	17.03	GCTGTCTTGCTGCTTA-CAGCCTCTAAATAG	TGCAATGGGGAT-TCTAGCCG	
chr7:25674153-25674369	Rps19-7	17.69	TCTGCAGAGTGAGTGCCA-GGACTATACA	CCATCCCAAGCTGTG-TAAACCTGG	
chr7:25674989-25675167	Rps19-8	18.19	TTGTCTCAAGACACCACT-GGAGCT	ATCTGCTCAACCGCACT-TGG	
chr7:25675686-25675847	Rps19-9	7.46	AATTCCAGCTCCTAGT-TCGTTGGGA	AACCCCTTCCAT-GAGCTTG	
chr5:143667188-143667378	Actin-3	6.15	TTGATAGTTCGCCATGGAT-GACGA	ATCGATCCCCAA-GAAAACCCCA	
chr5:143665065-143665289	Actin-5	3.75	TTAAAACTGGAACGGT-GAAGGCG	ATGGACGCGACCATC-CTCCTCTTA	
chr5:143664523-143664714	Actin-6	17.39	TGGGCTAGGCCTTGCTGAT-GGTAT	TGCAGATGGAG-GCAAAGGGT	
chr5:143664113-143664319	Actin-7	17.57	AGCTCCAGCCTTTGT-CAGTCGTGA	TTGTTTGGGTGGGT-GAAGGG	
chr5:143663236-143663455	Actin-8	16.62	GTAACAGAGGCCAGGTAA-CACAGACA	TGGCAAAGATG-GAAGGCGCA	
chr17:35640919-35641088	Oct4-4	11.84	GGAAGTGGGTGTGGGGAG-GTTGTA	AGCAGAT-TAAGGAAGGGCTAGGAC-GAGAG	Chew et al. 2005
chr17:35644291-35644494	Oct4-10	5.62	GGAGTCCCCTAG-GAAGGCATTAATAGTTT	GGATTCTCTCGGCTTCA-GACAGACTTT	Chew et al. 2005
chr17:35646759-35646950	Oct4-12	13.88	CCCTCCAGAGC-CCCTTTCAGTAA	GCACCAGGTCTC-CGATTTCATA	
chr6:122660079-122660280	Nanog-6	4.48	TCTGCTGCCTAAGCTCT-TGTGCTGT	CTGCTCTTCCAATC-CCTTCTCCC	
chr18:11051739-11051976	Gata6	5.54	CCTTCCCATACACCACAACC	CCCCTCCTTCCAAAT-TAAGC	Boyer et al. 2006
chr15:40487513-40487666	Zfp2	6.47	GGATGAAGTTCTCAGAGCT-GGT	GCGCGAAGTTTACAC-CTACTT	Bernstein et al. 2006
chr3:87773491-87773703	Nestin-4	1.64	TTTGCTATGTAGAGGCTG-GCCTCAA	GCATGGTGCACCTGTA-CACACACA	
chr3:87774987-87775219	Nestin-6	3.75	TCAGCTCGCTGCTGGAATC-CTCCG	TGCTCTCCAGCGTCT-TGACCC	
chr3:87781155-87781365	Nestin-10	1.50	ATGAGACCCAGGGTTCAT-TGAGGC	TCGTTCTCTAGCTT-GCTCTCAGCCA	
chr5:130137827-130138046	TF5-2	10.78	ACTACCCCATTTGTG-GCAGTCGATG	GATGGCACA-CAAAAGGAAGCATGG	
chr5:130139506-130139707	TF5-4	12.51	TGTGATGGCAGAATGCT-CAATCCTG	GAGCCAAATGAAC-CTTTTCTCCC	

chr2:172342091-172342310	TF2-4	10.13	TGGCGGCTAGAGCTTCTT-TAATGCA	TCGAAATCTCCGCCTGC-CAA	
chr1:72275693-72275847	TF_Oct4-43	10.34	GCAATGGTTAAGCGAGGT-TACAGGAACACA	AACCCTAAGGCCAG-GATGGTCAGTAG	Chen et al. 2008
chr1:182854966-182855138	TF_Nanog-34	7.46	GGGGAAATACAGTAATT-GACCAATCAGC	CCAAGACCACCAAAGT-TCCACCTACCCTCA	Chen et al. 2008
repeat	rDNA_90F_146R	3.42	CTTTTACTGGCTTG-GGTCTGTCG	GGAAGCGTGGCTCG-GGGAC	
repeat	rDNA_743F_811R	2.17	CTCTTCTCGTTCTGCCA-GCG	TGCCTTCCACACAC-CACCG	
repeat	rDNA_2801F_2860R	1.90	CGAGGGCAAAGGGAAA-GAGGC	CAACAGACCACCACAC-CGTCG	
repeat	rDNA_5319F_5451R	4.41	GATTCCGTGGGTGGTG-GTGC	AGAAGTTGGGGGACGC-CGAC	
repeat	rDNA_6007F_6060R	1.64	AAGGGGTGGGTGGGGTCG	AGGGGAAATCGGAG-GCGG	
repeat	rDNA_6814F_6878R	1.80	GGTCGTTGGGGGACTGT-GCC	CGTATCGGTATTTCG-GGTGTGAGC	
repeat	rDNA_7147F_7200R	1.20	GGCGGTTGTGCGGTGTGGC	GGCGGAGCGGGAA-GAAGC	
repeat	rDNA_9752F_9862R	2.99	CGGGCGTGGAATGC-GAGTG	GGGGTCTGAT-GAGCGTCGCGC	
repeat	rDNA_13366F_13419R	1.40	GTGTGGAGGCGGAGAG-GGTC	TTGGGGGGGAAAAAT-GGAAG	
chr1:136431721-136431906	TF_Nanog-37	9.16	TTCAAAACCTTTTCTAATGT-TAAGTATGGG	GCAGGCTAAACAATTG-GTGCTCTGACATC	Chen et al. 2008
chr4:135702804-135702993	TF_Nanog-36 (lincRNA)	7.75	ACACCTACGAGGCATTAT-TAAGCTAATTCC	GCCCATAAATGATTACAA-CAGCAGACTAA	Chen et al. 2008
repeat	MajSat	1.17	GACGACTTGAAAAATGAC-GAAATC	CATATTCCAGGTCCT-TCAGTGTGC	Martens et al. 2005
repeat	MinSat	1.03	CATGGAAAATGATAAAACC	CATCTAATATGTTCTA-CAGTGTGG	Martens et al. 2005

Table 5.1 continued

Table 5.2: Summary of ChIP-seq experiments

ChIP-seq method, cell type, antibody, and read numbers are indicated. Data correspond to files deposited in GEO (late 2009, accession number GSE16893).

sample number	ChIP, method, cell-type as listed in Figure S5	Antibody	GEO sample name	Tag- Size	totalRead	MappedRead
1	H3.3-HA, x-link, H3.3-HA ES	anti-HA (12CA5)	ES_H33_HA_antiHA_xChIPSeq	36	9035477	5459510
1	H3.3-HA, x-link, H3.3-HA ES	anti-HA (12CA5)	deeper sequencing replicate within sample directory above	36	7879052	3746969
1	H3.3-HA, x-link, H3.3-HA ES	anti-HA (12CA5)	deeper sequencing replicate within sample directory above	35	6936913	4505457
2	H3K36me3, x-link, H3.3-HA ES	H3K36me3 (Abcam ab9050)	ES_H33_HA_H3K36me3_xChIPSeq	36	6022772	3521419
3	H3K4me1, x-link, H3.3-HA ES	H3K4me1 (Abcam ab8895)	ES_H33_HA_H3K4me1_xChIPSeq	36	9101416	6203945
4	RNAPII CTD4H8, x-link, H3.3-HA ES	RNAPII CTD4H8 (Upstate/Millipore #05-623)	ES_H33_HA_RNAPII_CTD4H8_xChIPSeq	36	10380059	5466893
5	H3.2-HA, x-link, H3.2-HA ES	anti-HA (12CA5)	ES_H32_HA_antiHA_xChIPSeq	36	7175197	4189238
5	H3.2-HA, x-link, H3.2-HA ES	anti-HA (12CA5)	deeper sequencing replicate within sample directory above	35	7662575	4667154
5	H3.2-HA, x-link, H3.2-HA ES	anti-HA (12CA5)	deeper sequencing replicate within sample directory above	36	8022943	3754658
6	H3K36me3, x-link, H3.2-HA ES	H3K36me3 (Abcam ab9050)	ES_H32_HA_H3K36me3_xChIPSeq	36	8258652	4128488
7	H3.3-HA, x-link, H3.3-HA NPC	anti-HA (12CA5)	NP_H33_HA_antiHA_xChIPSeq	36	10468958	6374820
7	H3.3-HA, x-link, H3.3-HA NPC	anti-HA (12CA5)	deeper sequencing replicate within sample directory above	35	6507833	4593473
7	H3.3-HA, x-link, H3.3-HA NPC	anti-HA (12CA5)	deeper sequencing replicate within sample directory above	36	8474427	4224225
8	H3K36me3, x-link, H3.3-HA NPC	H3K36me3 (Abcam ab9050)	NP_H33_HA_H3K36me3_xChIPSeq	36	6099447	4124703
9	H3K4me1, x-link, H3.3-HA NPC	H3K4me1 (Abcam ab8895)	NP_H33_HA_H3K4me1_xChIPSeq	36	10408633	6652275
10	RNAPII CTD4H8, x-link, H3.3-HA ES	RNAPII CTD4H8 (Upstate/Millipore #05-623)	NP_H33_HA_RNAPII_CTD4H8_xChIPSeq	36	10340552	4836974
11	H3.2-HA, x-link, H3.2-HA NPC	anti-HA (12CA5)	NP_H32_HA_antiHA_xChIPSeq	36	8424873	3587825

11	H3.2-HA, x-link, H3.2-HA NPC	anti-HA (12CA5)	deeper sequencing replicate within sample directory above	36	6438827	2946810
11	H3.2-HA, x-link, H3.2-HA NPC	anti-HA (12CA5)	deeper sequencing replicate within sample directory above	35	6935320	4006635
12	H3K36me3, x-link, H3.2-HA, ES	H3K36me3 (Abcam ab9050)	NP_H32_HA_H3K36me3_xChIPSeq	36	6639414	4283010
13	Input, x-link, H3.3-HA hybrid ES	x-link ChIP input control. No antibody.	ESHyb_H33_HA_Input_xChIPSeq	36	9110275	5371094
14	H3.3-HA, x-link, H3.3-HA hybrid ES	anti-HA (12CA5)	ESHyb_H33_HA_antiHA_xChIPSeq	36	9676064	5894951
14	H3.3-HA, x-link, H3.3-HA hybrid ES	anti-HA (12CA5)	deeper sequencing replicate within sample directory above	36	9784425	5997655
14	H3.3-HA, x-link, H3.3-HA hybrid ES	anti-HA (12CA5)	deeper sequencing replicate within sample directory above	34	6268370	3836705
15	H3K36me3, x-link, H3.3-HA hybrid ES	H3K36me3 (Abcam ab9050)	ESHyb_H33_HA_H3K36me3_xChIPSeq	34	6463354	3789693
16	Input, x-link, H3.1S31-HA hybrid ES	x-link ChIP input control. No antibody.	ESHyb_H31S31_HA_Input_xChIPSeq	36	9173252	5397974
17	H3.1S31-HA, x-link, H3.1S31-HA hybrid ES	anti-HA (12CA5)	ESHyb_H31S31_HA_antiHA_xChIPSeq	36	5780596	3603934
17	H3.1S31-HA, x-link, H3.1S31-HA hybrid ES	anti-HA (12CA5)	deeper sequencing replicate within sample directory above	36	10048249	6076492
17	H3.1S31-HA, x-link, H3.1S31-HA hybrid ES	anti-HA (12CA5)	deeper sequencing replicate within sample directory above	36	8300087	4699124
18	H3K36me3, x-link, H3.1S31-HA hybrid ES	H3K36me3 (Abcam ab9050)	ESHyb_H31S31_HA_H3K36me3_xChIPSeq	36	8371173	4773789
19	H3.3-HA, native, H3.3-HA ES	anti-HA (12CA5)	ES_H33_HA_antiHA_nChIPSeq	36	11966966	6541403
19	H3.3-HA, native, H3.3-HA ES	anti-HA (12CA5)	deeper sequencing replicate within sample directory above	36	11704727	7346788
20	H3K4me3, native, H3.3-HA ES	H3K4me3 (Abcam ab8580)	ES_H33_HA_H3K4me3_nChIPSeq	36	8200550	5167942
21	Input, native, H3.3-HA ES	native ChIP input control. No antibody.	ES_H33_HA_input_nChIPSeq	36	11427929	6039408
22	H3.2-HA, native, H3.2-HA ES	anti-HA (12CA5)	ES_H32_HA_antiHA_nChIPSeq	36	2133418	1361872
22	H3.2-HA, native, H3.2-HA ES	anti-HA (12CA5)	deeper sequencing replicate within sample directory above	36	13362561	7074823

Table 5.2 contd.

22	H3.2-HA, native, H3.2-HA ES	anti-HA (12CA5)	deeper sequencing replicate within sample directory above	32	9664647	5528131
23	H3K4me3, native, H3.2-HA ES	H3K4me3 (Abcam ab8580)	ES_H32_HA_H3K4me3_nChIPSeq	32	6402480	4415020
24	H3K27me3, native, H3.2-HA ES	H3K27me3 (Upstate/Millipore #07-449)	ES_H32_HA_H3K27me3_nChIPSeq	32	8741662	5428037
25	Input, native, H3.2-HA ES	native ChIP input control. No antibody.	ES_H32_HA_input_nChIPSeq	32	8715354	5479449
26	H3.3-EYFP, native, H3.3-EYFP ES	anti-GFP (Cristea et al., 2005)	ES_H33_EYFP_anti-GFP_nChIPSeq	36	5193695	3744380
26	H3.3-EYFP, native, H3.3-EYFP ES	anti-GFP (Cristea et al., 2005)	deeper sequencing replicate within sample directory above	36	5595554	3395192
26	H3.3-EYFP, native, H3.3-EYFP ES	anti-GFP (Cristea et al., 2005)	deeper sequencing replicate within sample directory above	36	6047143	3191111
27	Input, native, H3.3-EYFP ES	native ChIP input control. No antibody.	ES_H33_EYFP_Input_nChIPSeq	36	5726370	3531774
28	H3K4me1, native, H3.3-EYFP ES	H3K4me1 (Abcam ab8895)	ES_H33_EYFP_antiH3K4me1_nChIPSeq	36	5780578	4385502
29	H3K36me3, native, H3.3-EYFP ES	H3K36me3 (Abcam ab9050)	ES_H33_EYFP_antiH3K36me3_nChIPSeq	36	8600269	5174908
30	H3.3-EYFP, native, Hira +/- H3.3-EYFP ES	anti-GFP (Cristea et al., 2005)	Hiranull_ES_H33_EYFP_antiGFP_nChIPSeq	36	11564437	6590765
30	H3.3-EYFP, native, Hira +/- H3.3-EYFP ES	anti-GFP (Cristea et al., 2005)	deeper sequencing replicate within sample directory above	36	12460001	5457471
31	H3K4me1, native, Hira +/- H3.3-EYFP ES	H3K4me1 (Abcam ab8895)	Hiranull_ES_H33_EYFP_antiH3K4me1_nChIPSeq	36	8915573	6420547
32	H3K36me3, native, Hira +/- H3.3-EYFP ES	H3K36me3 (Abcam ab9050)	Hiranull_ES_H33_EYFP_antiH3K36me3_nChIPSeq	36	7999631	4416253
33	Input, native, Hira +/- H3.3-EYFP ES	native ChIP input control. No antibody.	Hiranull_ES_H33_EYFP_Input_nChIPSeq	36	8156833	5292703
34	H3.3-EYFP, native, Atrx-flox H3.3-EYFP ES	anti-GFP (Cristea et al., 2005)	Atrx-flox_ES_H33_EYFP_antiGFP_nChIPSeq	36	6476017	4304375
34	H3.3-EYFP, native, Atrx-flox H3.3-EYFP ES	anti-GFP (Cristea et al., 2005)	deeper sequencing replicate within sample directory above	36	9088918	5313474
34	H3.3-EYFP, native, Atrx-flox H3.3-EYFP ES	anti-GFP (Cristea et al., 2005)	deeper sequencing replicate within sample directory above	36	9236425	6207716

Table 5.2 contd.

