Title
Stage-specific Chromatin Modification and Regulation in Cardiomyocyte Differentiation

Author
Alexander, Jeffrey M.

Publication Date
2012

Peer reviewed|Thesis/dissertation

Stage-specific Chromatin Modification and Regulation in Cardiomyocyte Differentiation

by

Jeffrey M. Alexander

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Acknowledgements

Publications and Contributions

Data presented in Chapter 2 is in press at *Cell* entitled, “Dynamic and Coordinated Epigenetic Regulation of Developmental Transitions in the Cardiac Lineage” by Joseph A. Wamstad*, Jeffrey M. Alexander*, Rebecca M. Truty, Avanti Shrikumar, Fugen Li, Kirsten E. Eilertson, Huiming Ding, John N. Wylie, Alexander R. Pico, John A. Capra, Genevieve Erwin, Steven J. Kattman, Gordon M. Keller, Deepak Srivastava, Stuart S. Levine, Katherine S. Pollard, Alisha K. Holloway, Laurie A. Boyer, & Benoit G. Bruneau.

* - authors contributed equally to this work

The work described in Chapter 2 was achieved through close collaboration with Dr. Alisha K. Holloway, Dr. Rebecca M. Truty, Dr. Alexander R. Pico, and John N. Wylie. Data analysis presented in Figures 2.10, 2.13-27 was performed by remaining authors, especially Dr. Joseph A. Wamstad and Avanti Shrikumar. This work has been included here with the authors’ permission.

Personal Acknowledgements

There are many to whom I owe many thanks and much gratitude. These people have offered mentorship, counseling, and friendship at times when I needed all of the above. I would not have completed the challenge of graduate work without their collective support.

First and foremost, I would like to thank my graduate mentor, Dr. Benoit Bruneau.

In agreeing to allow me to pursue research in his laboratory, Benoit took on the task of teaching a young student with high aspirations and very little research experience how to study biology. This was a risky undertaking, but remarkably I never felt a lack of confidence from Benoit that I would get to where I needed to be, even at times when I was questioning myself. Through the years, I have benefited from Benoit's mentorship and scientific vision. At the same time, Benoit has given me freedom to explore and refine my own scientific ideas, redirecting me when I have strayed too far off course. I have learned so much during my time in the lab and know I will take these lessons with me as I move to further develop my scientific career.

I would also like to express my gratitude to my committee members, Dr. Deepak Srivastava, Dr. Robert Blelloch, and Dr. Miguel Ramalho-Santos. You have been an important part of my graduate training, both before and after joining Benoit's lab. I have enjoyed our scientific discussions and believe these discussions have helped refined how I think about biological problems and ask scientific questions. I am lucky to have had such smart and generous scientists to turn to for advice and guidance.

UCSF and the Gladstone Institute have been stimulating environments for me as a graduate student, and numerous investigators have helped me along the way. I would like to thank Dr. Frank McCormick and Dr. Feroz Papa. I learned much rotating through your respective laboratories and appreciate the opportunity that you gave me. I would also like to thank Dr. Jeremy Reiter, Dr. Barbara Panning, and Dr. Tikashi Mikawa, who volunteered their time to serve on my qualifying exam and gave me excellent advice on my developing project. Lastly, I would like to thank Dr. Katie Pollard, Dr. Bruce Conklin, and Dr. Bob Farese. It has been a pleasure to work with such great collaborators and neighbors and I have been the beneficiary of your guidance and advice, whether they

came from bumping into one another in the lobby or on the shuttle or from feedback at Gladstone RIPS. Thank you for making UCSF and the Gladstone a great place to be a student.

I would like to thank my undergraduate advisors and mentors, specifically Dr. Monica Driscoll and Dr. Beate Gerstbrein. I learned many valuable lessons about how to do science during my time in the Driscoll lab. Furthermore, Monica gave me critical advice and guidance during my search for a graduate school, and she recommended the graduate division at UCSF, which is how I came to pursue my Ph.D at this institution.

I am indebted to many members of the Bruneau lab, including those who have come and gone. I have learned so much from our day to day interactions and you have made the lab a fun and exciting environment to learn how to practice science. Specifically, I would like to acknowledge Dr. Kiyonori Togi and Dr. Nathalie Gaborit. Kiyonori was my hands-on tutor in the lab for quite some time and I could not have asked for a more patient and careful mentor. Working with Kiyonori also built a friendship that I value, and we enjoyed taking breaks from work to take in some Philadelphia Eagles games. Nathalie was a very influential person during my graduate student career. She taught me many things about science, including how to carefully design experiments and analyze data. More than that, Nathalie was a good friend and helped talk through tough times in the lab when I needed help finding the light at the end of the tunnel. I am lucky to have made such good friends.

I must also thank my current colleagues in the lab. I have never been lacking a place for good discussions, valuable advice, and sharp constructive criticism, and you are the reason for this. You all have made the lab a great place to work and inspired my

work through your example. Specifically, I would like to thank Dr. Paul Delgado, Dr. Josh Wythe, and Dr. Patrick Devine. You have been important tutors for me in the laboratory and I have learned about how to be a careful and successful scientist from the lessons that you have taught me directly or by example. You have been generous with your time and for that I am grateful.

During my time in Benoit's lab, I have had the pleasure of working with many collaborators. I would like to thank Dr. Gordon Keller, Dr. Steve Kattman, and the entire Keller lab. You were incredibly generous to allow a young graduate student to set up shop in your laboratory for weeks and made me feel right at home. This experience was an important turning point for me and I cannot thank you enough. Also, I would like to thank Dr. Rebecca Truty, Dr. Joseph Wamstad, Dr. Laurie Boyer, and Dr. Alisha Holloway. It is through your hardwork and determination that many of the insights we discovered in our collaborative project came to be. In addition, I have learned many things about bioinformatics, chromatin biology, and how to successfully collaborate through our interactions and discussions. Thank you for making our collaboration enriching and exciting.

I consider myself incredibly fortunate to have been a part of the Gladstone Institute. This is truly a great place to learn and do science, and I have met some incredible people during my time here. I need to thank members of the Gladstone Flow Cytometry, Stem Cell, Genomics, and Bioinformatics Cores. You have given me important assistance and advice during my time here that have been very useful. While it is impossible to mention everyone who has impacted my time here, please know that I remember and will be forever grateful for the advice, discussions, and laughs that we have had during our overlapping time here. I am glad to have met so many smart and

good people. You have helped broaden my vision, both scientifically and personally. Thank you all.

Graduate school is a tough time that takes a lot of support from friends and family. I am lucky to have met so many incredible people since moving to the Bay Area for graduate school. Classmates, roommates, coworkers, and others who I have met along the way have been an important part of my time in graduate school. Whether it was willing to offer a word of encouragement when needed or grabbing a beer and getting my mind off lab, you have been there.

Furthermore, I have many friends and family in New Jersey who have supported me through graduate school. I thank my parents, Chuck and Joan. You had always been there while I was growing up, and you have continued to support me even though the distance has perhaps made it more difficult. To my sister, Sara, and my brother, Dan, thank you for being there both past and present. It has been fun watching your lives change and you grow during these past years. I look forward to more great times ahead. I love you and thank you for always believing that I could do anything if I put my mind to it.

Lastly, I thank my loving girlfriend, Sarah. You have been so important to me over the past two and a half years. You have always believed in me and have been there to listen to me complain when I needed it but also been insistent that I should not feel sorry for myself. You have given me encouragement that has seen me through tough times. Importantly, you have helped me rediscover some of the joys of life that are found outside the lab. I am incredibly lucky to have you as a part of my life. Thanks for being the wonderful girl that you are.

To anyone that I have unfortunately left out, I apologize for not giving you the recognition you deserve. I will end with a final thank you. Thanks to all who have helped me live, laugh, think, and persevere on way to producing this dissertation. It is, indeed, not my accomplishment but our accomplishment.

Abstract

Development of a properly formed heart is vital to life and defects in cardiogenesis lead to congenital heart disease. Central to this process is the commitment and differentiation of cardiovascular cell types from pluripotent progenitors, which depends on activation of entire gene expression programs. Chromatin structure is essential for the modulation of gene expression, yet we know little about how chromatin is modified and regulated during cardiogenesis. We have described the changing chromatin landscape of cardiac differentiation and investigated how chromatin regulators, such as chromatin remodelers, impact gene expression during this process. Using an efficient directed differentiation of cardiomyocytes from embryonic stem cells, we identified four stages of cardiac differentiation and profiled genome-wide occupancy of histone modifications. We found multiple, distinct chromatin patterns and have demonstrated the relationship between these patterns and gene expression. In addition, we identified a novel pre-activation chromatin pattern found at many cardiac muscle genes. Using histone modification signature, we have identified numerous putative enhancer regions, which allowed for the discovery of novel transcriptional regulatory networks and the identification of transcriptional synergism between the transcription factors *Gata4* and *Meis1*. Furthermore, we studied the role of the chromatin remodeling factor, *Brg1*, in cardiac differentiation. We determined that *Brg1* is required for cardiac differentiation and early loss of *Brg1* led to the derepression of many Polycomb target genes. Further investigation revealed that *Brg1* is required for H3K27me3 levels at many derepressed genes, suggesting a potential cooperativity between *Brg1* and Polycomb repressive

complexes. Taken together, our studies provide an important framework for future study of chromatin and its regulatory factors in the developing heart.