35	Input, native, Atrxflox H3.3-EYFP ES	native ChIP input control. No antibody.	Atrxflox_ES_H33_EYFP_Input_nChIPSeq	36	6060126	4117751
36	H3.3-EYFP, native, Atrx-null H3.3-EYFP ES	anti-GFP (Cristea et al., 2005)	Atrxnull_ES_H33_EYFP_antiGFP_nChIPSeq	36	7806156	4811353
36	H3.3-EYFP, native, Atrx-null H3.3-EYFP ES	anti-GFP (Cristea et al., 2005)	deeper sequencing replicate within sample directory above	36	9685286	6402420
36	H3.3-EYFP, native, Atrx-null H3.3-EYFP ES	anti-GFP (Cristea et al., 2005)	deeper sequencing replicate within sample directory above	36	8575771	5495831
37	Input, native, Atrxnull H3.3-EYFP ES	native ChIP input control. No antibody.	Atrxnull_ES_H33_EYFP_Input_nChIPSeq	36	7102030	4758929
38	anti-HA, x-link, W9.5 wild-type ES	anti-HA (12CA5) control. No HA epitope in untagged ES cells	ES_wild_type_antiHA_control_xChIPSeq	36	7099906	3065689

Table 5.2 contd.

Data Analysis

All ChIP-seq data analysis was performed in collaboration with Dr. Deyou Zheng. ChIP-seq or input reads were mapped to the mouse genome (build 37, or mm9) using the ELAND alignment software within the Illumina Analysis Pipeline. Profiles in specific genomic regions were displayed in the Affymetrix Integrated Genome Browser. For analysis of repetitive elements, reads were aligned directly to a library of mouse consensus repetitive sequences (<http://www.girinst.org>). For analysis at TSS and TES, we segregated reference gene (refSeq) annotation into low, medium and high expression based on a previous microarray analysis (Mikkelsen et al., 2007). For analysis of TFBS, binding sites for 13 TFs in mouse ES cells were obtained from a previous ChIP-seq analysis (Chen et al., 2008). To generate density profiles \pm 5kb around TSS, TES, bivalent genes, and TFBS, we used a sliding window method which summed the number of ChIP-seq reads in each 200 bp window. The sums were then normalized by the number of genes or TFBS in each group and the total mapped reads obtained in a ChIP-seq experiment.

Genome-wide density maps of H3 variants and determination of enriched regions. All reads were mapped to the mouse genome (build 37, or mm9) using the ELAND alignment software within the Illumina Analysis Pipeline in multi/extended mode. Unique reads mapped to a single best-matching location with no more than two mismatches were kept. During our data analysis, each read was extended 100 nt in its 3' direction to account for the estimated ChIP DNA fragment size of ~300 bp. Using these extensions, we generated genome-wide distributions of H3 variants, histone modifications, and RNAPII represented by the number of reads spanning individual genomic locations. These profiles were displayed in the Affymetrix Integrated Genome Browser (IGB) and also used for the determinations of genomic regions enriched with ChIP-Seq signals. Like H3K36me3, the signals for H3.3 are

very dispersed and often found in large extended regions (**Figure 2.11, 2.19., 2.22, 2.26**). As a result, my collaborators and I tried several existing peak identification programs and found that they were not very effective in identification of H3.3 enriched regions. We therefore employed a simple segmentation algorithm. We first applied a Poisson distribution to model the genome-wide ChIP-Seq profile, and then selected positions with read coverage above a threshold that was determined for each experiment specifically at the p-value of 0.1. Adjacent positions above this threshold and less than 600 bp apart were then merged to form genomic blocks, which were subsequently filtered by a minimal length of 300 bp to generate candidates of enriched regions. The final enriched regions, furthermore, must also have twice more reads in a ChIP assay than its control input run (after normalization by read depths). As a comparison, 95% and 97% of the peaks identified by our method for H3K4me1 and H3K4me3 ChIP-Seq experiments, respectively, were also determined as peaks by a popular ChIP-Seq peak analysis program, MACS (Zhang et al., 2008), suggesting that we only called peaks of high confidence. **Figure 2.12** and **Figure 2.14** summarize the numbers of enriched regions for selected ChIP-experiments and their spatial relationships to gene annotation and transcription factor binding sites.

The relative depositions of H3 variants in repetitive elements. Reads were aligned directly to a library of mouse consensus repetitive sequences (<http://www.girinst.org>). The proportion of reads aligning to each class of repeats was computed for H3.3, H3.2, and H3.1S31, on the condition that the alignment length and sequence identity must be ≥ 25 -bp and $\geq 90\%$, respectively. The resulting numbers were compared to the corresponding proportions calculated for input controls to yield relative fold-enrichments.

Distributions of H3 variants in transcription start and end sites. The reference gene (refSeq or known genes) annotation was downloaded from the UCSC browser (<http://genome.ucsc.edu>), and genes were segregated into low, medium and high expression based on expression data obtained from a previous microarray analysis (Mikkelsen et al., 2007). The expression values for these three categories were defined as 5~20, 50~200, and >500, respectively. Based on the CpG contents in their promoters, genes were further divided into three types: high (HCP), intermediate (ICP) and low CpG (LCP) promoters (Mikkelsen et al., 2007). Data shown here are mostly for low, medium and high expression HCP genes (~1800 genes in each group). The three groups of genes were comparatively analyzed for density profiles of H3 variants or modifications around their transcription start sites (TSSs) and transcription end sites (TESs). Briefly, after dividing the ± 5 kb regions across TSSs (or TESs) into 50 non-overlapping windows (200-bp in size), we summed the number of ChIP-Seq reads ending in each window over all genes. The sums were then normalized by the gene number in each group and the total of mapped reads obtained in a ChIP-Seq experiment to generate final aggregated ChIP-Seq profiles of H3 variants or modifications. This sliding window method was also applied to analyze a set of genes with “bivalent” marks at their promoters in mouse ES cells but not in mouse NPC (Mikkelsen et al., 2007).

The relative enrichment of H3 variants in TFBS. The lists of binding sites for 13 transcription factors in mouse ES cells were obtained from a previous ChIP-Seq analysis (Chen et al., 2008). RefSeq gene annotation was used to segregate TFBS into three groups: promoter (± 5 kb of TSS), intergenic (5 kb away from genes), and otherwise gene body. After extending each of these TFBS from its 5' and 3' direction to span 10-kb sequence, we applied a sliding-window approach, similar to the one for TSSs, to generate density profiles over TFBS for different

H3 variants or histone modifications. (Chen et al., 2008) has also defined a set of multiple transcription factor-binding loci (MTL) that were bound by more than four different transcription factors. For these MTL, we computed the read density (number of reads per 5-kb sequence) to characterize the correlation between number of bound factors and the abundance of different H3 variants.

APPENDIX

CHROMATIN IMMUNOPRECIPITATION (CHIP) AND DETAILED PROTOCOLS FOR CHIP-SEQ

The crosslinking ChIP-seq protocol described below was adapted from a ChIP-chip protocol originally published by the laboratory of Dr. Richard Young (Lee et al., 2006b). In addition, a number of the steps in this protocol grew out of detailed discussions with Dr. Tom Milne, Sarah Whitcomb, and Dr. Nicolai Siegel. For native ChIP-seq, I adapted a native ChIP protocol from Dr. Keji Zhao's laboratory (Barski et al., 2007).

Crosslinking and native ChIP

ChIP is generally performed with either formaldehyde-mediated crosslinking of proteins and nucleic acids within living cells followed by sonication (crosslinking, x-ChIP) or by micrococcal nuclease digestion of harvested cell nuclei and isolation of solubilized mono or short oligonucleosomes (native ChIP). Crosslinking ChIP 'freezes' protein and nucleic acid interactions and enables the analysis of nucleosomal as well as non-nucleosomal proteins such as TFs and RNA polymerase, while native ChIP is more suitable for the analysis of core chromatin and nucleosomal components such as histones and histone modifications (Kuo and Allis, 1999; O'Neill and Turner, 2003) (See **Figure 1.8**). I used a slightly modified version of described protocols (Barski et al., 2007; Lee et al., 2006b) to perform both native and crosslinking ChIP. Both protocols use Invitrogen Dynal magnetic beads (Invitrogen Dynabeads anti-mouse M-280 #112-02, or Dynabeads anti-rabbit M-280 #112-04, or Dynabeads Protein A #100-02D). I have found that these magnetic beads generate low background when washed as described (Barski et

al., 2007; Lee et al., 2006b), and also are free of salmon sperm DNA which might otherwise contaminate ChIP-seq results. The use of a protocol that omits salmon sperm or other non-specific DNA blocking agents is absolutely critical for ChIP-seq, as such non-specific nucleic acids overwhelm the actual sequence reads from specifically immunoprecipitated DNA (D Zheng, personal communication).

For x-ChIP, I harvested embryonic stem (ES) cells and neural precursor cells (NPCs) immediately prior to fixation with 1% paraformaldehyde, as I found that crosslinking on gelatin-coated plates also crosslinks gelatin. Approximately 1×10^7 cells were used for each ChIP. For sonication, I used a Bioruptor (Diagenode) and optimized sonication conditions to generate DNA fragments of approximately 300-500bp. Following a published x-ChIP protocol (Lee et al., 2006b) through overnight crosslink reversal at 65°C, RNase digestion, and proteinase K digestion, I then isolated ChIP DNA and ChIP input DNA using the Qiagen Qiaquick PCR purification kit (28104), eluting in 50µl 10mM Tris pH8.5. For native ChIP, I used the published protocol (Barski et al., 2007) as described, with slight modifications. Following the final wash, I eluted off of magnetic beads using 50mM Tris-HCl pH 8.0, 10mM EDTA, and 1% SDS, as described in the x-link ChIP protocol (Lee et al., 2006b). Following elution, I digested proteins with proteinase K (Roche, recombinant, PCR grade, 03115828001), and isolated DNA with Qiagen PCR purification kit as described for x-ChIP protocol. Following DNA isolation from native ChIP, ChIP-seq library preparation is largely identical to the crosslinking ChIP-seq protocol.

Validation of ChIP prior to ChIP-seq

To validate the ChIP, I used real-time PCR with SYBR Green PCR master mix (Applied Biosystems) to amplify ChIP samples and ChIP inputs, determining % input of ChIP samples at target locations and control intergenic locations. Real-

time PCR is critical for ChIP quality control prior to ChIP-seq. Primer sequences and ES % inputs are listed in **Table 5.1**. To confirm telomere enrichment, ChIP-seq DNA libraries were also probed with a TTAGGG repeat probe and BamHI repeat probe as described (Sfeir et al., 2009).

ChIP-seq library preparation

For ChIP-seq of x-link or native ChIP DNA, I took 30 ul of remaining ChIP DNA or ~180ng of input DNA, repaired DNA ends to generate blunt-ended DNA using the Epicentre DNA ENDRepair kit (Epicentre Biotechnologies, cat# ER0720), and purified repaired DNA using Qiagen PCR purification kit (28104). Following DNA END Repair, I added A bases to the 3' end of the DNA fragments using Klenow Fragment (NEB M0212L), and purified DNA using Qiagen MinElute (28004). I ligated Illumina/Solexa adapters (#FC-102-1003) to DNA fragments overnight, using T4 DNA ligase (NEB M0202L). Following overnight ligation, I purified adapter-ligated DNA fragments with Qiagen MinElute. To generate libraries for Solexa sequencing, I performed 18 cycles of PCR with Illumina/Solexa primers 1.1 and 2.1, checked the fragment size for 1/10 of the amplified library on an agarose gel, and purified the remaining ChIP-seq library using Qiagen MinElute. I did not use gel purification to isolate the ChIP-seq library, as standard methods of gel purification have been shown to decrease the representation of A+T-rich sequences (Quail et al., 2008), and I have also found decreased DNA yield following gel purification (data not shown). Detailed protocols for all of these steps are provided below. For details of Solexa sequencing and ChIP-seq data analysis, please see Chapter Five.

Detailed protocols for crosslinking ChIP-seq

Prepare cells and paraformaldehyde

1. For ChIP-seq, seed cells so as to reach $\sim 1 \times 10^7$ cells / IP.

Prepare 10% paraformaldehyde stock as follows:

Microwave 18ml of PBS in a 50ml tube placed in a beaker with water until water starts to boil (leave lid of tube slightly ajar). Do the subsequent steps in a hood. Add 2 grams of paraformaldehyde (wear a mask when weighing this out) and 140 μ l of 1M KOH. Vortex (be careful since this is a hot solution) and the paraformaldehyde should dissolve. QS to 20 ml. Cool to 37° before use. Can be stored frozen indefinitely.

Cross-link and fix cells

2. Fix cells for 10 minutes with 1% freshly made or frozen paraformaldehyde in media @ 37°. Count another plate in parallel to have some idea of cell #.

* aside for BM (and other non-adherent) cells - fix in a small volume of Pf/media (~5ml) and then dilute 10X with 45 ml RPMI.

Fix trypsinized ES or non-adherent cells in 10ml warm media + 1ml 10% PF, rotating in a 50ml conical tube at RT. (Fixing ES cells on a gelatinized plate crosslinks the gelatin, and generates a goopy, disastrous mess. Hence, trypsinize ES cells first.)

For adherent cells grown on tissue culture dishes without gelatin, add 1 ml 10% paraformaldehyde to 10 ml media for 10 min at 37.

3. Quench unreacted paraformaldehyde with 2.5M glycine, 5 min @ RT, rotating if the cells are already out of the tissue culture dish. (2.5M glycine is ~20X, ie 500 μ l / 10ml media with 1% paraformaldehyde.)
4. Dilute with 40ml cold media with FBS. This step may neutralize any residual trypsin or formaldehyde, but is not required.
5. Spin fixed cells down. Remove media/paraformaldehyde and rinse and resuspend cells with 5ml PBS + protease inhibitors (PI). Add PI just before use. Transfer to a new tube (15ml or 1.5 ml tube). Cold PBS + inhibitors rinse. (Make sure to include protease inhibitors in last PBS rinse prior to flash freeze. I use 1 tablet of PI / 7ml of PBS. Roche, complete mini EDTA-free 11-836-170-001)
6. Pellet cells @ 1200rpm, 5'. Flash-freeze pellet in liquid nitrogen and store @ -80°. Alternately, you can continue with protocol. If continuing, note that a single freeze/thaw helps break up cells.

Bind antibody to Dynabeads/ block beads (afternoon prior to ChIP)

Block Solution (for 1 ChIP)

500ml	1x PBS
2.5g BSA	--> 0.5% w/v

1. 150uL Dynabeads (anti-mouse M-280, cat# 112-02, or anti-rabbit M-280 cat#112-04, to eppendorff (for 2 ChIPs)
2. 3x wash 1ml block solution: gently mix, collect beads, remove sup
3. resuspend beads in 500uL block solution + ~15ug antibody (20ul H3K4me3 ab8580) (scale antibody and beads accordingly: 100ul beads, ~10ug antibody – aim for more rather than less antibody – you will wash any remaining antibody away after binding to the beads). Per IP, you will use > 50ul of antibody bound beads, >5ug antibody, for ~ 1×10^7 cells.
4. O/N rotating @ 4 deg
5. 3x wash in 1ml block solution
6. resuspend beads in 150uL block solution

Lysis and Sonication

LB1 for 100mL

5mL	1M Hepes-KOH, pH 7.5	-->	50mM
2.8mL	5M NaCl	-->	140mM
0.2mL	0.5M EDTA	-->	1mM
20mL	50% glycerol	-->	10%
5mL	10% NP-40	-->	0.5%
2.5mL	10% Triton x-100	-->	0.25%
64.5mL	ddH2O		

1. pipette resuspend x-linked cell pellet in 1mL LB1 + **protease inhibitors (PI)** (per 1×10^7 cells! ie, for 9×10^7 cells, use 9ml) -- if using large cell #s, the next few steps can be done in a conical -- just make sure to sufficiently resuspend pellet.
2. 10 min rotating @ 4deg
3. spin 5 min @ 1,350g @ 4deg

LB2 for 100mL

1mL	1M Tris-HCl, pH 8.0	-->	10mM
4mL	5M NaCl	-->	200mM
0.2mL	0.5M EDTA	-->	1mM
0.1mL	0.5M EGTA	-->	0.5mM
94.7mL	ddH2O		

1. pipette resuspend pellet in 1mL LB2 + **PI**
2. 10 min rock @ RT
3. spin 5 min @ 1,350g @ 4deg

LB3 for 100mL

1mL	1M Tris-HCl, pH 8.0	-->	10mM
2mL	5M NaCl	-->	100mM
0.2mL	0.5M EDTA	-->	1mM
0.1mL	0.5M EGTA	-->	0.5mM
1mL	10% Na-Deoxycholate	-->	0.1%

2.5mL	20% N-lauroylsarcosine	-->	0.5%
93.2mL	ddH ₂ O		

1. resuspend pellet in 600µl LB3 + PI, for 1*10⁷ cell# ChIP-seq)
Note: for multiple ChIPs from the same sample, resuspend in more LB3, for ChIP-seq, ~600ul LB3 / 1*10⁷ cells / IP, pool resuspended cells.
pass through 27 gauge needle 2-3x to break up cell clumps
-split into aliquots for Bioruptor sonication, no more than 300ul per tube
2. sonicate BioRuptor 7x 30"/30" in ice water, high power, replace ice, repeat 7x sonication.
3. Pool sonicated chromatin from one sample in a 15ml conical. Add 1/10 volume of 10% Triton X-100 to sonicated lysate (DO NOT FORGET TO ADD TRITON OR YOU WILL NOT SOLUBILIZE YOUR NUCLEAR MEMBRANE) (for ex., 80ul for 800ul).
4. Aliquot and spin 10min @ 20,000g @ 4deg to crack nuclei; save sup = IP input

Aliquot 600µl sup for each IP.
save 30uL (5%) freeze for sonication check (agarose)
save 30uL (5%) freeze for ChIP INput (compare w/ChIP elute)
Freeze remaining sup in aliquots for future IPs.

IP Chromatin

1. add 70uL antibody bound bead mix to sonication sup (600ul) – if you are using fewer than 1*10⁷ cells per IP, you can scale your antibody and beads down accordingly
2. rotate O/N @4deg

Wash

RIPA Wash Buffer (for 250mL)

12.5mL	1M Hepes-KOH, pH 7.6	-->	50mM
125mL	1M LiCl	-->	500mM
0.5mL	0.5M EDTA	-->	1mM
25mL	10% NP-40	-->	1%
17.5mL	10% Na-Deoxycholate	-->	0.7%
69.5mL	ddH ₂ O		

@ 4 deg

0. start water bath to 65deg for elution and rev X-link!
1. 3-7x (Suggestion: 8x !) wash 1mL RIPA wash buffer - on magnetic stand
2. wash 1mL TE + 50mM NaCl
3. spin (3 min @ 960g @ 4deg and carefully remove residual TE with a pipet – (do NOT attempt to use magnet at this step or you will lose your sample)

Elute

Elution Buffer (for 100mL)

5mL	1M Tris-HCL, pH 8.0	-->	50mM
2mL	0.5M EDTA	-->	10mM
10mL	10% SDS	-->	1%
83mL	ddH ₂ O		

1. add 210uL elution buffer
2. elute 30 min @ 65deg- resuspend beads every 2 min by briefly vortexing – **make sure caps on tubes are not loose** – also, the 65 degree water bath can bleed the marker – write in strong marker and keep track of tubes
3. spin down 1 min @ 16,000g @ RT
4. remove 200uL sup transfer to new eppendorff

Reverse Crosslinks

1. bring ChIP IN samples (30uL) and sonication check samples (30uL) frozen yesterday to 200uL with elution buffer
2. ChIP elutions (IP), ChIP IN, sonication check samples O/N @ 65 deg

Digest RNA and protein

1. add 2uL 500 ug/ml RNase (Roche, RNase, DNase free, Cat. No. 11 119 915 001), vortex briefly to mix
2. 1 hour @ 37deg (bring water bath to 55 deg)
3. add 2uL 14-22 mg/mL Proteinase K (Roche, proteinase K, recombinant, PCR grade, Cat. No. 03 115 828 001) proteinaseK --> 0.2ug/mL final, vortex briefly.
4. 2 hours @ 55 deg

DNA purification

5. Recover DNA by Qiagen PCR purification kit. Add 1ml of Binding reagent PBI to each RNased, PK digested sample. Mix.
6. Bind DNA through Qiagen spin columns as described in protocol, using no more than 0.75ml of sample at one time / column. I.e., apply the first 0.7ml, spin, discard flow-through, then apply then second ~0.5ml, spin, etc. Continue as described in Qiagen PCR purification protocol. **Elute with 50ul Qiagen elution buffer EB. Take 5 ul of ChIPed, purified DNA for real-time PCR analysis. Store remaining 45ul at -20, and use for ChIP-seq only pending real-time PCR analysis.**

Real-time PCR analysis

Make 1:100 dilution of each input and sample (5ul + 495ul H₂O) for real-time PCR. I use SYBR green (Applied Biosystems SYBR Green PCR Master Mix, #4309155) for real-time quantification.

For each primer set, make the appropriate mix. Keep on ice until adding to plate. Use 15λ of each mix / well of 96-well real-time PCR plate. Add 10λ of appropriate

1:100 dilution to each well for 25 λ total. Make sure to pipet up and down. Quick spin 96 well plate for 30 sec. at 450rpm to make sure mix is even in all wells.

	1:100 DNA	10.0 λ
2X SYBR Mix:	12.5 λ	
	Fwd primer	0.5 λ
	Rev primer	0.5 λ
	H2O	<u>1.5 λ</u>
	Total	25.0 λ

Preparing ChIP samples for ChIP-seq

Note on yield of ChIP DNA: ChIP DNA amounts are generally <10ng / μ l, and not accurately quantifiable by nanodrop. Therefore, I do not base my decision about whether to use the sample for ChIP-seq on the relative DNA yield, but rather on real-time PCR analysis, looking at % input at an enriched location vs % input at an unenriched location. Antibody quality and the relative biological abundance of their epitopes can be variable, but the real-time PCR analysis is critical to make sure that you are sequencing high quality, specifically IPed material.

Note on yield of input DNA: Typical yields of purified input DNA (pre-IP) are ~50-100+ ng / μ l, with 50 μ l EB representing 5% (2.5 to 5 μ g) of the original input (~50-100 μ g.)

DNA end repair

1. Repair DNA ends to generate blunt-ended DNA using the Epicentre DNA ENDRepair

kit (Epicentre Biotechnologies, cat# ER0720)

1-34 μ l DNA (0.3 μ g)

5 μ l 10x End repair buffer (330 mM Tris-acetate, pH7.8, 660 mM potassium acetate, 100 mM magnesium acetate, 5 mM DTT)

5 μ l 2.5 mM each dNTPs

5 μ l 10 mM ATP

x μ l H2O to adjust the reaction volume to 49 μ l

1 μ l End-Repair Enzyme mix (T4 DNA Pol + T4 PNK)

For ex.:

30 μ l ChIP DNA from remaining 45 μ l (~180ng of input DNA, add H2O to reach 30 μ l)

Make mix qs to 19 μ l 5 μ l 10x End repair buffer

5 μ l 2.5 mM dNTP mix

5 μ l 10 mM ATP

4 μ l H2O to adjust the reaction volume to 49 μ l

1 μ l End-Repair Enzyme mix (T4 DNA Pol + T4 PNK)

Keep at room temperature for 45 min.

2. Use QIAquick PCR purification kit and elute in 34 μ l Qiagen EB

Add 'A' bases to the 3' end of the DNA fragments

3. Add "A" to 3' ends

34 μ l DNA from above

Make mix qs to 13 μ l

5 μ l Klenow buffer NEB2 (10X)

1 μ l dATP (10mM)

7 μ l dH₂O

3 μ l Klenow enzyme (Klenow Fragment (3' \rightarrow 5' exo-), NEB M0212L 5000 un/ml)

Total volume: 50 μ l

Incubate for 30 min at 37°C.

4. Purify DNA with Qiagen MinElute (cat. 28004) and elute in 11 μ l Qiagen EB.

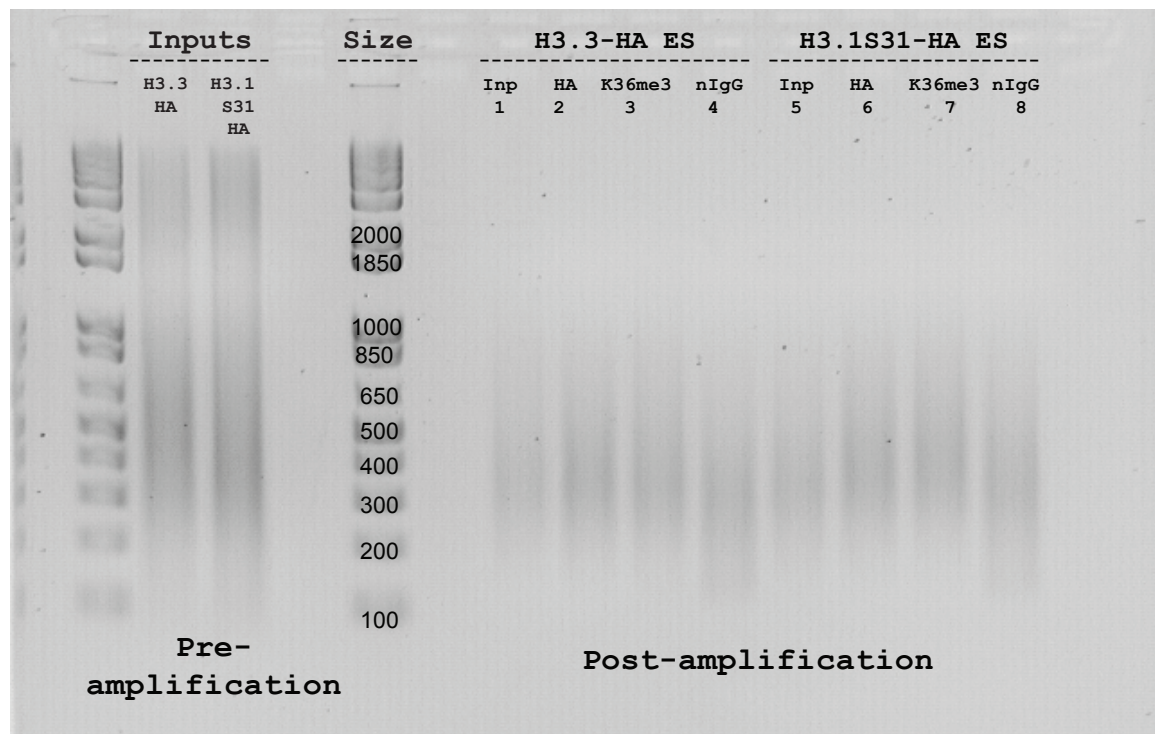


Figure A1: Gel image of crosslinking ChIP-seq DNA pre and post-amplification for Solexa sequencing

Sonicated ChIP inputs and ChIP samples were amplified and prepared for Solexa sequencing. ~180ng of ChIP input DNA and 35 μ l of each ChIP sample were used in end repair, A addition, adapter ligation, and amplification (see detailed steps above). 300 ng of ChIP input DNA was loaded for comparison (pre-amplification, left). Following 18 cycle PCR amplification with Solexa primers 1.1 and 2.1, 4 μ l of unpurified 50 μ l PCR reaction was loaded in each lane to assess fragment sizes of the amplified ChIP-seq libraries (post-amplification, right). DNA ladder (center) indicates that the ChIP-seq libraries are predominantly between 300-500bp. The remaining 46 μ l post-amplification PCR reaction was then directly purified with Qiagen MinElute (without gel purification) and eluted into 15 μ l EB for Solexa sequencing.

Ligate adapters to DNA fragments

5. Linker ligation. The oligo mix should contain 10x molar excess of the DNA templates.

10 µl DNA
Make mix qs to 17µl
3 µl 10x T4 DNA ligase buffer
13 µl H₂O (nuclease free)
1 µl Solexa adapter oligo mix (try 1:10 dilution)
3 µl T4 DNA ligase (NEB M0202L, 400 units/µl)
Total volume: 30 µl

Incubate at room temperature for 15 min and then overnight in PCR program below.

1) 4°C for 2 hrs. 2) 0.1° / s to 16°C. 3) 16°C for 20min. 4) 0.1° / s to 4°C.
5) Go to 1, 6 times. 6) 4°C forever.

6. Purify DNA with QIAGEN MinElute and elute in 11µl Qiagen EB.

Amplify adapter-modified DNA fragments by PCR to generate ChIP-seq library

7. Amplify the DNA using Solexa primers and enzyme mix

10 µl of DNA
Make mix qs to 39.5µl
10 µl of 5X Phusion HF buffer
1.5 µl of dNTP mix (10mM dNTP)
0.75 µl **Solexa primer 1.1**
0.75 µl **Solexa primer 2.1**
1.5 µl DMSO
25 µl dH₂O (nuclease free)

0.5 µl Phusion

Total volume: 50 µl

Denature at 98°C for 30 sec. 98°C, 10sec; 65°C, 30sec; 72°C, 30sec. (18 cycles) 72°C, 5 min. hold @ 4 degrees.

8. Try 18 cycles first, check 4 µl of product on 2% gel. If the band is not clearly visible, do 3 more cycles. Check again.

9. Purify remaining 46µl DNA with Qiagen MinElute (cat. 28004) and elute in 15µl Qiagen EB.

10. Nanodrop and store at -20 prior to Solexa sequencing. Typical nanodrop yields post amplification and MinElute purification are 170-270 ng / µl, 15 µl total volume in Qiagen EB.