Table of Contents

Acknowledgements iii
Abstract ix
Table of Contents xi
List of Tables xv
List of Figures xvi
Abbreviations xix

Chapter 1
Introduction: Applying Embryonic Stem Cells as a Model to Understand Chromatin States and Their Regulation during Cardiomyocyte Differentiation 1
1.1 Medical significance of studying the role of chromatin regulation in cardiac differentiation 1
1.2 Cellular basis of mammalian heart development 2
1.3 Transcriptional control of cardiac differentiation 5
1.4 Chromatin structure and regulation 7
1.4.1 Histone modifications 10
1.4.2 Polycomb repressive complexes 12
1.4.3 Chromatin remodeling 14
1.4.4 BAF complexes 15
1.5 Embryonic stem cells as a developmental model 17
1.6 Summary of our studies to investigate chromatin modification and regulation in cardiac differentiation 19
Chapter 2
Dynamic and Coordinated Epigenetic Regulation of Developmental Transitions in the Cardiac Lineage 22
2.1 Contributions 22
2.2 Abstract 22
2.3 Introduction 23
2.3 Results 25
2.4.1 Characterization of directed cardiomyocyte differentiation 25

2.4.2 Expression and chromatin states in cardiac differentiation 36
2.4.3 Chromatin state dynamics during cardiac differentiation 39
2.4.4 Dynamic chromatin states correlate with distinct expression patterns 40
2.4.5 A novel chromatin state transition during CM differentiation 45
2.4.6 Enhancer activity correlates with cardiac specific programs 50
2.4.7 Enhancer transitions during CM differentiation 64
2.4.8 Integrating enhancers into gene networks 67
2.5 Discussion 78
2.5.1 Dynamic epigenetic transitions in differentiation 79
2.5.2 A novel dynamic pattern of histone modifications 79
2.5.3 Identification of transcriptional networks based on enhancer predictions 80
2.6 Materials and methods 81
2.6.1 Cardiomyocyte differentiation 81
2.6.2 Immunofluorescence microscopy 83
2.6.3 Flow cytometry 83
2.6.4 Quantitative PCR 84
2.6.5 RNA-seq 85
2.6.6 RNA-seq analysis pipeline 85
2.6.7 miRNA analysis 86
2.6.8 ChIP-seq 86
2.6.9 ChIP-seq analysis pipeline 88
2.6.10 Analysis of chromatin marks at TSSs 88
2.6.11 Evaluation of the correlation of chromatin and expression clusters 90
2.6.12 GO analysis of chromatin clusters 91
2.6.13 Chromatin expression co-clusters networks 91
2.6.14 Enhancer identification 91
2.6.15 Enhancer analysis 92
2.6.16 Conservation of cardiac enhancers 93
2.6.17 Dip finding algorithm 94
2.6.18 Identification of enriched motifs 95

2.6.19 Transcription factor expression analysis 96
2.6.20 Transcription factor and motif enrichment correlation 96
2.6.21 Gene network analysis 97
Chapter 3
Brg1 Is Required for Repression of Developmental Regulators by Polycomb Repressive Complexes in Mesoderm Differentiation 99
3.1 Contributions 99
3.2 Abstract 99
3.3 Introduction 100
3.4 Results 102
3.4.1 BAF complex subunits are downregulated with cardiac differentiation 102
3.4.2 Brg1 is required for cardiomyocyte differentiation 103
3.4.3 Loss of Brg1 causes dysregulation of developmental regulators 110
3.4.4 Brg1-dependent genes are Polycomb target genes 116
3.4.5 Brg1 is required for H3K27me3 levels at upregulated genes 119
3.5 Discussion 120
3.5.1 An expanded role for Brg1 in cardiac differentiation 120
3.5.2 Brg1 and Polycomb-mediated silencing 124
3.5.3 Brg1 classification as a Trithorax-group gene 125
3.6 Materials and methods 128
3.6.1 Cardiomyocyte differentiation 128
3.6.2 Immunofluorescence microscopy 129
3.6.3 Flow cytometry 130
3.6.4 Quantitative PCR 130
3.6.5 Western blotting 131
3.6.6 RNA-seq 131
3.6.7 RNA-seq analysis 132
3.6.8 Metagene analysis 132
3.6.9 Polycomb target analysis 133
3.6.10 ChIP-seq 133
3.6.11 ChIP-seq analysis pipeline 134

Chapter 4
Summary, Future Directions, and Perspective 136
4.1 A bird's-eye view of chromatin and transcription in cardiac differentiation 137
4.2 Temporal analysis of *Brg1* function reveals novel aspects of its role in cardiogenesis 139
4.3 Concluding remarks 142
References 144
Appendix: Supplementary Tables 162
Library Publishing Agreement 174

List of Tables

Supplementary Table 1. Enriched gene ontology terms for expression clusters..... 162
Supplementary Table 2. Enriched gene ontology terms for chromatin clusters..... 166
Supplementary Table 3. List of antibodies used in study..... 172

List of Figures

Figure 1.1. Multiple stages of mouse cardiac development. 3
Figure 1.2. Chromatin structure and histone modifications. 8
Figure 1.3. BAF and related complexes. 16
Figure 2.1. Timeline of cardiac differentiation protocol. 27
Figure 2.2. Characterization of cardiac markers during cardiac differentiation. 29
Figure 2.3. Differentiated cultures express cardiac muscle proteins. 30
Figure 2.4. Emergence of FLK-1- and PDGFRα-expressing cell population during mesoderm differentiation. 32
Figure 2.5. BMP4 concentration is critical for cardiomyocyte yield. 34
Figure 2.6. Gene expression analysis of directed cardiac differentiation. 35
Figure 2.7. Analysis of cardiomyocyte differentiations used for global gene expression and chromatin analysis. 37
Figure 2.8. Transcriptional analysis of cardiac differentiation. 38
Figure 2.9. Chromatin state transitions during cardiac differentiation. 41
Figure 2.10. Dynamic and highly correlated chromatin and gene expression patterns during cardiomyocyte differentiation. 43
Figure 2.11. Chromatin pattern distinguishes functional gene classes within genes specifically expressed at the cardiomyocyte stage. 46
Figure 2.12. H3K4me1 marks cardiac contractile genes prior to gene activation. 48
Figure 2.13. Identification of enhancer elements during cardiac differentiation. 51
Figure 2.14. Characterization of enhancers identified by chromatin signature during cardiac differentiation. 52
Figure 2.15. Example enhancer regions identified at genes expressed throughout cardiac development. 53
Figure 2.16. Comparison of cardiac enhancers with previous studies reveals significant overlap in predicted cardiac enhancers. 55
Figure 2.17. Cardiac enhancers display increased conservation within a limited phylogenetic distance. 58
Figure 2.18. Chromatin features and activity of enhancer elements during

cardiac differentiation. 59
Figure 2.19. H3K4me3, H3K27me3 and RNAPser5P overlap distinct subsets of enhancers identified during cardiomyocyte differentiation. 60
Figure 2.20. Active enhancer elements strongly associate with genes characteristic of each stage of differentiation. 63
Figure 2.21. Transitioning enhancer states during cardiac differentiation. 65
Figure 2.22. Active enhancer elements display rapid chromatin state transitions over the course of cardiac differentiation. 66
Figure 2.23. Identification of putative regulators of enhancer activity during cardiac differentiation. 68
Figure 2.24. Predicted transcription factor binding events significantly overlap with genome-wide TF occupancy. 71
Figure 2.25. Transcription Factor/Motif identified target gene networks. 72
Figure 2.26. Putative enhancer gene networks in heart development. 73
Figure 2.27. GATA and Meis factors are predicted to co-regulate many enhancers during cardiac differentiation. 75
Figure 2.28. MEIS1A and GATA4 synergistically activate multiple putative target enhancer elements in a luciferase reporter activation assay. 77
Figure 3.1. BAF complex subunits are downregulated with cardiac differentiation. 104
Figure 3.2. *Brg1* is downregulated during cardiomyocyte differentiation. 105
Figure 3.3. Treatment of *Brg1* fl/fl; Actin-CreER ES cells with 4-OHT leads to efficient deletion of *Brg1*. 106
Figure 3.4. *Brg1* is required for cardiomyocyte differentiation from ES cells. 108
Figure 3.5. Loss of *Brg1* leads to misregulation of a small subset of cardiac genes. 109
Figure 3.6. Loss of *Brg1* leads to an increase in cell death. 111
Figure 3.7. Cartoon representation of RNA-seq experimental design. 112
Figure 3.8. Loss of *Brg1* leads to significantly upregulated and downregulated gene expression. 113
Figure 3.9. Genes downregulated and upregulated by loss of *Brg1* show distinct regulation during mesoderm differentiation. 115
Figure 3.10. *Brg1* is required for the regulation of developmental genes during mesoderm induction. 117

Figure 3.11. *Brg1*-dependent genes are targets of Polycomb repressive complexes. 118
Figure 3.12. *Brg1* is required for H3K27me3 levels at genes upregulated by loss of *Brg1*. 121

Abbreviations

CHD – congenital heart disease

SHF – second heart field

ES – embryonic stem

TF – transcription factor

TSS – transcription start site

PRC – Polycomb repressive complex

ESC – embryonic stem cell

MES – mesoderm

CP – cardiac precursor

CM – cardiomyocyte

EB – embryoid body

cTnT – cardiac Troponin T

DP – double positive

BAF – Brg1/Brahma-associated factor

HDAC – histone deacetylase

iPS – induced pluripotent stem

RPKM – reads per kilobase exon per million reads

GO – gene ontology

Chapter 1

Introduction: Applying Embryonic Stem Cells as a Model to Understand Chromatin States and Their Regulation During Cardiomyocyte Differentiation

1.1 Medical significance of studying the role of chromatin regulation in cardiac differentiation

Abnormalities in embryonic development lead to congenital disease. Incidence of some form of congenital malformation is common, present in 2-5% of human births (Thorogood, 1997). Within these, perturbations of heart development are most common, as congenital heart disease (CHD) affects approximately 1% of live births (Hoffman and Kaplan, 2002). Given the essential function the heart plays during embryogenesis, it is likely that many more CHDs lead to lethality prior to birth. In addition, CHDs often require surgical correction and can lead to health complications throughout life. Thus, developing better treatments for CHDs remains a major challenge for human health.