REFERENCES

- Abmayr, S.M., and Workman, J.L. (2003). Transcription factors prominently in Lasker Award to Roeder. *Cell* **115**, 243-246.
- Adams, P.D. (2007). Remodeling of chromatin structure in senescent cells and its potential impact on tumor suppression and aging. *Gene* **397**, 84-93.
- Ahmad, K., and Henikoff, S. (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol Cell* **9**, 1191-1200.
- Alaiya, A., Al-Mohanna, M., and Linder, S. (2005). Clinical cancer proteomics: promises and pitfalls. *J Proteome Res* **4**, 1213-1222.
- Alberts, B. (2002). *Molecular biology of the cell*, 4th edn (New York, Garland Science).
- Albig, W., Bramlage, B., Gruber, K., Klobeck, H.G., Kunz, J., and Doenecke, D. (1995). The human replacement histone H3.3B gene (H3F3B). *Genomics* **30**, 264-272.
- Albig, W., and Doenecke, D. (1997). The human histone gene cluster at the D6S105 locus. *Hum Genet* **101**, 284-294.
- Allfrey, V.G., Faulkner, R., and Mirsky, A.E. (1964). Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. *Proc Natl Acad Sci U S A* **51**, 786-794.
- Allfrey, V.G., Littau, V.C., and Mirsky, A.E. (1963). On the role of of histones in regulation ribonucleic acid synthesis in the cell nucleus. *Proc Natl Acad Sci U S A* **49**, 414-421.
- Allis, C.D., Glover, C.V., Bowen, J.K., and Gorovsky, M.A. (1980). Histone variants specific to the transcriptionally active, amitotically dividing macronucleus of the unicellular eucaryote, *Tetrahymena thermophila*. *Cell* **20**, 609-617.
- Allis, C.D., Jenuwein, T., Reinberg, D., and Caparros, M.L., eds. (2007). *Epigenetics* (Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press).
- Allis, C.D., Richman, R., Gorovsky, M.A., Ziegler, Y.S., Touchstone, B., Bradley, W.A., and Cook, R.G. (1986). hv1 is an evolutionarily conserved H2A variant that is preferentially associated with active genes. *J Biol Chem* **261**, 1941-1948.
- Allis, C.D., Ziegler, Y.S., Gorovsky, M.A., and Olmsted, J.B. (1982). A conserved histone variant enriched in nucleoli of mammalian cells. *Cell* **31**, 131-136.

Allsopp, R.C., Vaziri, H., Patterson, C., Goldstein, S., Younglai, E.V., Futcher, A.B., Greider, C.W., and Harley, C.B. (1992). Telomere length predicts replicative capacity of human fibroblasts. *Proc Natl Acad Sci U S A* **89**, 10114-10118.

Almouzni, G., and Mechali, M. (1988). Assembly of spaced chromatin promoted by DNA synthesis in extracts from *Xenopus* eggs. *Embo J* **7**, 665-672.

Anderson, H.E., Wardle, J., Korkut, S.V., Murton, H.E., Lopez-Maury, L., Bahler, J., and Whitehall, S.K. (2009). The fission yeast HIRA histone chaperone is required for promoter silencing and the suppression of cryptic antisense transcripts. *Mol Cell Biol* **29**, 5158-5167.

Argentaro, A., Yang, J.C., Chapman, L., Kowalczyk, M.S., Gibbons, R.J., Higgs, D.R., Neuhaus, D., and Rhodes, D. (2007). Structural consequences of disease-causing mutations in the ATRX-DNMT3-DNMT3L (ADD) domain of the chromatin-associated protein ATRX. *Proc Natl Acad Sci U S A* **104**, 11939-11944.

Armstrong, J.A. (2007). Negotiating the nucleosome: factors that allow RNA polymerase II to elongate through chromatin. *Biochem Cell Biol* **85**, 426-434.

Avery, O.T., MacLeod, C.M., and McCarty, M. (1979). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Inductions of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J Exp Med* **149**, 297-326.

Azzalin, C.M., and Lingner, J. (2008). Telomeres: the silence is broken. *Cell Cycle* **7**, 1161-1165.

Azzalin, C.M., Reichenbach, P., Khoriantuli, L., Giulotto, E., and Lingner, J. (2007). Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* **318**, 798-801.

Banaszynski, L.A., Chen, L.C., Maynard-Smith, L.A., Ooi, A.G., and Wandless, T.J. (2006). A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* **126**, 995-1004.

Bannister, A.J., Schneider, R., Myers, F.A., Thorne, A.W., Crane-Robinson, C., and Kouzarides, T. (2005). Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J Biol Chem* **280**, 17732-17736.

Bao, Y., and Shen, X. (2006). Asf1, a loveseat for a histone couple. *Cell* **127**, 458-460.

Barman, H.K., Takami, Y., Nishijima, H., Shibahara, K., Sanematsu, F., and Nakayama, T. (2006). Histone acetyltransferase-1 regulates integrity of cytosolic histone H3-H4 containing complex. *Biochem Biophys Res Commun* **373**, 624-

630.

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.

Barski, A., and Zhao, K. (2009). Genomic location analysis by ChIP-Seq. *J Cell Biochem*.

Baxevanis, A.D., Godfrey, J.E., and Moudrianakis, E.N. (1991). Associative behavior of the histone (H3-H4)₂ tetramer: dependence on ionic environment. *Biochemistry* 30, 8817-8823.

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R., *et al.* (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* 120, 169-181.

Bernstein, B.E., Meissner, A., and Lander, E.S. (2007). The mammalian epigenome. *Cell* 128, 669-681.

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.* (2006a). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125, 315-326.

Bernstein, E., and Allis, C.D. (2005). RNA meets chromatin. *Genes Dev* 19, 1635-1655.

Bernstein, E., Duncan, E.M., Masui, O., Gil, J., Heard, E., and Allis, C.D. (2006b). Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Molecular and cellular biology* 26, 2560-2569.

Biddick, R., and Young, E.T. (2005). Yeast mediator and its role in transcriptional regulation. *C R Biol* 328, 773-782.

Black, J.C., Choi, J.E., Lombardo, S.R., and Carey, M. (2006). A mechanism for coordinating chromatin modification and preinitiation complex assembly. *Mol Cell* 23, 809-818.

Blackburn, E.H., Greider, C.W., and Szostak, J.W. (2006). Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging. *Nat Med* 12, 1133-1138.

Blackwell, C., Martin, K.A., Greenall, A., Pidoux, A., Allshire, R.C., and Whitehall, S.K. (2004). The Schizosaccharomyces pombe HIRA-like protein Hip1 is required

for the periodic expression of histone genes and contributes to the function of complex centromeres. *Mol Cell Biol* 24, 4309-4320.

Blasco, M.A. (2007). The epigenetic regulation of mammalian telomeres. *Nat Rev Genet* 8, 299-309.

Blethrow, J.D., Tang, C., Deng, C., and Krutchinsky, A.N. (2007). Modular mass spectrometric tool for analysis of composition and phosphorylation of protein complexes. *PLoS One* 2, e358.

Bloom, G., Yang, I.V., Boulware, D., Kwong, K.Y., Coppola, D., Eschrich, S., Quackenbush, J., and Yeatman, T.J. (2004). Multi-platform, multi-site, microarray-based human tumor classification. *Am J Pathol* 164, 9-16.

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

Bonnefoy, E., Orsi, G.A., Couble, P., and Loppin, B. (2007). The essential role of *Drosophila* HIRA for de novo assembly of paternal chromatin at fertilization. *PLoS Genet* 3, 1991-2006.

Borrelli, E., Nestler, E.J., Allis, C.D., and Sassone-Corsi, P. (2008). Decoding the epigenetic language of neuronal plasticity. *Neuron* 60, 961-974.

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.

Bramlage, B., Kosciessa, U., and Doenecke, D. (1997). Differential expression of the murine histone genes H3.3A and H3.3B. *Differentiation* 62, 13-20.

Briggs, S.D., Bryk, M., Strahl, B.D., Cheung, W.L., Davie, J.K., Dent, S.Y., Winston, F., and Allis, C.D. (2001). Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in *Saccharomyces cerevisiae*. *Genes Dev* 15, 3286-3295.

Brodsky, A.S., Meyer, C.A., Swinburne, I.A., Hall, G., Keenan, B.J., Liu, X.S., Fox, E.A., and Silver, P.A. (2005). Genomic mapping of RNA polymerase II reveals sites of co-transcriptional regulation in human cells. *Genome Biol* 6, R64.

Brownell, J.E., and Allis, C.D. (1995). An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in *Tetrahymena* macronuclei. *Proc Natl Acad Sci U S A* 92, 6364-6368.

Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y.,

and Allis, C.D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* 84, 843-851.

Burdon, T., Smith, A., and Savatier, P. (2002). Signalling, cell cycle and pluripotency in embryonic stem cells. *Trends Cell Biol* 12, 432-438.

Buschbeck, M., Uribesalgo, I., Wibowo, I., Rue, P., Martin, D., Gutierrez, A., Morey, L., Guigo, R., Lopez-Schier, H., and Di Croce, L. (2009). The histone variant macroH2A is an epigenetic regulator of key developmental genes. *Nat Struct Mol Biol*.

Bustin, M., and Cole, R.D. (1968). Species and organ specificity in very lysine-rich histones. *J Biol Chem* 243, 4500-4505.

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

Campbell, K.H., McWhir, J., Ritchie, W.A., and Wilmut, I. (1996). Sheep cloned by nuclear transfer from a cultured cell line. *Nature* 380, 64-66.

Carmo-Fonseca, M. (2002). The contribution of nuclear compartmentalization to gene regulation. *Cell* 108, 513-521.

Carozzi, N., Marashi, F., Plumb, M., Zimmerman, S., Zimmerman, A., Coles, L.S., Wells, J.R., Stein, G., and Stein, J. (1984). Clustering of human H1 and core histone genes. *Science* 224, 1115-1117.

Carroll, D. (2008). Progress and prospects: zinc-finger nucleases as gene therapy agents. *Gene Ther* 15, 1463-1468.

Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P., *et al.* (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* 123, 581-592.

Celeste, A., Petersen, S., Romanienko, P.J., Fernandez-Capetillo, O., Chen, H.T., Sedelnikova, O.A., Reina-San-Martin, B., Coppola, V., Meffre, E., Difilippantonio, M.J., *et al.* (2002). Genomic instability in mice lacking histone H2AX. *Science* 296, 922-927.

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

Chang, C.C., Lin, D.Y., Fang, H.I., Chen, R.H., and Shih, H.M. (2005). Daxx mediates the small ubiquitin-like modifier-dependent transcriptional repression of

Smad4. *J Biol Chem* 280, 10164-10173.

Chasis, J.A., and Mohandas, N. (2008). Erythroblastic islands: niches for erythropoiesis. *Blood* 112, 470-478.

Chen, C.C., Carson, J.J., Feser, J., Tamburini, B., Zabaronick, S., Linger, J., and Tyler, J.K. (2008a). Acetylated lysine 56 on histone H3 drives chromatin assembly after repair and signals for the completion of repair. *Cell* 134, 231-243.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., *et al.* (2008b). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133, 1106-1117.

Chew, J.L., Loh, Y.H., Zhang, W., Chen, X., Tam, W.L., Yeap, L.S., Li, P., Ang, Y.S., Lim, B., Robson, P., *et al.* (2005). Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. *Mol Cell Biol* 25, 6031-6046.

Chi, A.S., and Bernstein, B.E. (2009). Developmental biology. Pluripotent chromatin state. *Science* 323, 220-221.

Chow, C.M., Georgiou, A., Szutorisz, H., Maia e Silva, A., Pombo, A., Barahona, I., Dargelos, E., Canzonetta, C., and Dillon, N. (2005). Variant histone H3.3 marks promoters of transcriptionally active genes during mammalian cell division. *EMBO Rep* 6, 354-360.

Chow, J., and Heard, E. (2009). X inactivation and the complexities of silencing a sex chromosome. *Curr Opin Cell Biol* 21, 359-366.

Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. *Annual review of biochemistry* 78, 273-304.

Cloonan, N., Forrest, A.R., Kolle, G., Gardiner, B.B., Faulkner, G.J., Brown, M.K., Taylor, D.F., Steptoe, A.L., Wani, S., Bethel, G., *et al.* (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods* 5, 613-619.

Conti, L., Pollard, S.M., Gorba, T., Reitano, E., Toselli, M., Biella, G., Sun, Y., Sanzone, S., Ying, Q.L., Cattaneo, E., *et al.* (2005). Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biol* 3, e283.

Couldrey, C., Carlton, M.B., Nolan, P.M., Colledge, W.H., and Evans, M.J. (1999). A retroviral gene trap insertion into the histone 3.3A gene causes partial neonatal lethality, stunted growth, neuromuscular deficits and male sub-fertility in transgenic mice. *Hum Mol Genet* 8, 2489-2495.

Creyghton, M.P., Markoulaki, S., Levine, S.S., Hanna, J., Lodato, M.A., Sha, K.,

Young, R.A., Jaenisch, R., and Boyer, L.A. (2008). H2AZ is enriched at polycomb complex target genes in ES cells and is necessary for lineage commitment. *Cell* 135, 649-661.

Cristea, I.M., Williams, R., Chait, B.T., and Rout, M.P. (2005). Fluorescent proteins as proteomic probes. *Mol Cell Proteomics* 4, 1933-1941.

Croxton, R., Puto, L.A., de Belle, I., Thomas, M., Torii, S., Hanai, F., Cuddy, M., and Reed, J.C. (2006). Daxx represses expression of a subset of antiapoptotic genes regulated by nuclear factor-kappaB. *Cancer Res* 66, 9026-9035.

Cui, B., Liu, Y., and Gorovsky, M.A. (2006). Deposition and function of histone H3 variants in *Tetrahymena thermophila*. *Mol Cell Biol* 26, 7719-7730.

Cuzin, F., Grandjean, V., and Rassoulzadegan, M. (2008). Inherited variation at the epigenetic level: paramutation from the plant to the mouse. *Curr Opin Genet Dev* 18, 193-196.

Daury, L., Chailleux, C., Bonvallet, J., and Trouche, D. (2006). Histone H3.3 deposition at E2F-regulated genes is linked to transcription. *EMBO Rep* 7, 66-71.

Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., and Richmond, T.J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J Mol Biol* 319, 1097-1113.

de Jonge, H.J., Fehrmann, R.S., de Bont, E.S., Hofstra, R.M., Gerbens, F., Kamps, W.A., de Vries, E.G., van der Zee, A.G., te Meerman, G.J., and ter Elst, A. (2007). Evidence based selection of housekeeping genes. *PLoS One* 2, e898.

De Koning, L., Corpet, A., Haber, J.E., and Almouzni, G. (2007). Histone chaperones: an escort network regulating histone traffic. *Nat Struct Mol Biol* 14, 997-1007.

Dehe, P.M., and Geli, V. (2006). The multiple faces of Set1. *Biochem Cell Biol* 84, 536-548.

Deng, C., and Capecchi, M.R. (1992). Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Mol Cell Biol* 12, 3365-3371.

Dhalluin, C., Carlson, J.E., Zeng, L., He, C., Aggarwal, A.K., and Zhou, M.M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. *Nature* 399, 491-496.

Dobosy, J.R., and Selker, E.U. (2001). Emerging connections between DNA methylation and histone acetylation. *Cell Mol Life Sci* 58, 721-727.

Doyon, Y., McCammon, J.M., Miller, J.C., Faraji, F., Ngo, C., Katibah, G.E., Amora, R., Hocking, T.D., Zhang, L., Rebar, E.J., *et al.* (2008). Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat Biotechnol* 26, 702-708.

Drabent, B., Bode, C., Bramlage, B., and Doenecke, D. (1996). Expression of the mouse testicular histone gene H1t during spermatogenesis. *Histochem Cell Biol* 106, 247-251.

Duncan, E.M., Muratore-Schroeder, T.L., Cook, R.G., Garcia, B.A., Shabanowitz, J., Hunt, D.F., and Allis, C.D. (2008). Cathepsin L proteolytically processes histone H3 during mouse embryonic stem cell differentiation. *Cell* 135, 284-294.

Durrin, L.K., Mann, R.K., Kayne, P.S., and Grunstein, M. (1991). Yeast histone H4 N-terminal sequence is required for promoter activation in vivo. *Cell* 65, 1023-1031.

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

Eggan, K., Akutsu, H., Loring, J., Jackson-Grusby, L., Klemm, M., Rideout, W.M., 3rd, Yanagimachi, R., and Jaenisch, R. (2001). Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation. *Proc Natl Acad Sci U S A* 98, 6209-6214.

Elliott, B., Richardson, C., Winderbaum, J., Nickoloff, J.A., and Jasin, M. (1998). Gene conversion tracts from double-strand break repair in mammalian cells. *Mol Cell Biol* 18, 93-101.

Engelke, D.R., Ng, S.Y., Shastry, B.S., and Roeder, R.G. (1980). Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. *Cell* 19, 717-728.