Studies implicate genetic abnormalities in the etiology of CHD. Furthermore, these lesions often stem from haploinsufficiency of key cardiac transcriptional regulators. For instance, mutations in *TBX5*, *NKX2-5*, and *GATA4* have been identified in familial CHD (Schott et al., 1998, Basson et al., 1997, Li et al., 1997, Garg et al., 2003). *TBX1* loss of function is thought to contribute to the cardiac abnormalities observed in DiGeorge

syndrome (Lindsay et al., 2001). Hence, perturbation of transcriptional networks is central to cardiac malformation and CHD. Understanding how cardiac transcriptional programs are modulated by transcription factors and chromatin regulators during development is therefore fundamental to understanding how these programs are disrupted in CHDs. Interestingly, while the transcription factors that pattern cardiac gene expression have been studied for over a decade, the function of chromatin in cardiogenesis is still poorly understood. Not surprisingly, studies suggest chromatin regulators are likely involved in CHDs as well (Ng et al., 2010, Nimura et al., 2009, Vissers et al., 2004). New studies looking at chromatin organization and regulation in cardiomyocyte differentiation are necessary to provide a foundation for a more complete understanding of gene regulation during cardiogenesis and thus CHD etiology.

1.2 Cellular basis of mammalian heart development

The heart is the first organ to form and function and its development is now well-characterized (Figure 1.1)(Srivastava, 2006, Buckingham et al., 2005). In the mammalian embryo, cardiovascular cell types derive from the inner cell mass, a pluripotent group of cells in the blastocyst-staged embryo that gives rise to all embryonic tissues. Two days later in the mouse, these cells have differentiated into a cup-shaped epithelium known as the epiblast, which soon undergoes gastrulation. During gastrulation, cell migration through the primitive streak leads to the development of the mesoderm and endoderm, which along with the epiblast-derived ectoderm constitute the

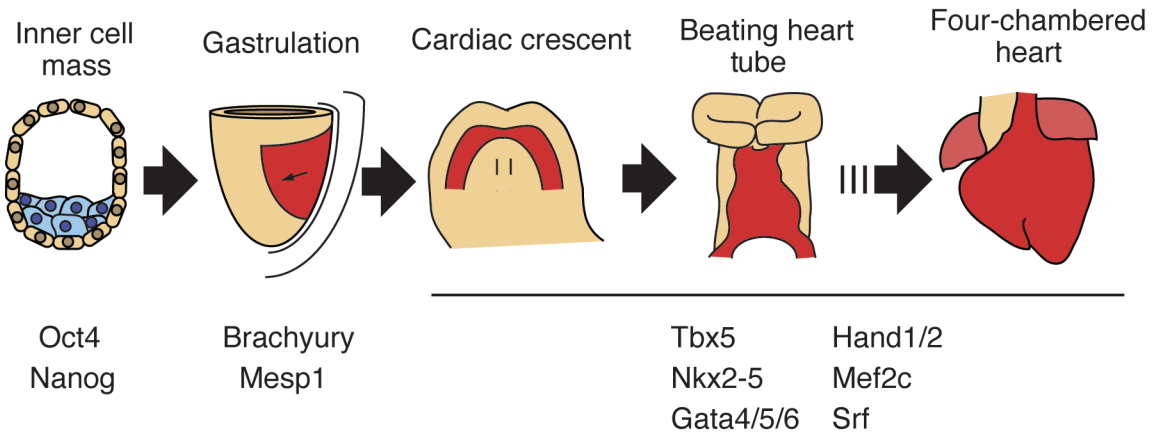


Figure 1.1. Multiple stages of mouse cardiac development. Cartoon of mouse heart development. Pluripotent cells of the inner cell mass (left most panel; blue cells) will differentiate into a cup-shaped epiblast (second panel). Cardiogenic mesoderm (red) will migrate as bilateral wings around the epiblast and populate an anatomical region known as the cardiac crescent (third panel; red). As development progresses, these cells will migrate to the midline and fuse to form a linear heart tube (fourth panel). This tube undergoes extensive remodeling to give rise to the four-chambered mammalian heart (right most panel). Dashed arrow represents multiple intermediate stages not depicted. Under each panel, an incomplete list of transcription factors known to function at this stage is shown.

three germ layers of the embryo. The heart derives from mesodermal cells that arise early during gastrulation at the posterior end of the embryo and migrate anteriorly as bilateral wings of cardiogenic mesoderm. Upon reaching the anterior pole of the embryo, these precursors populate an anatomical region known as the cardiac crescent (given its resemblance to an upside-down horseshoe) and initiate cardiac differentiation. As the embryo begins the process of folding, the differentiating cardiac precursors coalesce to form a linear heart tube, composed of an inner layer of endocardium (endothelial cells) surrounded by a myocardial sheath (cardiomyocytes). It is at this stage, after heart tube formation, that cardiac contraction initiates and first establishes the embryonic circulatory system.

In order to achieve the canonical four-chambered arrangement, the developing heart undergoes a complex series of morphological events. This starts with rightward looping of the heart tube, which continues to twist until the inflow tract (posterior end of the linear heart tube) resides dorsal of the outflow tract (anterior end of the linear heart tube) and thus aligns the future left and right atria with their corresponding ventricles. During looping, the heart continues to grow. This has now been shown to arise from the migration of a distinct pool of cardiac progenitors that reside in the neighboring pharyngeal mesoderm termed the second heart field (SHF) (Kelly et al., 2001, Mjaatvedt et al 2001., Waldo et al., 2001). These progenitors add to the heart via both the inflow and outflow tracts and begin cardiac differentiation. As the heart continues to mature, it undergoes extensive remodeling to ensure proper septation of the left and right ventricles and atria as well as formation of the four pairs cardiac valves.

1.3 Transcriptional control of cardiac differentiation

Fundamental to cardiogenesis is the progressive differentiation of progenitor cells. These cells transition from a pluripotent state and commit to mesodermal identity. This is followed by specification of mesodermal cells to the cardiac lineage. Finally, specified cardiac precursors differentiate to specialized, functional cardiovascular cell types.

The transition in cell identity mediated during cardiac differentiation is accomplished predominately through transcriptional changes, which are driven by groups of tissue-specific transcription factors. Many of these factors have been identified. For instance, *T* (*Brachyury*) is a T-box transcription factor that shows specific expression in cells within the primitive streak of the embryo (Wilkinson et al., 1990, Beddington et al., 1992). *T* mutant embryos show defects in mesoderm formation and mesoderm-derived structures, consistent with a required role in mesodermal cells (Wilkinson et al., 1990). A subset of mesoderm also expresses the basic-helix-loop-helix (bHLH) factor *Mesp1*, which has been proposed to indicate one of the earliest molecular signatures of the cardiac lineage (Saga et al., 2000). *Mesp1* is required for proper mesoderm migration and appears to directly activate cardiogenic TFs in ES cells (Bondue et al., 2008). As mesodermal cells further differentiate, *Mesp1* expression is downregulated, leading to a transient expression pattern in the embryo. Thus, *T* and *Mesp1* function atop a transcriptional hierarchy implemented during cardiac differentiation.

Downstream of mesoderm commitment, multiple families of cardiogenic transcription factors function to pattern cardiac gene expression and promote terminal differentiation. These include members of many transcription factor families, including

homeobox (*Nkx2-5*), T-box (*Tbx5*), basic-helix-loop-helix (*Hand1*, *Hand2*), zinc-finger (*Gata4, Gata5, Gata6*), and MADS-box (*Mef2c*, *Srf*) transcription factors (TF)(Bruneau, 2008, Srivastava, 2006). Additionally, *Islet1* (*Isl1*), a LIM/homeodomain TF appears to play an important role in SHF progenitors and specifically marks this cell population (Cai et al., 2003). In concert, these TFs regulate essential aspects of heart development, which is demonstrated by the fact that loss of many of these factors leads to cardiac malformation in the mouse (Lyons et al., 1995, Bruneau et al., 2001, Bi et al., 1999, Srivastava et al., 1997, Watt et al., 2004, Cai et al., 2003). Specifically, cardiogenic TFs are believed to stabilize a cardiac transcriptional program by reinforcing the expression of other cardiac TFs. There is also direct evidence that cardiac TFs activate cardiac genes. For instance, multiple cardiac TFs directly regulate *Nppa*, which encodes atrial natriuretic factor (ANF) (Durocher et al., 1996, Durocher et al., 1997, Hiroi et al., 2001). Furthermore, cardiac TFs bind and/or activate essential genes involved in cardiomyocyte contraction (Zhang et al., 2005, Ghosh et al., 2009, He et al., 2011). Together, these studies demonstrate the central role of transcription factors in establishing cardiac gene expression and modulating heart development.

Despite being instructive regulators, transcription factors represent only part of the complex regulatory apparatus cells utilize to fine-tune developmental gene expression; the function of these factors can be modified by extracellular signaling cues and chromatin structure. The latter is of particular interest, as it likely provides a template for propagating gene expression changes across many cell cycles and thus perdurance of developmental transcriptional programs. Despite numerous chromatin regulatory proteins being identified as required for heart development (Chang and Bruneau, 2012), still very little is known about the global organization of chromatin and how this is

regulated in cardiovascular cell types. This is fundamental for the understanding of transcriptional control during the development of the heart.