English, C.M., Adkins, M.W., Carson, J.J., Churchill, M.E., and Tyler, J.K. (2006). Structural basis for the histone chaperone activity of Asf1. *Cell* 127, 495-508.

Ezhkova, E., Pasolli, H.A., Parker, J.S., Stokes, N., Su, I.H., 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.

Fagiolini, M., Jensen, C.L., and Champagne, F.A. (2009). Epigenetic influences on brain development and plasticity. *Curr Opin Neurobiol* 19, 207-212.

Fan, Y., Nikitina, T., Morin-Kensicki, E.M., Zhao, J., Magnuson, T.R., Woodcock, C.L., and Skoultschi, A.I. (2003). H1 linker histones are essential for mouse

development and affect nucleosome spacing in vivo. *Mol Cell Biol* 23, 4559-4572.

Fan, Y., Nikitina, T., Zhao, J., Fleury, T.J., Bhattacharyya, R., Bouhassira, E.E., Stein, A., Woodcock, C.L., and Skoultschi, A.I. (2005). Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell* 123, 1199-1212.

Farnham, P.J. (2009). Insights from genomic profiling of transcription factors. *Nat Rev Genet* 10, 605-616.

Feil, R. (2009). Epigenetic asymmetry in the zygote and mammalian development. *Int J Dev Biol* 53, 191-201.

Felsenfeld, G., and Groudine, M. (2003). Controlling the double helix. *Nature* 421, 448-453.

Fernandez-Capetillo, O., Lee, A., Nussenzweig, M., and Nussenzweig, A. (2004). H2AX: the histone guardian of the genome. *DNA Repair (Amst)* 3, 959-967.

Field, Y., Fondufe-Mittendorf, Y., Moore, I.K., Mieczkowski, P., Kaplan, N., Lubling, Y., Lieb, J.D., Widom, J., and Segal, E. (2009). Gene expression divergence in yeast is coupled to evolution of DNA-encoded nucleosome organization. *Nat Genet* 41, 438-445.

Fillingham, J., and Greenblatt, J.F. (2008). A histone code for chromatin assembly. *Cell* 134, 206-208.

Fischle, W., Tseng, B.S., Dormann, H.L., Ueberheide, B.M., Garcia, B.A., Shabanowitz, J., Hunt, D.F., Funabiki, H., and Allis, C.D. (2005). Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature* 438, 1116-1122.

Flemming, W. (1882). *Zellsubstanz, Kern und Zelltheilung ... Mit ... Tafeln* (pp. viii. 424. Leipzig).

Frank, D., Doenecke, D., and Albig, W. (2003). Differential expression of human replacement and cell cycle dependent H3 histone genes. *Gene* 312, 135-143.

Franklin, S.G., and Zweidler, A. (1977). Non-allelic variants of histones 2a, 2b and 3 in mammals. *Nature* 266, 273-275.

Freitag, M., and Selker, E.U. (2005). Controlling DNA methylation: many roads to one modification. *Current opinion in genetics & development* 15, 191-199.

Gaillard, P.H., Martini, E.M., Kaufman, P.D., Stillman, B., Moustacchi, E., and Almouzni, G. (1996). Chromatin assembly coupled to DNA repair: a new role for

chromatin assembly factor I. *Cell* 86, 887-896.

Gaillard, P.H., Roche, D., and Almouzni, G. (1999). Nucleotide excision repair coupled to chromatin assembly. *Methods Mol Biol* 119, 231-243.

Galvani, A., Courbeyrette, R., Agez, M., Ochsenbein, F., Mann, C., and Thuret, J.Y. (2008). In vivo study of the nucleosome assembly functions of ASF1 histone chaperones in human cells. *Mol Cell Biol* 28, 3672-3685.

Garrick, D., Sharpe, J.A., Arkell, R., Dobbie, L., Smith, A.J., Wood, W.G., Higgs, D.R., and Gibbons, R.J. (2006). Loss of Atrx affects trophoblast development and the pattern of X-inactivation in extraembryonic tissues. *PLoS Genet* 2, e58.

Gaspar-Maia, A., Alajem, A., Polesso, F., Sridharan, R., Mason, M.J., Heidersbach, A., Ramalho-Santos, J., McManus, M.T., Plath, K., Meshorer, E., *et al.* (2009). Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature* 460, 863-868.

Gatewood, J.M., Cook, G.R., Balhorn, R., Schmid, C.W., and Bradbury, E.M. (1990). Isolation of four core histones from human sperm chromatin representing a minor subset of somatic histones. *J Biol Chem* 265, 20662-20666.

Geiser, M., Cebe, R., Drewello, D., and Schmitz, R. (2001). Integration of PCR fragments at any specific site within cloning vectors without the use of restriction enzymes and DNA ligase. *Biotechniques* 31, 88-90, 92.

Gibbons, R. (2006). Alpha thalassaemia-mental retardation, X linked. *Orphanet J Rare Dis* 1, 15.

Gibbons, R.J., McDowell, T.L., Raman, S., O'Rourke, D.M., Garrick, D., Ayyub, H., and Higgs, D.R. (2000). Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. *Nat Genet* 24, 368-371.

Gibbons, R.J., Wada, T., Fisher, C.A., Malik, N., Mitson, M.J., Steensma, D.P., Fryer, A., Goudie, D.R., Krantz, I.D., and Traeger-Synodinos, J. (2008). Mutations in the chromatin-associated protein ATRX. *Hum Mutat* 29, 796-802.

Girton, J.R., and Johansen, K.M. (2008). Chromatin structure and the regulation of gene expression: the lessons of PEV in *Drosophila*. *Adv Genet* 61, 1-43.

Goldberg, A.D., Allis, C.D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. *Cell* 128, 635-638.

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. *Annual review of biochemistry* 74, 481-514.

Govin, J., Caron, C., Lestrat, C., Rousseaux, S., and Khochbin, S. (2004). The role of histones in chromatin remodelling during mammalian spermiogenesis. *Eur J Biochem* 271, 3459-3469.

Green, E.M., Antczak, A.J., Bailey, A.O., Franco, A.A., Wu, K.J., Yates, J.R., 3rd, and Kaufman, P.D. (2005). Replication-independent histone deposition by the HIR complex and Asf1. *Curr Biol* 15, 2044-2049.

Grewal, S.I., and Jia, S. (2007). Heterochromatin revisited. *Nature reviews* 8, 35-46.

Groth, A., Corpet, A., Cook, A.J., Roche, D., Bartek, J., Lukas, J., and Almouzni, G. (2007a). Regulation of replication fork progression through histone supply and demand. *Science* 318, 1928-1931.

Groth, A., Rocha, W., Verreault, A., and Almouzni, G. (2007b). Chromatin challenges during DNA replication and repair. *Cell* 128, 721-733.

Grummt, I. (2003). Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. *Genes Dev* 17, 1691-1702.

Guenther, M.G., Jenner, R.G., Chevalier, B., Nakamura, T., Croce, C.M., Canaani, E., and Young, R.A. (2005). Global and Hox-specific roles for the MLL1 methyltransferase. *Proc Natl Acad Sci U S A* 102, 8603-8608.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* 130, 77-88.

Gurdon, J.B., and Byrne, J.A. (2003). The first half-century of nuclear transplantation. *Proc Natl Acad Sci U S A* 100, 8048-8052.

Gurdon, J.B., and Melton, D.A. (2008). Nuclear reprogramming in cells. *Science* 322, 1811-1815.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., *et al.* (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458, 223-227.

Hadjantonakis, A.K., and Papaioannou, V.E. (2004). Dynamic in vivo imaging and cell tracking using a histone fluorescent protein fusion in mice. *BMC Biotechnol* 4, 33.

Hajkova, P., Ancelin, K., Waldmann, T., Lacoste, N., Lange, U.C., Cesari, F., Lee, C., Almouzni, G., Schneider, R., and Surani, M.A. (2008). Chromatin dynamics

during epigenetic reprogramming in the mouse germ line. *Nature* **452**, 877-881.

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., and Surani, M.A. (2002). Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* **117**, 15-23.

Hake, S.B., and Allis, C.D. (2006). Histone H3 variants and their potential role in indexing mammalian genomes: The "H3 barcode hypothesis". *Proc Natl Acad Sci U S A*.

Hake, S.B., Garcia, B.A., Duncan, E.M., Kauer, M., Dellaire, G., Shabanowitz, J., Bazett-Jones, D.P., Allis, C.D., and Hunt, D.F. (2006). Expression patterns and post-translational modifications associated with mammalian histone H3 variants. *J Biol Chem* **281**, 559-568.

Hake, S.B., Garcia, B.A., Kauer, M., Baker, S.P., Shabanowitz, J., Hunt, D.F., and Allis, C.D. (2005). Serine 31 phosphorylation of histone variant H3.3 is specific to regions bordering centromeres in metaphase chromosomes. *Proc Natl Acad Sci U S A* **102**, 6344-6349.

Hammoud, S.S., Nix, D.A., Zhang, H., Purwar, J., Carrell, D.T., and Cairns, B.R. (2009). Distinctive chromatin in human sperm packages genes for embryo development. *Nature* **460**, 473-478.

Handel, M.A. (2004). The XY body: a specialized meiotic chromatin domain. *Exp Cell Res* **296**, 57-63.

Happel, N., and Doenecke, D. (2009). Histone H1 and its isoforms: contribution to chromatin structure and function. *Gene* **431**, 1-12.

Harley, C.B., Futcher, A.B., and Greider, C.W. (1990). Telomeres shorten during ageing of human fibroblasts. *Nature* **345**, 458-460.

Hasty, P., Rivera-Perez, J., and Bradley, A. (1991). The length of homology required for gene targeting in embryonic stem cells. *Mol Cell Biol* **11**, 5586-5591.

Hayashi, K., and Surani, M.A. (2009). Resetting the epigenome beyond pluripotency in the germline. *Cell Stem Cell* **4**, 493-498.

Heard, E. (2005). Delving into the diversity of facultative heterochromatin: the epigenetics of the inactive X chromosome. *Current opinion in genetics & development* **15**, 482-489.

Hebbes, T.R., Thorne, A.W., and Crane-Robinson, C. (1988). A direct link between core histone acetylation and transcriptionally active chromatin. *Embo J* **7**, 1395-1402.

Heintz, N., Zernik, M., and Roeder, R.G. (1981). The structure of the human histone genes: clustered but not tandemly repeated. *Cell* 24, 661-668.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., *et al.* (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459, 108-112.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., *et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 39, 311-318.

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.

Henikoff, S. (2008). Nucleosome destabilization in the epigenetic regulation of gene expression. *Nat Rev Genet* 9, 15-26.

Henikoff, S., and Ahmad, K. (2005). Assembly of variant histones into chromatin. *Annu Rev Cell Dev Biol* 21, 133-153.

Henikoff, S., Henikoff, J.G., Sakai, A., Loeb, G.B., and Ahmad, K. (2009). Genome-wide profiling of salt fractions maps physical properties of chromatin. *Genome Res* 19, 460-469.

Higgs, D.R., Garrick, D., Anguita, E., De Gobbi, M., Hughes, J., Muers, M., Vernimmen, D., Lower, K., Law, M., Argentaro, A., *et al.* (2005). Understanding alpha-globin gene regulation: Aiming to improve the management of thalassemia. *Ann N Y Acad Sci* 1054, 92-102.

Ho, L., and Crabtree, G.R. (2008). An EZ mark to miss. *Cell Stem Cell* 3, 577-578.

Ho, L., Jothi, R., Ronan, J.L., Cui, K., Zhao, K., and Crabtree, G.R. (2009a). An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proc Natl Acad Sci U S A*.

Ho, L., Ronan, J.L., Wu, J., Staahl, B.T., Chen, L., Kuo, A., Lessard, J., Nesvizhskii, A.I., Ranish, J., and Crabtree, G.R. (2009b). An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proc Natl Acad Sci U S A* 106, 5181-5186.

Hodl, M., and Basler, K. (2009). Transcription in the Absence of Histone H3.3. *Curr*

Biol.

Hogan, B. (2007). A shared vision. *Dev Cell* 13, 769-771.

Hollenbach, A.D., McPherson, C.J., Mientjes, E.J., Iyengar, R., and Grosveld, G. (2002). Daxx and histone deacetylase II associate with chromatin through an interaction with core histones and the chromatin-associated protein Dek. *J Cell Sci* 115, 3319-3330.

Hong, L., Schroth, G.P., Matthews, H.R., Yau, P., and Bradbury, E.M. (1993). Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA. *J Biol Chem* 268, 305-314.

Howman, E.V., Fowler, K.J., Newson, A.J., Redward, S., MacDonald, A.C., Kalitsis, P., and Choo, K.H. (2000). Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. *Proc Natl Acad Sci U S A* 97, 1148-1153.

Huang, Y.S., and Shih, H.M. (2009). Daxx positively modulates beta-catenin/TCF4-mediated transcriptional potential. *Biochem Biophys Res Commun* 386, 762-768.

Iguchi, N., Tanaka, H., Yomogida, K., and Nishimune, Y. (2003). Isolation and characterization of a novel cDNA encoding a DNA-binding protein (Hils1) specifically expressed in testicular haploid germ cells. *Int J Androl* 26, 354-365.

Ishov, A.M., Vladimirova, O.V., and Maul, G.G. (2004). Heterochromatin and ND10 are cell-cycle regulated and phosphorylation-dependent alternate nuclear sites of the transcription repressor Daxx and SWI/SNF protein ATRX. *J Cell Sci* 117, 3807-3820.

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

Jaenisch, R., and Young, R. (2008). Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* 132, 567-582.

Janicki, S.M., Tsukamoto, T., Salghetti, S.E., Tansey, W.P., Sachidanandam, R., Prasanth, K.V., Ried, T., Shav-Tal, Y., Bertrand, E., Singer, R.H., *et al.* (2004). From silencing to gene expression: real-time analysis in single cells. *Cell* 116, 683-698.

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

Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet* 10, 161-172.

Jin, C., and Felsenfeld, G. (2006). Distribution of histone H3.3 in hematopoietic cell lineages. *Proc Natl Acad Sci U S A* *103*, 574-579.

Jin, C., and Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes Dev* *21*, 1519-1529.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nat Genet*.

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

Joshi, A.A., and Struhl, K. (2005). Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Mol Cell* *20*, 971-978.

Kacem, S., and Feil, R. (2009). Chromatin mechanisms in genomic imprinting. *Mamm Genome*.

Kalitsis, P., Fowler, K.J., Earle, E., Griffiths, B., Howman, E., Newson, A.J., and Choo, K.H. (2003). Partially functional Cenpa-GFP fusion protein causes increased chromosome missegregation and apoptosis during mouse embryogenesis. *Chromosome Res* *11*, 345-357.

Kanoh, J., and Russell, P. (2000). Slm9, a novel nuclear protein involved in mitotic control in fission yeast. *Genetics* *155*, 623-631.

Keogh, M.C., Kim, J.A., Downey, M., Fillingham, J., Chowdhury, D., Harrison, J.C., Onishi, M., Datta, N., Galicia, S., Emili, A., *et al.* (2006). A phosphatase complex that dephosphorylates gammaH2AX regulates DNA damage checkpoint recovery. *Nature* *439*, 497-501.

Keogh, M.C., Kurdistani, S.K., Morris, S.A., Ahn, S.H., Podolny, V., Collins, S.R., Schuldiner, M., Chin, K., Punna, T., Thompson, N.J., *et al.* (2005). Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* *123*, 593-605.

Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A., *et al.* (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A* *106*, 11667-11672.

Kim, T., and Buratowski, S. (2009). Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions. *Cell* *137*, 259-272.

Kinkade, J.M., Jr., and Cole, R.D. (1966). A structural comparison of different lysine-rich histones of calf thymus. *J Biol Chem* **241**, 5798-5805.

Klose, R.J., and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* **31**, 89-97.