1.4 Chromatin structure and regulation

Eukaryotic DNA is packaged into higher-order chromatin, which facilitates its condensation into the limited three-dimensional space of the nucleus (Figure 1.2a). Consequently, chromatin structure presents a point of potential regulation of genomic processes such as RNA transcription. The fundamental unit of chromatin is the nucleosome core particle, which is composed of two copies of four distinct histone proteins: H2A, H2B, H3, and H4 (Kornberg and Lorch, 1999). These histones assemble as a H3-H4 tetramer and two H2A-H2B dimers into a symmetrical nucleosome that spools 147 basepairs of DNA in 1.67 left-handed superhelical turns (Luger et al., 1997). Nucleosomes are regularly distributed throughout the genome in roughly 200 bp intervals (Kornberg, 1974), which gives first-order chromatin structure the appearance of "beads on a string". Furthermore, chromatin can adopt additional orders of compaction leading to highly organized and spatially condensed regions of genomic DNA.

Chromatin structure has a repressive effect on transcription, a phenomenon that has been demonstrated both *in vitro* (Lorch et al., 1987) and *in vivo* (Han and Grunstein, 1988). Mechanistically, this is the result of two phenomena. First, nucleosomes interact with regions of DNA and thus can affect the affinity of transcription factors and RNA polymerase to regulatory regions (Cirillo et al., 1999, Cirillo et al., 2002. Adams and Workman, 1995). Secondly, higher-order structures dramatically compact chromatin and presumably bury DNA away from regulatory proteins (Robinson and Rhodes, 2006,

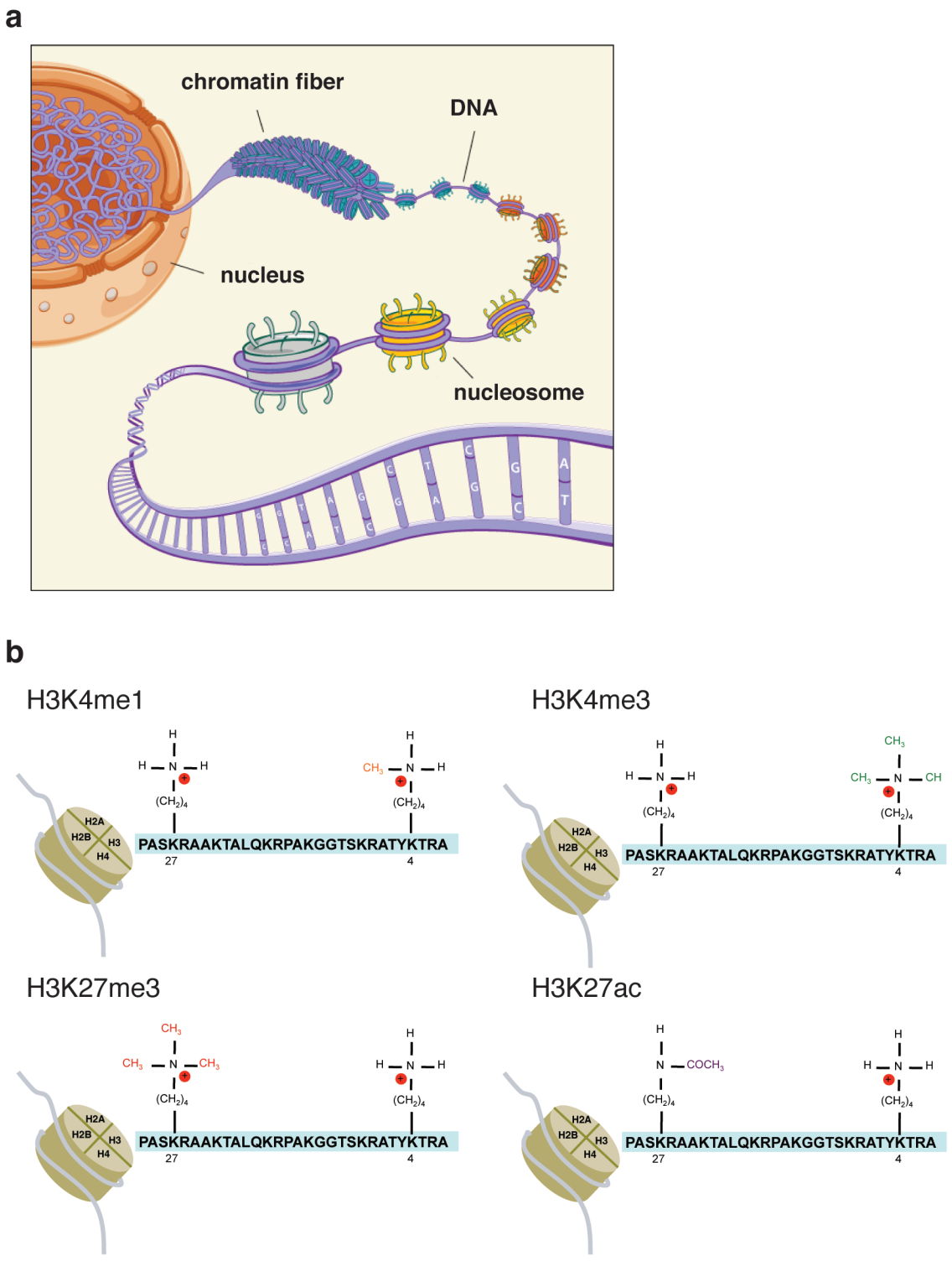


Figure 1.2. Chromatin structure and histone modifications. a. Cartoon depiction of chromatin and higher-order chromatin compaction. Adapted from the Broad Institute website (http://www.broadinstitute.org). b. Cartoon depictions of histone modifications known to associate with specific states of chromatin. Amino acid sequence in blue bar

represents N-terminal H3 tail peptide.

Tremethick, 2007). As such, condensed chromatin or heterochromatin is often devoid of expressed genes (Trojer and Reinberg, 2007).

Eukaryotic organisms have evolved mechanisms to modify the characteristics of chromatin and thereby affect transcriptional output. These include 1) the post-transcriptional modification of histones, 2) ATP-dependent chromatin remodeling, 3) incorporation of histone variants and linker histones, 4) the recruitment of scaffold molecules such as non-coding RNA species and proteins, and 5) modification of the DNA itself via DNA methylation. Together, these processes mediate transcription and DNA accessibility via relocalization of nucleosomes and regulation of higher-order chromatin organization. Of these, histone modification and chromatin remodeling are especially relevant for the studies discussed here, and so will be described in further detail.

1.4.1 Histone modifications

Histone proteins can be modified by a diverse group of post-translational chemical modifications that predominantly occur on unstructured histone tails (Figure 1.2b). These modifications include methylation, acetylation, phosphorylation, sumoylation, ADP-ribosylation, ubiquitination, citrullination, proline isomerization, and crotonylation (Kouzarides, 2007, Tan et al., 2011). In addition, modifications such as methylation can occur in multiple forms on a given residue; a residue can be mono-, di-, or trimethylated. Together, the diverse array of possible histone modifications presents the cell with a source of potentially complex regulation.

Histone modifications are believed to function primarily through two mechanisms. First, certain modifications, such as acetylation, can disrupt the assembly of condensed

chromatin structures, likely due to changes caused in the charge of the histone tail (Tse et al., 1998, Shogren-Knaak et al., 2006). Secondly, these modifications can be recognized by other regulatory proteins. For instance, chromatin remodeling complexes and the heterochromatin protein HP1 have both been shown to be recruited to chromatin through interactions with histone modifications (Ruthenburg et al., 2007, Zeng et al., 2010). In this way, signals encoded by single or combinations of histone modifications can be interpreted by secondary molecules to bring about changes to chromatin organization. This has led to the proposal of codes or languages based on histone modification status (Strahl and Allis, 2000, Wu et al., 2009).

Histone acetylation and methylation represent two of the best understood chromatin marks. While histone acetylation is overwhelming associated with open chromatin and active transcription (Wang et al., 2008), histone methylation had been implicated in both active and repressive chromatin. For instance, trimethylation of lysine 4 on histone H3 (H3K4me3) is associated with active chromatin in yeast, flies, and mammals (Santos-Rosa et al., 2002, Bernstein et al., 2005, Pokholok et al., 2005, Schubeler et al., 2004). Genome-wide studies have demonstrated that H3K4me3 localizes predominantly to transcription start sites (TSS). Furthermore, enrichment of H3K4me3 correlates positively with gene expression levels. In contrast, H3K27me3 is associated with transcriptionally silent chromatin and domains of heterochromatin. In contrast to H3K4me3, H3K27me3 is not limited to the TSS. Instead, H3K27me3 often coats large regions of repressive chromatin, exemplified by its distribution at Hox loci and the inactive X chromosome. Although functional significance of histone modifications is often difficult to identify, H3K27me3 seems to function in part though recruitment of Polycomb repressive complex 1 (PRC1), which can in turn directly

compact chromatin (Francis et al., 2004), as well as deposit other repressive modifications (de Napoles et al., 2004, Wang et al., 2004).

Given their correlation with specific transcriptional states and the relative ease in which their localization can be assessed experimentally, histone modifications have proven useful markers of chromatin state. This approach has been broadly applied to identify unique chromatin states (Kharchenko et al., 2011) and assess how these states vary across cell types or with differentiation (Ernst et al., 2011, Lien et al., 2011). Furthermore, multiple groups have demonstrated enhancer regions can also be identified by looking at histone modifications. Distal enhancers are often enriched for H3K4me1 and not H3K4me3 (Heintzman et al., 2007). Enhancers are also bound by nucleosomes carrying H3K27 acetylation and/or the histone acetyltransferase p300 responsible for catalyzing this mark (Heintzman et al., 2009). Notably, enhancers marked by H3K27ac correlate with high expression of neighboring genes (Creyghton et al., 2010, Rada-Iglesias et al., 2010, Hawkins et al., 2011). These data have led to a model where H3K4me1 marks enhancers in a competent or poised state whereas enrichment of H3K27ac identifies a subset of active enhancers in a given cell-type. Consistent with this model, a subset of poised enhancers in human ES cells is activated upon differentiation (Rada-Iglesias et al., 2010). Taken together, histone modification signatures can provide important information about chromatin state and the potential function of the underlying DNA sequence.

1.4.2 Polycomb repressive complexes

The Polycomb-group proteins represent classical repressors of transcription. They are well conserved throughout metazoa and function in many cellular processes and

pathologies such as X-inactivation, differentiation, and cancer (Surface et al., 2010, Sparmann and van Lohuizen, 2006). Within the cell, Polycomb proteins assemble into two distinct complexes named Polycomb repressive complex (PRC) 1 and 2. Each complex demonstrates histone modifying activity *in vitro* (Wang et al., 2004, Czermin et al., 2002, Kuzmichev et al., 2002, Muller et al., 2002) and mediates repressive chromatin.

Mammalian PRC2 is composed of core subunits *Suz12*, *Eed*, and *Ezh2*. PRC2 is required for embryonic development, as loss of *Suz12*, *Eed*, or *Ezh2* leads to early lethality in the mouse embryo (Pasini et al., 2004, O'Carroll et al., 2001, Faust et al., 1995). In each case, embryos arrest at the peri-implantation stage and fail to initiate or complete gastrulation. Moreover, tissue-specific deletion of *Ezh2* has elucidated roles for PRC2 in heart development (He et al., 2012, Delgado-Olguin et al., 2012, Chen et al., 2012). Thus, PRC2 likely plays important roles throughout embryonic development.

Evidence suggests that PRC2 functions predominantly during development through the regulation of a broad group of developmental regulators. This is likely achieved through its role in mediating di- and trimethylation of H3K27. Indeed, both PRC2 subunits and H3K27me3 are bound near developmental TFs and signaling genes on chromatin in ES cells (Boyer et al., 2006, Lee et al., 2006, Ku et al., 2008, Bernstein et al., 2006, Mikkelsen et al., 2007, Pan et al., 2007). Loss of PRC2 subunits does not affect ES cell pluripotency; rather, it results in derepression of target genes and ultimately differentiation defects (Boyer et al., 2006, Pasini et al., 2007). Importantly, differentiation defects have been reported in other models of differentiation as well (Asp et al., 2011), suggesting a broad role for PRCs in the regulation of cell differentiation.

1.4.3 Chromatin remodeling

Cells employ a diverse group of chromatin remodeling machines to alter chromatin structure. Chromatin remodelers exist as protein complexes, which include an enzymatic subunit that converts the chemical energy released by ATP hydrolysis into translational movement of nucleosomal DNA. In metazoans, there are four main families of remodeling ATPases: SWI/SNF, ISWI, CHD, and INO80. Indeed, these remodelers particular in myriad DNA-dependent processes, including replication, repair, and RNA transcription (Morrison and Shen, 2009, Hargreaves and Crabtree, 2011).

The precise mechanism by which chromatin remodelers alter DNA-nucleosome contacts is still largely unresolved (Cairns, 2007). However, remodelers function through at least three mechanisms: 1) sliding nucleosomes laterally on the DNA (Kassabov et al., 2003), 2) creating DNA bulges or looping within the nucleosomal DNA (Zhang et al., 2006), or 3) displacing nucleosomes altogether from the DNA template (Lorch et al., 1999). These actions open or alter DNA accessibility, which can mediate transcription factor binding (Burns and Peterson, 1997, Ostlund Farrants et al., 1997, Wang et al., 1996a). However, not all remodelers promote chromatin accessibility. For instance, studies on ISWI remodelers suggest that these enzymes sample linker DNA on either side of the nucleosome, ultimately sliding the nucleosome until the free DNA is of equal length on either side (Kagalwala et al., 2004, Yang et al., 2006, Racki et al., 2009). This is thought to result in evenly spaced nucleosomes that facilitate chromatin compaction and transcriptional repression (Sun et al., 2001). Thus, chromatin remodelers function broadly while employing related yet distinct mechanisms to alter chromatin.

1.4.4 BAF complexes

In mammals, *Brahma*/*Brg1* associated factor (BAF) complexes are a major class of chromatin remodelers. BAF complexes share many homologous subunits with yeast SWI/SNF (Wang et al., 1996a, Wang et al., 1996b, Ho and Crabtree, 2010). However, BAF complexes have also diverged during evolution (Figure 1.3); many of their 10-12 subunits are represented by gene families which can be incorporated in a cell type specific manner. Thus, BAF complexes are combinatorial, which has been proposed to facilitate the recognition of complex patterns on chromatin and generate specialized functions of BAF complexes during development (Wu et al 2009). Supporting this, while the catalytic subunit *Brg1* (also known as *Smarca4*) is expressed broadly (Randazzo et al., 1994), other subunits show dynamic and tissue specific expression in the developing mouse (Lickert et al., 2004, Lange et al., 2008, Lessard et al., 2007, Lamba et al., 2008). These subunits can modify complex function by mediating novel TF interactions (Lickert et al., 2004, Debril et al., 2004, Forcales et al., 2012). Thus, BAF complexes are adaptable and likely function distinctly in different cellular environments through unique combinations of subunit composition.

BAF complexes modulate gene expression. Consistent with identified roles for yeast SWI/SNF, BAF complexes cooperate with many factors in transcriptional activation. These include nuclear receptors (Trotter and Archer, 2007), lineage specific TFs (Lickert et al., 2004, Takeuchi and Bruneau, 2009) and viral proteins (Agbottah et al., 2006, Ariumi et al., 2006, Mahmoudi et al., 2006, Treand et al., 2006). In some cases, these factors function in recruitment of BAF to target loci and thereby facilitate chromatin remodeling. However, BAF complexes are also required for TF binding *in vitro* (Ostlund Farrants et al., 1997) and have been suggested to mediate of TF

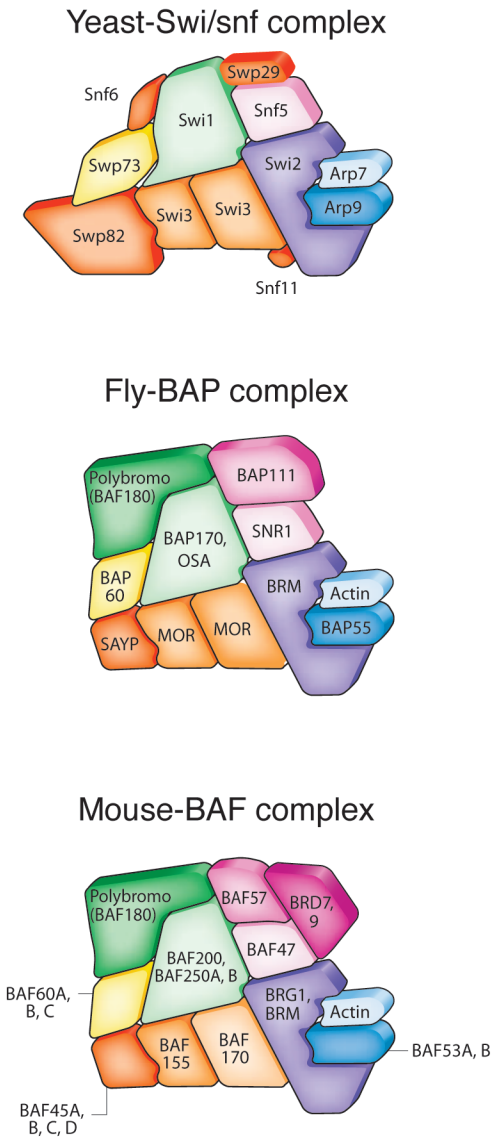


Figure 1.3. BAF and related complexes. a. Cartoon depiction of an archetypical BAF complex (bottom) and homologous complexes from *Saccharomyces cerevisiae* (top) and *Drosophila melanogaster* (middle). Note that BAF complexes have multiple subunits that can be represented by more than one isoform (e.g. BAF60A, BAF60B, or BAF60C). Figure adapted from Ho and Crabtree, 2010.

recruitment *in vivo* (Takeuchi et al., 2009, Agalioti et al., 2000, Lomvardas and Thanos, 2001). Adding to the complexity, BAF complexes have been implicated in gene repression as well (Chi et al., 2003, Bilodeau et al., 2006, Stankunas et al., 2008, Hang et al., 2010, Trouche et al., 1997). *Brg1*/*Brm* has been shown to interact with histone deacetylase (HDAC) complexes (Sif et al., 2001, Kuzmichev et al., 2002, Underhill et al., 2000, Pal et al., 2003), and HDAC recruitment is *Brg1*-dependent on at least one gene (Hang et al., 2010). Thus, BAF complexes play both activating and repressive roles in the regulation of mammalian gene expression.

1.5 Embryonic stem cells as a developmental model

Embryonic stem cells have the unique properties of potentially limitless self-renewal and pluripotency, the latter of which is shared with the cells of the inner cell mass from which they are derived. A hallmark of pluripotent ES cells is their ability to contribute to all cell types, including the germ line, when injected into early mouse embryos. For this reason, ES cells have been widely utilized as a tool for the manipulation of the mouse genome. Beyond this, ES cell differentiation is a valuable tool for the study of development. (Keller, 2002, Murry and Keller, 2006, Niwa, 2010). ES cells readily differentiate into cardiomyocytes when aggregated into embryoid bodies. This has made ES cells an important model for studying cardiac differentiation (Boheler et al., 2002).

Multiple lines of evidence demonstrate the similarity between ES-derived cardiomyocytes and those found *in vivo*. Studies have demonstrated similarities in chronotropic response to adrenergic and cholinergic agents (Wobus et al., 1991), electrical activity (Maltsev et al., 1993, Banach, 2003), and organization of sarcomeric

proteins (Guan et al., 1999). Furthermore, cardiomyocytes show changes in contraction rate and action potentials during differentiation from ES cells that are consistent with what is observed *in vivo*. Gene expression patterns during differentiation of cardiomyocytes from ES cells resemble those seen *in vivo* (Boheler et al., 2002). For example, transcriptional regulators of cardiac myocytes such as *Nkx2-5* and *Gata4* appear prior to more mature markers like ANF, *Mlc2v*, *Myh6*, and *Myh7* in differentiating EBs. Thus, ES cell differentiation reproduces many features of cardiomyocyte differentiation seen in the developing embryo.