Klug, A. (2005). Towards therapeutic applications of engineered zinc finger proteins. *FEBS Lett* **579**, 892-894.

Koch, F., Jourquin, F., Ferrier, P., and Andrau, J.C. (2008). Genome-wide RNA polymerase II: not genes only! *Trends Biochem Sci* **33**, 265-273.

Konev, A.Y., Tribus, M., Park, S.Y., Podhraski, V., Lim, C.Y., Emelyanov, A.V., Vershilova, E., Pirrotta, V., Kadonaga, J.T., Lusser, A., *et al.* (2007). CHD1 motor protein is required for deposition of histone variant H3.3 into chromatin in vivo. *Science* **317**, 1087-1090.

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

Kornberg, R.D. (2005). Mediator and the mechanism of transcriptional activation. *Trends Biochem Sci* **30**, 235-239.

Kornberg, R.D., and Thomas, J.O. (1974). Chromatin structure; oligomers of the histones. *Science* **184**, 865-868.

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* **128**, 693-705.

Krimer, D.B., Cheng, G., and Skoultchi, A.I. (1993). Induction of H3.3 replacement histone mRNAs during the precommitment period of murine erythroleukemia cell differentiation. *Nucleic Acids Res* **21**, 2873-2879.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., *et al.* (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* **4**, e1000242.

Kuo, M.H., and Allis, C.D. (1999). In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. *Methods* **19**, 425-433.

Kurdistani, S.K., and Grunstein, M. (2003). Histone acetylation and deacetylation in yeast. *Nat Rev Mol Cell Biol* **4**, 276-284.

Kurimoto, K., Yabuta, Y., Ohinata, Y., Shigeta, M., Yamanaka, K., and Saitou, M.

(2008). Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes Dev* 22, 1617-1635.

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.

Lamour, V., Lecluse, Y., Desmaze, C., Spector, M., Bodescot, M., Aurias, A., Osley, M.A., and Lipinski, M. (1995). A human homolog of the *S. cerevisiae* HIR1 and HIR2 transcriptional repressors cloned from the DiGeorge syndrome critical region. *Hum Mol Genet* 4, 791-799.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860-921.

Laskey, R.A., and Gurdon, J.B. (1970). Genetic content of adult somatic cells tested by nuclear transplantation from cultured cells. *Nature* 228, 1332-1334.

Law, M. (2006). Targets and targeting of ATRX. In *Medical Sciences Division* (University of Oxford), pp. 196.

Lee, D.Y., Hayes, J.J., Pruss, D., and Wolffe, A.P. (1993). A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell* 72, 73-84.

Lee, J., Inoue, K., Ono, R., Ogonuki, N., Kohda, T., Kaneko-Ishino, T., Ogura, A., and Ishino, F. (2002a). Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. *Development* 129, 1807-1817.

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.* (2006a). Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 125, 301-313.

Lee, T.I., Johnstone, S.E., and Young, R.A. (2006b). Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat Protoc* 1, 729-748.

Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., *et al.* (2002b). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298, 799-804.

Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., *et al.* (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445, 168-176.

Lejnine, S., Makarov, V.L., and Langmore, J.P. (1995). Conserved nucleoprotein structure at the ends of vertebrate and invertebrate chromosomes. *Proc Natl Acad Sci U S A* 92, 2393-2397.

Levasseur, D.N., Wang, J., Dorschner, M.O., Stamatoyannopoulos, J.A., and Orkin, S.H. (2008). Oct4 dependence of chromatin structure within the extended Nanog locus in ES cells. *Genes Dev* 22, 575-580.

Levenson, J.M., and Sweatt, J.D. (2005). Epigenetic mechanisms in memory formation. *Nat Rev Neurosci* 6, 108-118.

Levy, M.A., Fernandes, A.D., Tremblay, D.C., Seah, C., and Berube, N.G. (2008). The SWI/SNF protein ATRX co-regulates pseudoautosomal genes that have translocated to autosomes in the mouse genome. *BMC Genomics* 9, 468.

Levy, M.Z., Allsopp, R.C., Futcher, A.B., Greider, C.W., and Harley, C.B. (1992). Telomere end-replication problem and cell aging. *J Mol Biol* 225, 951-960.

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

Li, B., Gogol, M., Carey, M., Lee, D., Seidel, C., and Workman, J.L. (2007b). Combined action of PHD and chromo domains directs the Rpd3S HDAC to transcribed chromatin. *Science* 316, 1050-1054.

Li, B., Gogol, M., Carey, M., Pattenden, S.G., Seidel, C., and Workman, J.L. (2007c). Infrequently transcribed long genes depend on the Set2/Rpd3S pathway for accurate transcription. *Genes Dev* 21, 1422-1430.

Li, Q., Zhou, H., Wurtele, H., Davies, B., Horazdovsky, B., Verreault, A., and Zhang, Z. (2008). Acetylation of histone H3 lysine 56 regulates replication-coupled nucleosome assembly. *Cell* 134, 244-255.

Lian, Z., Karpikov, A., Lian, J., Mahajan, M.C., Hartman, S., Gerstein, M., Snyder, M., and Weissman, S.M. (2008). A genomic analysis of RNA polymerase II modification and chromatin architecture related to 3' end RNA polyadenylation. *Genome Res* 18, 1224-1237.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., *et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289-293.

Litt, M., Qiu, Y., and Huang, S. (2009). Histone arginine methylations: their roles in chromatin dynamics and transcriptional regulation. *Biosci Rep* 29, 131-141.

Lopez-Fernandez, L.A., Lopez-Alanon, D.M., Castaneda, V., Krimer, D.B., and del Mazo, J. (1997). Developmental expression of H3.3A variant histone mRNA in mouse. *Int J Dev Biol* 41, 699-703.

Loppin, B., Bonnefoy, E., Anselme, C., Laurencon, A., Karr, T.L., and Couble, P. (2005). The histone H3.3 chaperone HIRA is essential for chromatin assembly in the male pronucleus. *Nature* 437, 1386-1390.

Loppin, B., Docquier, M., Bonneton, F., and Couble, P. (2000). The maternal effect mutation sesame affects the formation of the male pronucleus in *Drosophila melanogaster*. *Dev Biol* 222, 392-404.

Lorain, S., Demczuk, S., Lamour, V., Toth, S., Aurias, A., Roe, B.A., and Lipinski, M. (1996). Structural Organization of the WD repeat protein-encoding gene HIRA in the DiGeorge syndrome critical region of human chromosome 22. *Genome Res* 6, 43-50.

Lorain, S., Quivy, J.P., Monier-Gavelle, F., Scamps, C., Lecluse, Y., Almouzni, G., and Lipinski, M. (1998). Core histones and HIRIP3, a novel histone-binding protein, directly interact with WD repeat protein HIRA. *Mol Cell Biol* 18, 5546-5556.

Lord, K.A., Abdollahi, A., Hoffman-Liebermann, B., and Liebermann, D.A. (1990). Dissection of the immediate early response of myeloid leukemia cells to terminal differentiation and growth inhibitory stimuli. *Cell Growth Differ* 1, 637-645.

Loyola, A., and Almouzni, G. (2007). Marking histone H3 variants: how, when and why? *Trends Biochem Sci* 32, 425-433.

Loyola, A., Bonaldi, T., Roche, D., Imhof, A., and Almouzni, G. (2006). PTMs on H3 variants before chromatin assembly potentiate their final epigenetic state. *Mol Cell* 24, 309-316.

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

Luke, B., and Lingner, J. (2009). TERRA: telomeric repeat-containing RNA. *EMBO J*.

Luke, B., Panza, A., Redon, S., Iglesias, N., Li, Z., and Lingner, J. (2008). The Rat1p 5' to 3' exonuclease degrades telomeric repeat-containing RNA and promotes telomere elongation in *Saccharomyces cerevisiae*. *Mol Cell* 32, 465-477.

Luo, Y., Li, T., Yu, F., Kramer, T., and Cristea, I.M. (2009). Resolving the Composition of Protein Complexes using a MALDI LTQ Orbitrap. *Journal of the American Society for Mass Spectrometry* *In press*.

- Luse, D.S., Haynes, J.R., VanLeeuwen, D., Schon, E.A., Cleary, M.L., Shapiro, S.G., Lingrel, J.B., and Roeder, R.G. (1981). Transcription of the beta-like globin genes and pseudogenes of the goat in a cell-free system. *Nucleic Acids Res* 9, 4339-4354.
- Lusser, A., and Kadonaga, J.T. (2003). Chromatin remodeling by ATP-dependent molecular machines. *Bioessays* 25, 1192-1200.
- Magnaghi, P., Roberts, C., Lorain, S., Lipinski, M., and Scambler, P.J. (1998). HIRA, a mammalian homologue of *Saccharomyces cerevisiae* transcriptional co-repressors, interacts with Pax3. *Nat Genet* 20, 74-77.
- Makarov, V.L., Lejnine, S., Bedoyan, J., and Langmore, J.P. (1993). Nucleosomal organization of telomere-specific chromatin in rat. *Cell* 73, 775-787.
- Malik, H.S., and Henikoff, S. (2003). Phylogenomics of the nucleosome. *Nat Struct Biol* 10, 882-891.
- Malik, S., and Roeder, R.G. (2005). Dynamic regulation of pol II transcription by the mammalian Mediator complex. *Trends Biochem Sci* 30, 256-263.
- Marion, R.M., Strati, K., Li, H., Tejera, A., Schoeftner, S., Ortega, S., Serrano, M., and Blasco, M.A. (2009). Telomeres acquire embryonic stem cell characteristics in induced pluripotent stem cells. *Cell Stem Cell* 4, 141-154.
- Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., *et al.* (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* 134, 521-533.
- Martens, J.H., 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.
- Marzluff, W.F., Gongidi, P., Woods, K.R., Jin, J., and Maltais, L.J. (2002). The human and mouse replication-dependent histone genes. *Genomics* 80, 487-498.
- Matsui, T., Segall, J., Weil, P.A., and Roeder, R.G. (1980). Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *J Biol Chem* 255, 11992-11996.
- Matthews, A.G., Kuo, A.J., Ramon-Maiques, S., Han, S., Champagne, K.S., Ivanov, D., Gallardo, M., Carney, D., Cheung, P., Ciccone, D.N., *et al.* (2007). RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature* 450, 1106-1110.

McDowell, T.L., Gibbons, R.J., Sutherland, H., O'Rourke, D.M., Bickmore, W.A., Pombo, A., Turley, H., Gatter, K., Picketts, D.J., Buckle, V.J., *et al.* (1999). Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes. *Proc Natl Acad Sci U S A* **96**, 13983-13988.

McKittrick, E., Gafken, P.R., Ahmad, K., and Henikoff, S. (2004). Histone H3.3 is enriched in covalent modifications associated with active chromatin. *Proc Natl Acad Sci U S A* **101**, 1525-1530.

McStay, B., and Grummt, I. (2008). The epigenetics of rRNA genes: from molecular to chromosome biology. *Annu Rev Cell Dev Biol* **24**, 131-157.

Meissner, A., Eminli, S., and Jaenisch, R. (2009). Derivation and manipulation of murine embryonic stem cells. *Methods Mol Biol* **482**, 3-19.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., *et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766-770.

Meisterernst, M., Roy, A.L., Lieu, H.M., and Roeder, R.G. (1991). Activation of class II gene transcription by regulatory factors is potentiated by a novel activity. *Cell* **66**, 981-993.

Mendenhall, E.M., and Bernstein, B.E. (2008). Chromatin state maps: new technologies, new insights. *Curr Opin Genet Dev* **18**, 109-115.

Meneghini, M.D., Wu, M., and Madhani, H.D. (2003). Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. *Cell* **112**, 725-736.

Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev Cell* **10**, 105-116.

Metzler-Guillemain, C., Luciani, J., Depetris, D., Guichaoua, M.R., and Mattei, M.G. (2003). HP1beta and HP1gamma, but not HP1alpha, decorate the entire XY body during human male meiosis. *Chromosome Res* **11**, 73-81.

Meyne, J., Ratliff, R.L., and Moyzis, R.K. (1989). Conservation of the human telomere sequence (TTAGGG)_n among vertebrates. *Proc Natl Acad Sci U S A* **86**, 7049-7053.

Michaelson, J.S., Bader, D., Kuo, F., Kozak, C., and Leder, P. (1999). Loss of Daxx, a promiscuously interacting protein, results in extensive apoptosis in early mouse development. *Genes Dev* **13**, 1918-1923.

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.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., *et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448, 553-560.

Miller, J.C., Holmes, M.C., Wang, J., Guschin, D.Y., Lee, Y.L., Rupniewski, I., Beausejour, C.M., Waite, A.J., Wang, N.S., Kim, K.A., *et al.* (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol* 25, 778-785.

Minden, J.S., Dowd, S.R., Meyer, H.E., and Stuhler, K. (2009). Difference gel electrophoresis. *Electrophoresis* 30 Suppl 1, S156-161.

Minucci, S., and Pelicci, P.G. (2006). Histone deacetylase inhibitors and the promise of epigenetic (and more) treatments for cancer. *Nat Rev Cancer* 6, 38-51.

Mito, Y., Henikoff, J.G., and Henikoff, S. (2005). Genome-scale profiling of histone H3.3 replacement patterns. *Nat Genet* 37, 1090-1097.

Mito, Y., Henikoff, J.G., and Henikoff, S. (2007). Histone replacement marks the boundaries of cis-regulatory domains. *Science* 315, 1408-1411.

Mizuta, H., and Kuroda, Y. (2004). Cloning and functional characterization of a rat Daxx that functions as a corepressor for the androgen receptor. *Cell Biol Int* 28, 609-614.

Monod, J., and Jacob, F. (1961). Teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harb Symp Quant Biol* 26, 389-401.

Montgomery, N.D., Yee, D., Chen, A., Kalantry, S., Chamberlain, S.J., Otte, A.P., and Magnuson, T. (2005). The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation. *Curr Biol* 15, 942-947.

Moore, M., Klug, A., and Choo, Y. (2001). Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units. *Proc Natl Acad Sci U S A* 98, 1437-1441.

Morgan, H.D., Santos, F., Green, K., Dean, W., and Reik, W. (2005). Epigenetic reprogramming in mammals. *Hum Mol Genet* 14 Spec No 1, R47-58.

Moriniere, J., Rousseaux, S., Steuerwald, U., Soler-Lopez, M., Curtet, S., Vitte, A.L., Govin, J., Gaucher, J., Sadoul, K., Hart, D.J., *et al.* (2009). Cooperative

binding of two acetylation marks on a histone tail by a single bromodomain. *Nature* **461**, 664-668.

Murata, T., Kurokawa, R., Krones, A., Tatsumi, K., Ishii, M., Taki, T., Masuno, M., Ohashi, H., Yanagisawa, M., Rosenfeld, M.G., *et al.* (2001). Defect of histone acetyltransferase activity of the nuclear transcriptional coactivator CBP in Rubinstein-Taybi syndrome. *Hum Mol Genet* **10**, 1071-1076.

Muromoto, R., Nakao, K., Watanabe, T., Sato, N., Sekine, Y., Sugiyama, K., Oritani, K., Shimoda, K., and Matsuda, T. (2006). Physical and functional interactions between Daxx and STAT3. *Oncogene* **25**, 2131-2136.

Nakayama, T., Nishioka, K., Dong, Y.X., Shimojima, T., and Hirose, S. (2007). *Drosophila* GAGA factor directs histone H3.3 replacement that prevents the heterochromatin spreading. *Genes Dev* **21**, 552-561.

Natsume, R., Eitoku, M., Akai, Y., Sano, N., Horikoshi, M., and Senda, T. (2007). Structure and function of the histone chaperone CIA/ASF1 complexed with histones H3 and H4. *Nature* **446**, 338-341.

Ng, H.H., and Bird, A. (1999). DNA methylation and chromatin modification. *Curr Opin Genet Dev* **9**, 158-163.