While ES cells readily differentiate into cardiomyocytes, they do so with low efficiency using standard protocols. However, there now exist multiple approaches for differentiating cardiomyocytes from ES cells under defined conditions (Laflamme et al., 2007, Kattman et al., 2011, Lian et al., 2012), which have led to clear improvements in the yield and purity of cardiomyocyte differentiations. Of particular importance for the studies described below is the work of Kattman and colleagues, which describes the efficient differentiation of cardiomyocytes from mouse ES cells, human ES cells, and induced pluripotent stem (iPS) cells through the titration of Nodal (via Activin A) and BMP signaling pathway activity. These cultures pass through a progenitor state marked by two cell surface molecules, FLK-1 (also known as *Kdr*) and PDGFRα. This progenitor state also expresses high levels of early mesodermal markers such as *Mesp1*. As with previous studies, temporal activation of cardiac markers is consistent with that seen in the embryo, again suggesting similarity between cardiac differentiation *in vitro* and *in vivo*. As a result, such directed differentiation approaches serve as important tools in the study of cardiac differentiation and development.

1.6 Summary of our studies to investigate chromatin modification and regulation in cardiac differentiation

We sought to investigate the regulation of chromatin during cardiac differentiation. To accomplish our goal, we first adapted an ES cell differentiation protocol that directed the differentiation of cardiomyocytes by defined factors (Chapter 2). We found this approach yielded highly enriched cultures of contractile cells that expressed many markers of the cardiac lineage. Using marker gene expression, we identified four stages of cardiac differentiation during this protocol that we focused on in later studies: ES cells (ESC), mesoderm (MES), cardiac precursors (CP), and cardiomyocytes (CM).

We next measured global gene expression and genome-wide occupancy of multiple histone modifications at each stage of differentiation in order to understand the dynamic chromatin landscape of cardiac differentiation and how changing chromatin states correlate with gene expression (Chapter 2). This was done in a collaborative effort with the laboratories of Dr. Katie Pollard at the Gladstone and Dr. Laurie Boyer at MIT. Using this approach together with hierarchical clustering, we identified many patterns of gene expression and histone modifications during differentiation. We uncovered complex relationships between chromatin regulation and gene expression. For instance, we found gene expression patterns during differentiation could be associated with one or many modes of chromatin regulation. Furthermore, we showed that groups of similarly regulated genes show specific stages and transitions where their chromatin modifications are highly correlated and demonstrated that chromatin pattern can differentiate functionally distinct groups of genes that have similar expression profiles. Lastly, we observed that genes that encode contractile proteins display a novel chromatin pattern. At these genes, we find that H3K4me1 becomes enriched prior to

H3K4me3 and gene activation. We hypothesize this transition may represent a novel point of regulation for the activation of these genes.

Using genome-wide maps of histone modifications, we identified thousands of putative enhancer regions based on chromatin signature. These enhancers were classified as poised or active based on enrichment for H3K4me1 and H3K27ac. We found enhancer utilization to be highly stage-specific during cardiomyocyte differentiation. Furthermore, we found active enhancers are associated with expressed genes, and gene ontology analysis demonstrated these genes function in important and unique processes for each stage of differentiation. We found few active enhancers transition through a poised state during differentiation, but also that these transitions are most frequent between closely related cell types. Lastly, most poised enhancers fail to become activated, and we speculate these enhancers may function during the differentiation of alternative lineages. Together, these findings suggest complex regulation and dynamic usage of enhancers to modulate stage-specific gene expression.

We next identified potential regulatory DNA-binding transcription factors for each stage of differentiation. To this end, we identified dips in H3K27ac enrichment within enhancer regions and looked for enriched transcription factor motifs within these dips. This analysis led to the identification of factors likely to modulate enhancer activity at each stage of differentiation. From this analysis, we uncovered that GATA_Q6 and MEIS1BHOXA9 motifs were found at many of the same enhancers in CP. This observation suggested that GATA factors and Meis factors might cooperate to regulate gene expression. Consistent with this hypothesis, we found that 4 of 5 enhancers containing both motifs responded synergistically to these factors when cloned into a luciferase reporter. Thus, our analysis has revealed novel insights into the regulation of

cardiac differentiation by transcription factors.

In addition, we investigated the importance of a chromatin regulator, *Brg1*, during directed differentiation of cardiomyocytes (Chapter 3). We observed that *Brg1* was most abundant in mesoderm cultures and was downregulated with terminal cardiomyocyte differentiation. Furthermore, we found that *Brg1* was dispensable for cardiomyocyte survival when deleted after cardiomyocytes had already differentiated. Its loss led to the misregulation of a subset of cardiac markers. In contrast, deletion of *Brg1* prior to cardiomyocyte differentiation led to a significant reduction in cardiomyocyte number and increases in cell death.

We focused on the role of *Brg1* during mesoderm induction to better understand how early loss of *Brg1* affected cardiomyocyte differentiation. We found that loss of *Brg1* led to the dysregulation of gene expression, including upregulation of many developmental regulators. In addition, we found *Brg1*-dependent genes to be enriched for H3K27me3 and include many target genes of Polycomb repressive complexes. Interestingly, *Brg1* was specifically required for H3K27me3 levels at genes upregulated by loss of *Brg1*. These results implicate *Brg1* in the regulation of Polycomb silencing and gene repression during mesoderm induction.

Chapter 2

Dynamic and Coordinated Epigenetic Regulation of Developmental Transitions in the Cardiac Lineage

2.1 Contributions

The author has contributed significantly to the work described in this chapter. All cardiac differentiations, including characterization, identification, and selection of differentiation stages were performed by the author. Furthermore, bioinformatics analysis (described in Figures 2.8, 2.9, 2.11, 2.12) was performed cooperatively with Dr. Rebecca Truty, Dr. Alisha Holloway, Dr. Alex Pico, and Dr. Kirsten Eilertson. Reporter assays described in 2.3.8 were performed with John Wylie. RNA-seq and ChIP-seq experiments were performed by Joseph A. Wamstad and bioinformatics analysis regarding enhancer usage (Figures 2.10, 2.13-27) was performed by remaining authors, especially Dr. Joseph A. Wamstad and Avanti Shrikumar.

2.2 Abstract

Heart development is exquisitely sensitive to the precise temporal regulation of thousands of genes that govern developmental decisions during differentiation. However, we currently lack a detailed understanding of how chromatin and gene expression patterns are coordinated during developmental transitions in the cardiac lineage. Here, we interrogated the transcriptome and several histone modifications

across the genome during defined stages of cardiac differentiation. We find distinct chromatin patterns that are coordinated with stage-specific expression of functionally related genes, including many human disease-associated genes. Moreover, we discover a novel pre-activation chromatin pattern at the promoters of genes associated with heart development and cardiac function. We further identify stage-specific distal enhancer elements and find enriched DNA binding motifs within these regions that predict sets of transcription factors that orchestrate cardiac differentiation. Together, these findings form a basis for understanding developmentally regulated chromatin transitions during lineage commitment and the molecular etiology of congenital heart disease.

2.3 Introduction

Developmental decisions during lineage commitment are precisely coordinated at the genome level, with broad gene expression programs being jointly activated or repressed (Davidson, 2010). Heart development requires the concurrent differentiation of several cardiovascular cell types including endothelial cells, smooth muscle cells, and cardiomyocytes that must be organized into a complex organ. This process involves specification of cells from the pluripotent inner cell mass that become progressively committed to mesodermal and cardiac precursors prior to terminal differentiation (Evans et al., 2010; Murry and Keller, 2008; Srivastava, 2006). Thus, heart development depends critically on precise spatial and temporal control of gene expression patterns, and disruption of transcriptional networks in heart development underlies congenital heart disease (CHD) (Bruneau, 2008; Evans et al., 2010; Srivastava, 2006). It is not known how groups of genes are co-regulated during lineage commitment in the cardiac

lineage.

Chromatin regulation is fundamental to how organisms specify different cell types in embryonic development and generate cellular responses to environmental factors and stress. Studies in a range of mammalian cell types have demonstrated that histone modification patterns are correlated with active, repressed, and poised expression states and may be markers of cell state (Barski et al., 2007; Cui et al., 2009; Ernst et al., 2011; Guenther et al., 2007; Mikkelsen et al., 2007; Zhou et al., 2011). Histone marks can also predict non-coding DNA elements, such as distal enhancers, that regulate tissue-specific gene expression (Creyghton et al., 2010; Ernst et al., 2011; Heintzman et al., 2009; Heintzman et al., 2007; Rada-Iglesias et al., 2011; Zentner et al., 2011). Although we have considerable knowledge of the epigenetic landscape of specific cell types, how chromatin states are coordinated with gene expression during lineage commitment is not known for any cell types.

Emerging evidence indicates that epigenetic regulation is critical for normal heart development and that faulty regulation contributes to congenital heart disease (Chang and Bruneau, 2012). For example, mutations in the histone methyltransferase *MLL2* in humans cause congenital heart defects in Kabuki syndrome (Ng et al., 2010). Moreover, transcription factors implicated in inherited congenital heart disease, such as *Tbx5* and *Nkx2-5*, interact with histone modifying enzymes to regulate gene expression (Miller et al., 2008; Miller et al., 2010; Nimura et al., 2009). Recent studies have demonstrated that the H3K27 methyltransferase *Ezh2* regulates cardiac gene expression programs important for heart development and homeostasis (Delgado-Olguin et al., 2012; He et al., 2012). In addition, direct reprogramming of cardiac fibroblasts into cardiomyocytes is accompanied by epigenetic changes at cardiac-specific genes (Ieda et al., 2010).