Ng, R.K., and Gurdon, J.B. (2005). Epigenetic memory of active gene transcription is inherited through somatic cell nuclear transfer. *Proc Natl Acad Sci U S A* **102**, 1957-1962.

Ng, R.K., and Gurdon, J.B. (2008). Epigenetic memory of an active gene state depends on histone H3.3 incorporation into chromatin in the absence of transcription. *Nat Cell Biol* **10**, 102-109.

Niwa, H. (2007). How is pluripotency determined and maintained? *Development* **134**, 635-646.

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet* **24**, 372-376.

Nottke, A., Colaiacovo, M.P., and Shi, Y. (2009). Developmental roles of the histone lysine demethylases. *Development* **136**, 879-889.

Nourani, A., Robert, F., and Winston, F. (2006). Evidence that Spt2/Sin1, an HMG-like factor, plays roles in transcription elongation, chromatin structure, and genome stability in *Saccharomyces cerevisiae*. *Mol Cell Biol* **26**, 1496-1509.

O'Carroll, D., Erhardt, S., Pagani, M., Barton, S.C., Surani, M.A., and Jenuwein, T.

(2001). The polycomb-group gene *Ezh2* is required for early mouse development. *Mol Cell Biol* 21, 4330-4336.

O'Neill, L.P., and Turner, B.M. (2003). Immunoprecipitation of native chromatin: NChIP. *Methods* 31, 76-82.

Odom, D.T., Dowell, R.D., Jacobsen, E.S., Nekludova, L., Rolfe, P.A., Danford, T.W., Gifford, D.K., Fraenkel, E., Bell, G.I., and Young, R.A. (2006). Core transcriptional regulatory circuitry in human hepatocytes. *Mol Syst Biol* 2, 2006 0017.

Ohinata, Y., Payer, B., O'Carroll, D., Ancelin, K., Ono, Y., Sano, M., Barton, S.C., Obukhanych, T., Nussenzweig, M., Tarakhovsky, A., *et al.* (2005). Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature* 436, 207-213.

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

Olins, D.E., and Olins, A.L. (2003). Chromatin history: our view from the bridge. *Nat Rev Mol Cell Biol* 4, 809-814.

Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D., *et al.* (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* 448, 714-717.

Ooi, S.L., Priess, J.R., and Henikoff, S. (2006). Histone H3.3 variant dynamics in the germline of *Caenorhabditis elegans*. *PLoS Genet* 2, e97.

Orsi, G.A., Couble, P., and Loppin, B. (2009). Epigenetic and replacement roles of histone variant H3.3 in reproduction and development. *Int J Dev Biol* 53, 231-243.

Ottaviani, A., Gilson, E., and Magdinier, F. (2008). Telomeric position effect: from the yeast paradigm to human pathologies? *Biochimie* 90, 93-107.

Oudet, P., Gross-Bellard, M., and Chambon, P. (1975). Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell* 4, 281-300.

Palm, W., and de Lange, T. (2008). How shelterin protects mammalian telomeres. *Annu Rev Genet* 42, 301-334.

Perez-Burgos, L., Peters, A.H., Opravil, S., Kauer, M., Mechtler, K., and Jenuwein, T. (2004). Generation and characterization of methyl-lysine histone antibodies. *Methods Enzymol* 376, 234-254.

Phillips, D.M. (1963). The presence of acetyl groups of histones. *Biochem J* 87,

258-263.

Picketts, D.J., Higgs, D.R., Bachoo, S., Blake, D.J., Quarrell, O.W., and Gibbons, R.J. (1996). ATRX encodes a novel member of the SNF2 family of proteins: mutations point to a common mechanism underlying the ATR-X syndrome. *Hum Mol Genet* 5, 1899-1907.

Pina, B., and Suau, P. (1987). Changes in histones H2A and H3 variant composition in differentiating and mature rat brain cortical neurons. *Dev Biol* 123, 51-58.

Pisano, S., Marchioni, E., Galati, A., Mechelli, R., Savino, M., and Cacchione, S. (2007). Telomeric nucleosomes are intrinsically mobile. *J Mol Biol* 369, 1153-1162.

Plumb, M., Stein, J., and Stein, G. (1983). Influence of DNA synthesis inhibition on the coordinate expression of core human histone genes during S phase. *Nucleic Acids Res* 11, 7927-7945.

Pogo, B.G., Allfrey, V.G., and Mirsky, A.E. (1966). RNA synthesis and histone acetylation during the course of gene activation in lymphocytes. *Proc Natl Acad Sci U S A* 55, 805-812.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolzheimer, E., *et al.* (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* 122, 517-527.

Polo, S.E., and Almouzni, G. (2006). Chromatin assembly: a basic recipe with various flavours. *Curr Opin Genet Dev* 16, 104-111.

Polo, S.E., Roche, D., and Almouzni, G. (2006). New histone incorporation marks sites of UV repair in human cells. *Cell* 127, 481-493.

Prochasson, P., Florens, L., Swanson, S.K., Washburn, M.P., and Workman, J.L. (2005). The HIR corepressor complex binds to nucleosomes generating a distinct protein/DNA complex resistant to remodeling by SWI/SNF. *Genes Dev* 19, 2534-2539.

Pusarla, R.H., and Bhargava, P. (2005). Histones in functional diversification. Core histone variants. *FEBS J* 272, 5149-5168.

Puto, L.A., and Reed, J.C. (2008). Daxx represses RelB target promoters via DNA methyltransferase recruitment and DNA hypermethylation. *Genes Dev* 22, 998-1010.

Quail, M.A., Kozarewa, I., Smith, F., Scally, A., Stephens, P.J., Durbin, R., Swerdlow, H., and Turner, D.J. (2008). A large genome center's improvements to the Illumina

sequencing system. *Nat Methods* 5, 1005-1010.

Raisner, R.M., Hartley, P.D., Meneghini, M.D., Bao, M.Z., Liu, C.L., Schreiber, S.L., Rando, O.J., and Madhani, H.D. (2005). Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell* 123, 233-248.

Ramirez-Carrozzi, V.R., Braas, D., Bhatt, D.M., Cheng, C.S., Hong, C., Doty, K.R., Black, J.C., Hoffmann, A., Carey, M., and Smale, S.T. (2009). A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* 138, 114-128.

Rando, O.J., and Chang, H.Y. (2009). Genome-Wide Views of Chromatin Structure. *Annual review of biochemistry*.

Ray-Gallet, D., Quivy, J.P., Scamps, C., Martini, E.M., Lipinski, M., and Almouzni, G. (2002). HIRA is critical for a nucleosome assembly pathway independent of DNA synthesis. *Mol Cell* 9, 1091-1100.

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.

Recht, J., Tsubota, T., Tanny, J.C., Diaz, R.L., Berger, J.M., Zhang, X., Garcia, B.A., Shabanowitz, J., Burlingame, A.L., Hunt, D.F., *et al.* (2006). Histone chaperone Asf1 is required for histone H3 lysine 56 acetylation, a modification associated with S phase in mitosis and meiosis. *Proc Natl Acad Sci U S A* 103, 6988-6993.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., *et al.* (2000). Genome-wide location and function of DNA binding proteins. *Science* 290, 2306-2309.

Ringrose, L., and Paro, R. (2004). Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annual review of genetics* 38, 413-443.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., *et al.* (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311-1323.

Roberts, C., Sutherland, H.F., Farmer, H., Kimber, W., Halford, S., Carey, A., Brickman, J.M., Wynshaw-Boris, A., and Scambler, P.J. (2002). Targeted mutagenesis of the Hira gene results in gastrulation defects and patterning abnormalities of mesoendodermal derivatives prior to early embryonic lethality. *Mol Cell Biol* 22, 2318-2328.

Robertson, A.G., Bilenky, M., Tam, A., Zhao, Y., Zeng, T., Thiessen, N., Cezard, T.,

Fejes, A.P., Wederell, E.D., Cullum, R., *et al.* (2008). Genome-wide relationship between histone H3 lysine 4 mono- and tri-methylation and transcription factor binding. *Genome Res* 18, 1906-1917.

Roeder, R.G. (2003). Lasker Basic Medical Research Award. The eukaryotic transcriptional machinery: complexities and mechanisms unforeseen. *Nat Med* 9, 1239-1244.

Roeder, R.G., and Rutter, W.J. (1969). Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature* 224, 234-237.

Roeder, R.G., and Rutter, W.J. (1970). Specific nucleolar and nucleoplasmic RNA polymerases. *Proc Natl Acad Sci U S A* 65, 675-682.

Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., and Bonner, W.M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem* 273, 5858-5868.

Rogers, R.S., Inselman, A., Handel, M.A., and Matunis, M.J. (2004). SUMO modified proteins localize to the XY body of pachytene spermatocytes. *Chromosoma* 113, 233-243.

Rosenfeld, N., Aharonov, R., Meiri, E., Rosenwald, S., Spector, Y., Zepeniuk, M., Benjamin, H., Shabes, N., Tabak, S., Levy, A., *et al.* (2008). MicroRNAs accurately identify cancer tissue origin. *Nat Biotechnol* 26, 462-469.

Ross, M.H., Pawlina, W., and Barnash, T.A. (2009). Atlas of descriptive histology (Sunderland, Mass., Sinauer Associates).

Rossant, J. (2009). Reprogramming to pluripotency: from frogs to stem cells. *Cell* 138, 1047-1050.

Ruthenburg, A.J., Allis, C.D., and Wysocka, J. (2007a). Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol Cell* 25, 15-30.

Ruthenburg, A.J., Li, H., Patel, D.J., and Allis, C.D. (2007b). Multivalent engagement of chromatin modifications by linked binding modules. *Nat Rev Mol Cell Biol* 8, 983-994.

Ruthenburg, A.J., Wang, W., Graybosch, D.M., Li, H., Allis, C.D., Patel, D.J., and Verdine, G.L. (2006). Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex. *Nat Struct Mol Biol* 13, 704-712.

Saga, Y. (2008). Mouse germ cell development during embryogenesis. *Curr Opin Genet Dev* 18, 337-341.

Sakai, A., Schwartz, B.E., Goldstein, S., and Ahmad, K. (2009). Transcriptional and Developmental Functions of the H3.3 Histone Variant in *Drosophila*. *Curr Biol*.

Salomoni, P., and Khelifi, A.F. (2006). Daxx: death or survival protein? *Trends Cell Biol* **16**, 97-104.

Sanematsu, F., Takami, Y., Barman, H.K., Fukagawa, T., Ono, T., Shibahara, K., and Nakayama, T. (2006). Asf1 is required for viability and chromatin assembly during DNA replication in vertebrate cells. *J Biol Chem* **281**, 13817-13827.

Santenard, A., and Torres-Padilla, M.E. (2009). Epigenetic reprogramming in mammalian reproduction: Contribution from histone variants. *Epigenetics* **4**.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are trimethylated at K4 of histone H3. *Nature* **419**, 407-411.

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

Scaturro, M., Cestelli, A., Castiglia, D., Nastasi, T., and Di Liegro, I. (1995). Posttranscriptional regulation of H1 zero and H3.3B histone genes in differentiating rat cortical neurons. *Neurochem Res* **20**, 969-976.

Schneiderman, J.I., Sakai, A., Goldstein, S., and Ahmad, K. (2009). The XNP remodeler targets dynamic chromatin in *Drosophila*. *Proc Natl Acad Sci U S A* **106**, 14472-14477.

Schoeftner, S., and Blasco, M.A. (2008). Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat Cell Biol* **10**, 228-236.

Schoeftner, S., and Blasco, M.A. (2009). A 'higher order' of telomere regulation: telomere heterochromatin and telomeric RNAs. *EMBO J*.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887-898.

Schotta, G., Ebert, A., Dorn, R., and Reuter, G. (2003). Position-effect variegation and the genetic dissection of chromatin regulation in *Drosophila*. *Semin Cell Dev Biol* **14**, 67-75.

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

745.

Schwartz, B.E., and Ahmad, K. (2005). Transcriptional activation triggers deposition and removal of the histone variant H3.3. *Genes Dev* 19, 804-814.

Schwartz, B.E., and Ahmad, K. (2006). 2. Chromatin assembly with H3 histones: full throttle down multiple pathways. *Curr Top Dev Biol* 74, 31-55.

Sealy, L., and Chalkley, R. (1978). DNA associated with hyperacetylated histone is preferentially digested by DNase I. *Nucleic Acids Res* 5, 1863-1876.

Seandel, M., James, D., Shmelkov, S.V., Falciatori, I., Kim, J., Chavala, S., Scherr, D.S., Zhang, F., Torres, R., Gale, N.W., *et al.* (2007). Generation of functional multipotent adult stem cells from GPR125+ germline progenitors. *Nature* 449, 346-350.

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thastrom, A., Field, Y., Moore, I.K., Wang, J.P., and Widom, J. (2006). A genomic code for nucleosome positioning. *Nature* 442, 772-778.

Segal, E., and Widom, J. (2009). What controls nucleosome positions? *Trends Genet* 25, 335-343.

Segall, J., Matsui, T., and Roeder, R.G. (1980). Multiple factors are required for the accurate transcription of purified genes by RNA polymerase III. *J Biol Chem* 255, 11986-11991.

Sfeir, A., Kosiyatrakul, S.T., Hockemeyer, D., MacRae, S.L., Karlseder, J., Schildkraut, C.L., and de Lange, T. (2009). Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell* 138, 90-103.

Shahbazian, M.D., and Grunstein, M. (2007). Functions of site-specific histone acetylation and deacetylation. *Annual review of biochemistry* 76, 75-100.

Shen, X., Liu, Y., Hsu, Y.J., Fujiwara, Y., Kim, J., Mao, X., Yuan, G.C., and Orkin, S.H. (2008). EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol Cell* 32, 491-502.

Shi, X., Hong, T., Walter, K.L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Pena, P., Lan, F., Kaadige, M.R., *et al.* (2006). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* 442, 96-99.

Shibahara, K., and Stillman, B. (1999). Replication-dependent marking of DNA by PCNA facilitates CAF-1-coupled inheritance of chromatin. *Cell* 96, 575-585.

Shilatifard, A. (2008). Molecular implementation and physiological roles for histone

H3 lysine 4 (H3K4) methylation. *Curr Opin Cell Biol* 20, 341-348.

Shukla, V.K., Doyon, Y., Miller, J.C., DeKolver, R.C., Moehle, E.A., Worden, S.E., Mitchell, J.C., Arnold, N.L., Gopalan, S., Meng, X., *et al.* (2009). Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature* 459, 437-441.

Siegel, T.N., Hekstra, D.R., Kemp, L.E., Figueiredo, L.M., Lowell, J.E., Fenyo, D., Wang, X., Dewell, S., and Cross, G.A. (2009). Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. *Genes Dev.*

Sikorski, T.W., and Buratowski, S. (2009). The basal initiation machinery: beyond the general transcription factors. *Curr Opin Cell Biol* 21, 344-351.

Sims, R.J., 3rd, Chen, C.F., Santos-Rosa, H., Kouzarides, T., Patel, S.S., and Reinberg, D. (2005). Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *J Biol Chem* 280, 41789-41792.

Sims, R.J., 3rd, and Reinberg, D. (2009). Stem cells: Escaping fates with open states. *Nature* 460, 802-803.

Sklar, V.E., Schwartz, L.B., and Roeder, R.G. (1975). Distinct molecular structures of nuclear class I, II, and III DNA-dependent RNA polymerases. *Proc Natl Acad Sci U S A* 72, 348-352.

Smith, C.L., and Peterson, C.L. (2005). ATP-dependent chromatin remodeling. *Curr Top Dev Biol* 65, 115-148.

Smith, S., and Stillman, B. (1989). Purification and characterization of CAF-I, a human cell factor required for chromatin assembly during DNA replication in vitro. *Cell* 58, 15-25.