Therefore, dissecting the dynamic chromatin and transcriptional landscapes during cardiomyocyte differentiation is critical for understanding heart development and will improve our ability to design stem cell-based therapies for cardiac-related diseases.

Here, we have defined the dynamic epigenetic and gene expression landscapes during cardiac differentiation. We used a directed differentiation system representing the stepwise differentiation of mouse embryonic stem cells (ESCs) into cardiomyocytes (CM) that allows for isolation of developmental intermediates, including mesoderm (MES) and cardiac precursors (CP). We analyzed the distribution of histone modifications at promoters during cardiac differentiation to define the chromatin states that accompany changes in gene expression during these developmental transitions. The use of a dynamic model of lineage determination and differentiation has allowed us to discover previously unknown chromatin state transitions, including a novel pre-activation pattern associated with a set of genes with cardiac functions. We also discovered thousands of stage-specific enhancers, leading to the identification of new transcriptional networks that are deployed during cardiac differentiation. Together, our data illustrate the strength of analyzing a differentiation time course rather than isolated cell types to understand epigenetic regulatory network transitions and reveal a chromatin-level determination of cell fates in the earliest stages of cardiac differentiation that may be key to elucidating temporal control of heart development.

2.4 Results

2.4.1 Characterization of directed cardiomyocyte differentiation

We sought to elucidate how global patterns of gene expression and chromatin

organization are coordinated during cardiac development. However, progenitor populations for the heart are limited in number in the developing embryo, making the study of chromatin in these cells especially challenging. To overcome this limitation, we utilized mouse ES cell differentiation as a model system to dissect chromatin regulation during cardiomyocyte differentiation. We used a directed differentiation approach to drive ES cells to efficiently differentiate down the cardiac lineage. The protocol for this approach is summarized in Figure 2.1. Briefly, ES cells are aggregated to form embryoid bodies (EBs) and allowed to differentiate without exogenous growth factors for two days in a serum-free media. After this incubation, EBs are dissociated and reaggregated in the presence of growth factors VEGF, Activin A (which signals through the Nodal pathway), and BMP4. BMP and Nodal signaling ligands are known morphogens in the gastrulation-staged mouse embryo involved in inducing and specifying mesendodermal subpopulations (Gadue et al., 2005). Following this mesoderm induction phase, embryoid bodies are dissociated and plated as a monolayer in the presence of VEGF, basic FGF, and FGF10 growth factors. These factors have known roles in regulating cardiac differentiation and development (Barron et al., 2000, Reifers et al., 2000) and further promote the efficient differentiation of these monolayer cultures to cardiomyocytes.

Although ES cells have been used as a model for cardiac differentiation for some time, directed differentiation approaches for cardiomyocyte differentiation have been developed only recently (Kattman et al., 2011). Thus, we first focused on characterization of directed cardiomyocyte differentiation. Although variable in efficiency, our directed differentiation approach did generate cultures highly enriched for functional cardiomyocytes. We observed features of cardiomyocytes such as

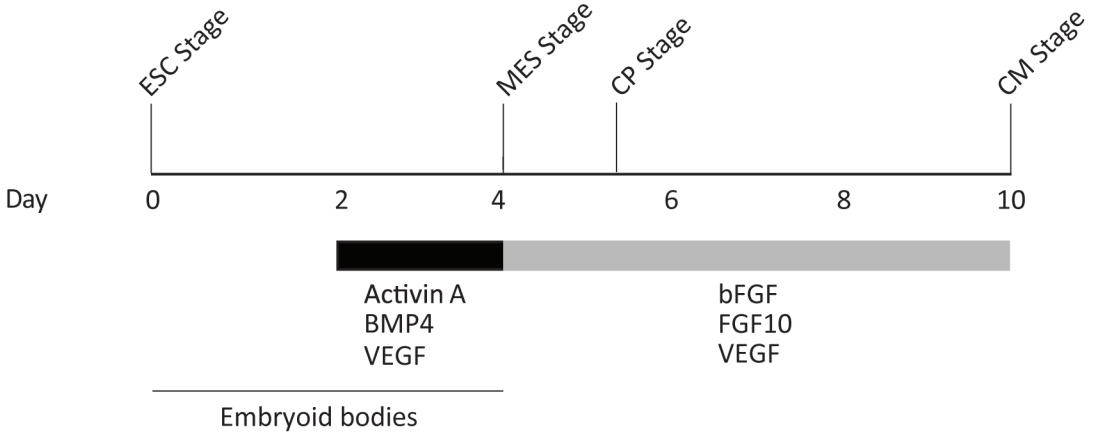


Figure 2.1. Timeline of cardiac differentiation protocol. Black and grey bars represent the time period where differentiating cultures were treated with the growth factors listed below each respective bar. Stages collected for profiling of gene expression and genome-wide occupancy of chromatin hallmarks are indicated.

spontaneously contracting cells beginning on Day 6 of differentiation. Due to apparent cell-cell adhesion, highly efficient differentiations generated sheets of beating cardiomyocytes. Furthermore, the characteristics of contraction reproducibly changed during our protocol (Day 6-Day 12), starting as contracting waves reminiscent of the peristaltic beating observed in the linear heart tube and transitioning to more synchronized contraction at later timepoints (data not shown).

We assessed the presence of a number of cardiac markers by immunofluorescence. TBX5 and ISL1, transcription factors that mark first and second heart fields respectively, were present in most nuclei at Day 5 of differentiation (Figure 2.2a). In addition, we observed many cells that were positive for GFP, a reporter driven by *Nkx2-5* regulatory elements and thus a marker of endogenous *Nkx2-5* expression (Hsiao et al., 2008). We also observed staining for cardiac Troponin T (cTnT), a contractile protein specific to cardiac myocytes. Interestingly, cTnT staining at Day 5 of differentiation was diffuse and poorly organized into filaments, consistent with an immature myocyte (Figure 2.2a). In contrast, by Day 10, cTnT staining revealed a clear sarcomeric organization, demonstrating cardiomyocyte maturation during the differentiation protocol (Figure 2.2b). Tropomyosin (TPM), myosin heavy chain, and myosin light chain 2a (MLC2a) were also detectable in Day 10 cultures (Figure 2.3). We also detected myosin light chain 2v (MLC2v) (Figure 2.3), although its staining appeared more diffuse and was therefore more difficult to distinguish from background staining. In order to measure the purity of cardiomyocytes in our cultures quantitatively, we performed intracellular flow cytometry for cTnT. Again, although our directed differentiation approach had considerable variability, we found that a significant number of differentiations were highly enriched for $cTnT^+$ cells (12 of 20, >60%, Figure 2.2c)

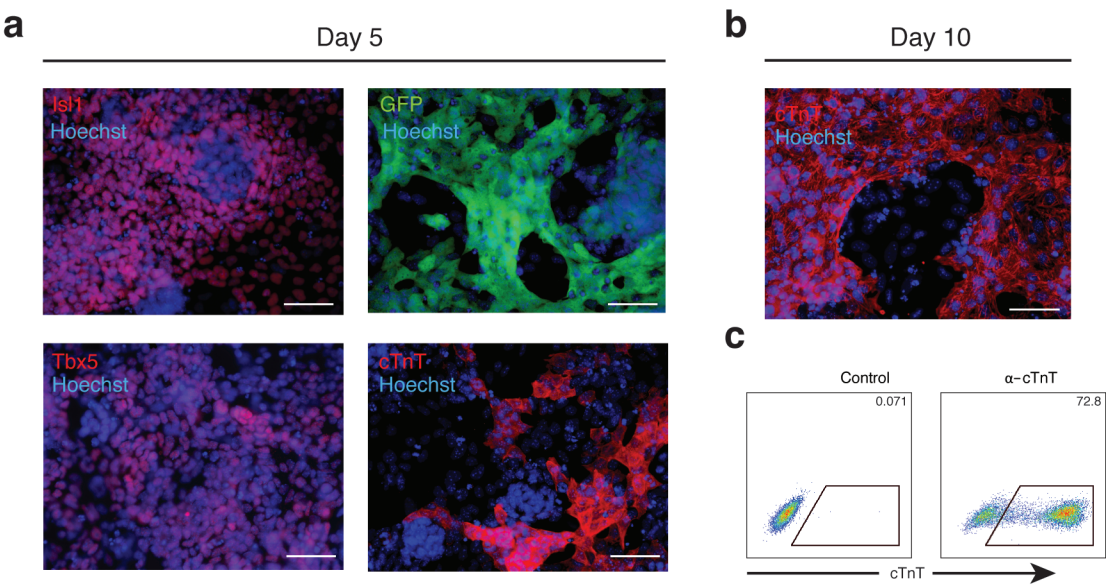


Figure 2.2. Characterization of cardiac markers during cardiac differentiation. a. Immunofluorescence of Day 5 cultures for ISL1, a Nkx2-5 GFP reporter, TBX5, and cTnT counterstained with Hoechst demonstrates high purity of cardiac precursors. b. Immunofluorescence of Day 10 cultures for cTnT counterstained with Hoechst. c. Intracellular flow cytometry of Day 10 cultures stained for cTnT or isotype control demonstrates high purity of cardiomyocytes for a representative differentiation. Scale bar is 50 um.