Sobel, R.E., Cook, R.G., and Allis, C.D. (1994). Non-random acetylation of histone H4 by a cytoplasmic histone acetyltransferase as determined by novel methodology. *J Biol Chem* 269, 18576-18582.

Sobel, R.E., Cook, R.G., Perry, C.A., Annunziato, A.T., and Allis, C.D. (1995). Conservation of deposition-related acetylation sites in newly synthesized histones H3 and H4. *Proc Natl Acad Sci U S A* 92, 1237-1241.

Spector, M.S., Raff, A., DeSilva, H., Lee, K., and Osley, M.A. (1997). Hir1p and Hir2p function as transcriptional corepressors to regulate histone gene transcription in the *Saccharomyces cerevisiae* cell cycle. *Mol Cell Biol* 17, 545-552.

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.

Straathof, K.C., Pule, M.A., Yotnda, P., Dotti, G., Vanin, E.F., Brenner, M.K., Heslop, H.E., Spencer, D.M., and Rooney, C.M. (2005). An inducible caspase 9 safety switch for T-cell therapy. *Blood* 105, 4247-4254.

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

Struhl, K. (1998). Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev* 12, 599-606.

Sutcliffe, E.L., Parish, I.A., He, Y.Q., Juelich, T., Tierney, M.L., Rangasamy, D., Milburn, P.J., Parish, C.R., Tremethick, D.J., and Rao, S. (2009). Dynamic histone variant exchange accompanies gene induction in T cells. *Mol Cell Biol* 29, 1972-1986.

Tagami, H., Ray-Gallet, D., Almouzni, G., and Nakatani, Y. (2004). Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell* 116, 51-61.

Tahiliani, M., Mei, P., Fang, R., Leonor, T., Rutenberg, M., Shimizu, F., Li, J., Rao, A., and Shi, Y. (2007). The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation. *Nature* 447, 601-605.

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.

Talbert, P.B., and Henikoff, S. (2009). Chromatin-based transcriptional punctuation. *Genes Dev* 23, 1037-1041.

Tamura, T., Smith, M., Kanno, T., Dasenbrock, H., Nishiyama, A., and Ozato, K. (2009). Inducible deposition of the histone variant H3.3 in interferon-stimulated genes. *J Biol Chem*.

Tanaka, M., Hennebold, J.D., Macfarlane, J., and Adashi, E.Y. (2001). A mammalian oocyte-specific linker histone gene H1oo: homology with the genes for the oocyte-specific cleavage stage histone (cs-H1) of sea urchin and the B4/H1M histone of the frog. *Development* 128, 655-664.

Tanaka, T.S. (2009). Transcriptional heterogeneity in mouse embryonic stem cells. *Reprod Fertil Dev* 21, 67-75.

Tang, J., Wu, S., Liu, H., Stratt, R., Barak, O.G., Shiekhata, R., Picketts, D.J., and Yang, X. (2004). A novel transcription regulatory complex containing death domain-associated protein and the ATR-X syndrome protein. *J Biol Chem* 279, 20369-20377.

Tang, Y., Poustovoitov, M.V., Zhao, K., Garfinkel, M., Canutescu, A., Dunbrack, R., Adams, P.D., and Marmorstein, R. (2006). Structure of a human ASF1a-HIRA complex and insights into specificity of histone chaperone complex assembly. *Nat Struct Mol Biol* 13, 921-929.

Taunton, J., Hassig, C.A., and Schreiber, S.L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* 272, 408-411.

te Riele, H., Maandag, E.R., and Berns, A. (1992). Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs. *Proc Natl Acad Sci U S A* 89, 5128-5132.

Teng, G., and Papavasiliou, F.N. (2007). Immunoglobulin somatic hypermutation. *Annu Rev Genet* 41, 107-120.

Terranova, R., Agherbi, H., Boned, A., Meresse, S., and Djabali, M. (2006). Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll. *Proc Natl Acad Sci U S A* 103, 6629-6634.

Thatcher, T.H., MacGaffey, J., Bowen, J., Horowitz, S., Shapiro, D.L., and Gorovsky, M.A. (1994). Independent evolutionary origin of histone H3.3-like variants of animals and Tetrahymena. *Nucleic Acids Res* 22, 180-186.

Tommerup, H., Dousmanis, A., and de Lange, T. (1994). Unusual chromatin in human telomeres. *Mol Cell Biol* 14, 5777-5785.

Torres-Padilla, M.E., Bannister, A.J., Hurd, P.J., Kouzarides, T., and Zernicka-Goetz, M. (2006). Dynamic distribution of the replacement histone variant H3.3 in the mouse oocyte and preimplantation embryos. *Int J Dev Biol* 50, 455-461.

Tumbar, T., Guasch, G., Greco, V., Blanpain, C., Lowry, W.E., Rendl, M., and Fuchs, E. (2004). Defining the epithelial stem cell niche in skin. *Science* 303, 359-363.

Tupler, R., Perini, G., and Green, M.R. (2001). Expressing the human genome. *Nature* 409, 832-833.

Turner, B.M. (2000). Histone acetylation and an epigenetic code. *Bioessays* 22, 836-845.

Urnov, F.D., Miller, J.C., Lee, Y.L., Beausejour, C.M., Rock, J.M., Augustus, S.,

Jamieson, A.C., Porteus, M.H., Gregory, P.D., and Holmes, M.C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* 435, 646-651.

van der Heijden, G.W., Derijck, A.A., Posfai, E., Giele, M., Pelczar, P., Ramos, L., Wansink, D.G., van der Vlag, J., Peters, A.H., and de Boer, P. (2007). Chromosome-wide nucleosome replacement and H3.3 incorporation during mammalian meiotic sex chromosome inactivation. *Nat Genet* 39, 251-258.

van der Heijden, G.W., Dieker, J.W., Derijck, A.A., Muller, S., Berden, J.H., Braat, D.D., van der Vlag, J., and de Boer, P. (2005). Asymmetry in histone H3 variants and lysine methylation between paternal and maternal chromatin of the early mouse zygote. *Mech Dev* 122, 1008-1022.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* 10, 252-263.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., *et al.* (2001). The sequence of the human genome. *Science* 291, 1304-1351.

Venters, B.J., and Pugh, B.F. (2009). How eukaryotic genes are transcribed. *Crit Rev Biochem Mol Biol* 44, 117-141.

Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W., van Schaik, F.M., Varier, R.A., Baltissen, M.P., Stunnenberg, H.G., Mann, M., and Timmers, H.T. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* 131, 58-69.

Verreault, A., Kaufman, P.D., Kobayashi, R., and Stillman, B. (1996). Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. *Cell* 87, 95-104.

Verreault, A., Kaufman, P.D., Kobayashi, R., and Stillman, B. (1998). Nucleosomal DNA regulates the core-histone-binding subunit of the human Hat1 acetyltransferase. *Curr Biol* 8, 96-108.

Vickaryous, M.K., and Hall, B.K. (2006). Human cell type diversity, evolution, development, and classification with special reference to cells derived from the neural crest. *Biol Rev Camb Philos Soc* 81, 425-455.

Vidali, G., Boffa, L.C., Bradbury, E.M., and Allfrey, V.G. (1978). Butyrate suppression of histone deacetylation leads to accumulation of multiacetylated forms of histones H3 and H4 and increased DNase I sensitivity of the associated DNA sequences. *Proc Natl Acad Sci U S A* 75, 2239-2243.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., *et al.* (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457, 854-858.

Waddington, C.H. (1942). The epigenotype. *Endeavour* 1, 18–20.

Waddington, C.H. (1957). The strategy of the genes; a discussion of some aspects of theoretical biology (London,, Allen & Unwin).

Wang, A., Kurdistani, S.K., and Grunstein, M. (2002). Requirement of Hos2 histone deacetylase for gene activity in yeast. *Science* 298, 1412-1414.

Wang, Z., Gerstein, M., and Snyder, M. (2009a). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10, 57-63.

Wang, Z., Zang, C., Cui, K., Schones, D.E., Barski, A., Peng, W., and Zhao, K. (2009b). Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* 138, 1019-1031.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q., *et al.* (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* 40, 897-903.

Waterborg, J.H., and Robertson, A.J. (1996). Common features of analogous replacement histone H3 genes in animals and plants. *J Mol Evol* 43, 194-206.

Watson, J.D. (2003). Celebrating the genetic jubilee: a conversation with James D. Watson. Interviewed by John Rennie. *Scientific American* 288, 66-69.

Weake, V.M., and Workman, J.L. (2008). Histone ubiquitination: triggering gene activity. *Mol Cell* 29, 653-663.

Weaver, J.R., Susiarjo, M., and Bartolomei, M.S. (2009). Imprinting and epigenetic changes in the early embryo. *Mamm Genome*.

Wederell, E.D., Bilenky, M., Cullum, R., Thiessen, N., Dagpinar, M., Delaney, A., Varhol, R., Zhao, Y., Zeng, T., Bernier, B., *et al.* (2008). Global analysis of in vivo Foxa2-binding sites in mouse adult liver using massively parallel sequencing. *Nucleic Acids Res* 36, 4549-4564.

Weinmann, R., Raskas, H.J., and Roeder, R.G. (1974). Role of DNA-dependent RNA polymerases II and III in transcription of the adenovirus genome late in productive infection. *Proc Natl Acad Sci U S A* 71, 3426-3439.

Weinmann, R., and Roeder, R.G. (1974). Role of DNA-dependent RNA polymerase

3 in the transcription of the tRNA and 5S RNA genes. *Proc Natl Acad Sci U S A* **71**, 1790-1794.

Weintraub, H., and Groudine, M. (1976). Chromosomal subunits in active genes have an altered conformation. *Science* **193**, 848-856.

Weintraub, H., Tapscott, S.J., Davis, R.L., Thayer, M.J., Adam, M.A., Lassar, A.B., and Miller, A.D. (1989). Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. *Proc Natl Acad Sci U S A* **86**, 5434-5438.

Wells, D., Hoffman, D., and Kedes, L. (1987). Unusual structure, evolutionary conservation of non-coding sequences and numerous pseudogenes characterize the human H3.3 histone multigene family. *Nucleic Acids Res* **15**, 2871-2889.

Wethkamp, N., and Klempnauer, K.H. (2009). Daxx Is a Transcriptional Repressor of CCAAT/Enhancer-binding Protein {beta}. *J Biol Chem* **284**, 28783-28794.

Whitcomb, S.J., Basu, A., Allis, C.D., and Bernstein, E. (2007). Polycomb Group proteins: an evolutionary perspective. *Trends Genet* **23**, 494-502.

Williams, S.K., Truong, D., and Tyler, J.K. (2008). Acetylation in the globular core of histone H3 on lysine-56 promotes chromatin disassembly during transcriptional activation. *Proc Natl Acad Sci U S A* **105**, 9000-9005.

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.

Wilusz, J.E., Sunwoo, H., and Spector, D.L. (2009). Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev* **23**, 1494-1504.

Wirbelauer, C., Bell, O., and Schubeler, D. (2005). Variant histone H3.3 is deposited at sites of nucleosomal displacement throughout transcribed genes while active histone modifications show a promoter-proximal bias. *Genes Dev* **19**, 1761-1766.

Wolffe, A. (1992). *Chromatin Structure and Function* (San Diego, Academic Press).

Wong, L.H., Ren, H., Williams, E., McGhie, J., Ahn, S., Sim, M., Tam, A., Earle, E., Anderson, M.A., Mann, J., *et al.* (2009). Histone H3.3 incorporation provides a unique and functionally essential telomeric chromatin in embryonic stem cells. *Genome Res* **19**, 404-414.

Wu, R.S., Tsai, S., and Bonner, W.M. (1982). Patterns of histone variant synthesis can distinguish G0 from G1 cells. *Cell* **31**, 367-374.

Wunsch, A.M., and Lough, J. (1987). Modulation of histone H3 variant synthesis during the myoblast-myotube transition of chicken myogenesis. *Dev Biol* 119, 94-99.

Wysocka, J., Swigut, T., Milne, T.A., Dou, Y., Zhang, X., Burlingame, A.L., Roeder, R.G., Brivanlou, A.H., and Allis, C.D. (2005). WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* 121, 859-872.

Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P., *et al.* (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* 442, 86-90.

Xie, W., Song, C., Young, N.L., Sperling, A.S., Xu, F., Sridharan, R., Conway, A.E., Garcia, B.A., Plath, K., Clark, A.T., *et al.* (2009). Histone h3 lysine 56 acetylation is linked to the core transcriptional network in human embryonic stem cells. *Mol Cell* 33, 417-427.

Xue, Y., Gibbons, R., Yan, Z., Yang, D., McDowell, T.L., Sechi, S., Qin, J., Zhou, S., Higgs, D., and Wang, W. (2003). The ATRX syndrome protein forms a chromatin-remodeling complex with Daxx and localizes in promyelocytic leukemia nuclear bodies. *Proc Natl Acad Sci U S A* 100, 10635-10640.

Yamanaka, S. (2009). A fresh look at iPS cells. *Cell* 137, 13-17.

Yan, W., Ma, L., Burns, K.H., and Matzuk, M.M. (2003). HILS1 is a spermatid-specific linker histone H1-like protein implicated in chromatin remodeling during mammalian spermiogenesis. *Proc Natl Acad Sci U S A* 100, 10546-10551.

Yang, P.K., and Kuroda, M.I. (2007). Noncoding RNAs and intranuclear positioning in monoallelic gene expression. *Cell* 128, 777-786.

Yang, X., Smith, S.L., Tian, X.C., Lewin, H.A., Renard, J.P., and Wakayama, T. (2007). Nuclear reprogramming of cloned embryos and its implications for therapeutic cloning. *Nat Genet* 39, 295-302.

Ye, X., Zerlanko, B., Zhang, R., Somaiah, N., Lipinski, M., Salomoni, P., and Adams, P.D. (2007). Definition of pRB- and p53-dependent and -independent steps in HIRA/ASF1a-mediated formation of senescence-associated heterochromatin foci. *Mol Cell Biol* 27, 2452-2465.

Yehezkel, S., Segev, Y., Viegas-Pequignot, E., Skorecki, K., and Selig, S. (2008). Hypomethylation of subtelomeric regions in ICF syndrome is associated with abnormally short telomeres and enhanced transcription from telomeric regions. *Hum Mol Genet* 17, 2776-2789.

Yuan, G.C., Liu, Y.J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., and Rando, O.J. (2005). Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* 309, 626-630.

Yuan, J., Pu, M., Zhang, Z., and Lou, Z. (2009). Histone H3-K56 acetylation is important for genomic stability in mammals. *Cell Cycle* 8, 1747-1753.

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

Zhang, R., Chen, W., and Adams, P.D. (2007). Molecular dissection of formation of senescence-associated heterochromatin foci. *Mol Cell Biol* 27, 2343-2358.

Zhang, R., Poustovoitov, M.V., Ye, X., Santos, H.A., Chen, W., Daganzo, S.M., Erzberger, J.P., Serebriiskii, I.G., Canutescu, A.A., Dunbrack, R.L., *et al.* (2005). Formation of MacroH2A-containing senescence-associated heterochromatin foci and senescence driven by ASF1a and HIRA. *Dev Cell* 8, 19-30.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., *et al.* (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9, R137.

Zhao, R., Bodnar, M.S., and Spector, D.L. (2009). Nuclear neighborhoods and gene expression. *Curr Opin Genet Dev* 19, 172-179.

Zheng, D., Zhao, K., and Mehler, M.F. (2009). Profiling RE1/REST-mediated histone modifications in the human genome. *Genome Biol* 10, R9.

Zofall, M., Fischer, T., Zhang, K., Zhou, M., Cui, B., Veenstra, T.D., and Grewal, S.I. (2009). Histone H2A.Z cooperates with RNAi and heterochromatin factors to suppress antisense RNAs. *Nature* 461, 419-422.