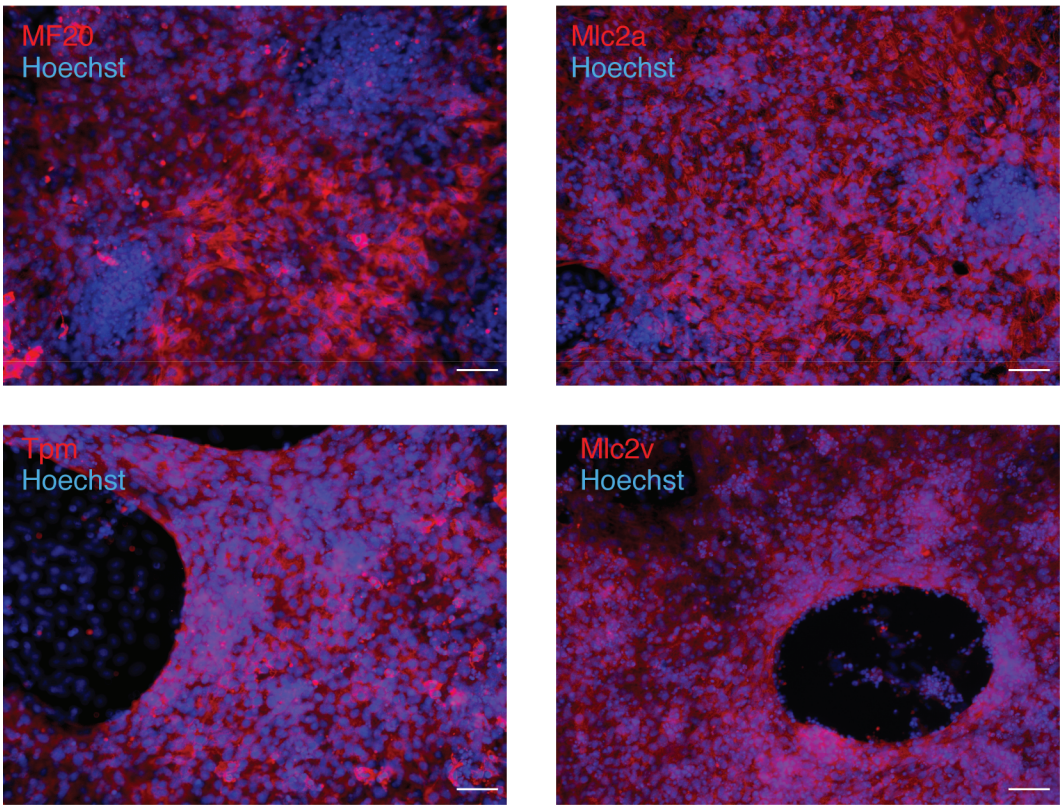


Figure 2.3. Differentiated cultures express cardiac muscle proteins.
Immunofluorescence of Day 10 cultures for MF20 (which detects multiple myosin heavy chain isoforms), MLC2a, TPM, and MLC2v demonstrates expression of multiple cardiac muscle proteins in end stage cardiac differentiations. DNA is stained using Hoechst. Scale bar represents 50 um.

Taken together, these data demonstrate that ES cultures express a broad number of cardiomyocyte markers during directed differentiation.

Activation of TGF-β superfamily signaling pathways BMP and Nodal induce mesoderm differentiation in EBs. Furthermore, the surface markers FLK-1 and PDGFRα are thought to mark an emerging cardiogenic mesodermal population during this stage of differentiation (Kattman et al., 2006, Kattman et al., 2011). Consistent with expectation, we observed an emergence of a FLK-1/PDGFRα double positive (DP) population concurrent with mesoderm induction by BMP and Activin A (Figure 2.4). We observed first appearance of a FLK-1$^-$/PDGFRα$^+$ population, followed by FLK-1$^+$/PDGFRα$^+$ and finally FLK-1$^+$/PDGFRα$^-$ population during mesoderm differentiation (Figure 2.4). Interestingly, although cultures with large numbers of DP cells consistently differentiated more efficiently to cardiomyocytes than cultures that have predominantly FLK-1$^-$/PDGFRα$^-$ cells, we did not observe a clear correlation between DP cell number and ultimate cardiomyocyte yield. That is, large numbers of DP cells did not necessarily translate to robust cardiomyocyte differentiation, and in some circumstances, we noticed cultures predominated by FLK-1$^-$/PDGFRα$^+$ cells yield very pure cultures of cardiomyocytes. Therefore, while FLK-1 and PDGFRα may indeed mark a cardiogenic mesodermal pool, we find that incidence of FLK-1/PDGFRα co-expression does not assure cardiac potential, at least in the particular context of directed ES cell differentiation. Thus, we believe additional markers are required to uniquely identify cardiogenic precursor populations.

Consistent with previous reports, we observed the efficacy of cardiomyocyte differentiation was affected by the relative concentrations of Activin A and BMP4 (Kattman et al 2011). In particular, BMP4 concentration was critical for successful

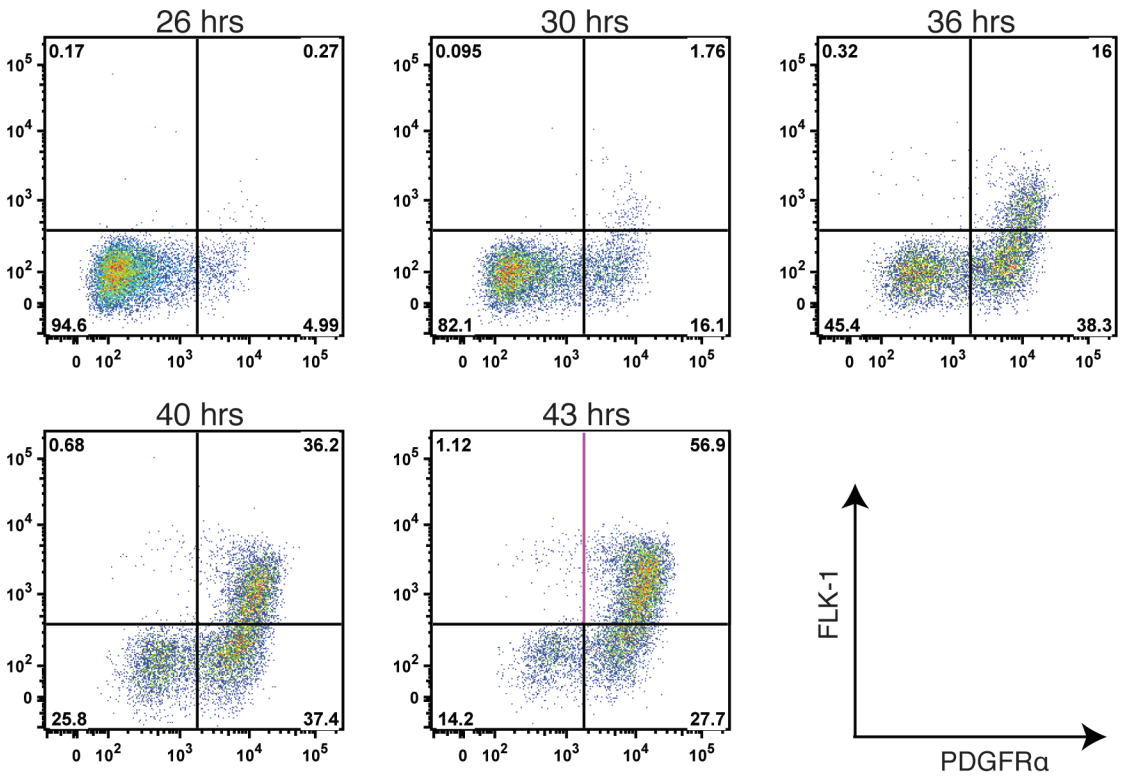


Figure 2.4. Emergence of FLK-1- and PDGFRα-expressing cell population during mesoderm differentiation. Flow cytometry analysis of differentiating E14 ES cell cultures demonstrates the emergence of FLK-1/PDGFRα double positive cells. PDGFRα staining is plotted on the x-axis and FLK-1 staining is plotted on the y-axis. Time listed above each graph indicates hours cultured in the presence of VEGF, Activin A, and BMP4 (hours after Day 2).

cardiac differentiation (Figure 2.5). Modification of BMP4 concentration during mesoderm differentiation led to differences in FLK-1/PDGFRα expression profiles (with higher levels of BMP4 leading to greater number of FLK-1$^+$ cells) and ultimately cTnT$^+$ percentage in end-staged cultures. Therefore, we optimized Activin A and BMP4 along concentration gradients in order to ensure robust cardiac differentiation.

In order to further characterize our directed differentiation protocol, we measured gene expression changes using quantitative PCR. We observed high levels of pluripotency genes *Pou5f1* (herein referred to as *Oct4)* and *Nanog* specifically at Day 0 (Figure 2.6a), consistent with this stage being composed of pluripotent ESCs. Mesodermal marker genes *T* and *Mesp1* showed a sharp peak in expression at Day 4 and were otherwise expressed at low levels (Figure 2.6b). These results are consistent with the appearance of FLK-1 and PDGFRα at this stage and demonstrate that our cultures are enriched for mesodermal precursors at Day 4. We next measured cardiac genes, including *Nkx2-5*, *Isl1*, *Tbx5*, *Actc1*, *Myh7*, *Myh6*, *Myl7*, *Myl2*, and *Tnnt2*. We detected upregulation of *Isl1* and *Tbx5* beginning at Day 4 of differentiation and reaching maximum expression level at Day 6 (Figure 2.6c). *Nkx2-5* and other more mature markers of cardiomyocytes were robustly expressed starting at Day 6 (Figure 2.6c,d). Our cultures showed very low levels of *Gfap*, a marker of ectodermal differentiation (Figure 2.6e). Interestingly, we did measure activation of *Sox17* and *Afp*, two markers of the endoderm lineage (Figure 2.6e). Thus, our direct cardiac differentiation demonstrates robust activation of mesodermal and cardiac marker genes in a temporal pattern consistent with that seen in the developing embryo. We also detect expression of endodermal genes, suggesting partial heterogeneity with the endoderm lineage.

Based on marker gene expression, we identified four stages of differentiation that

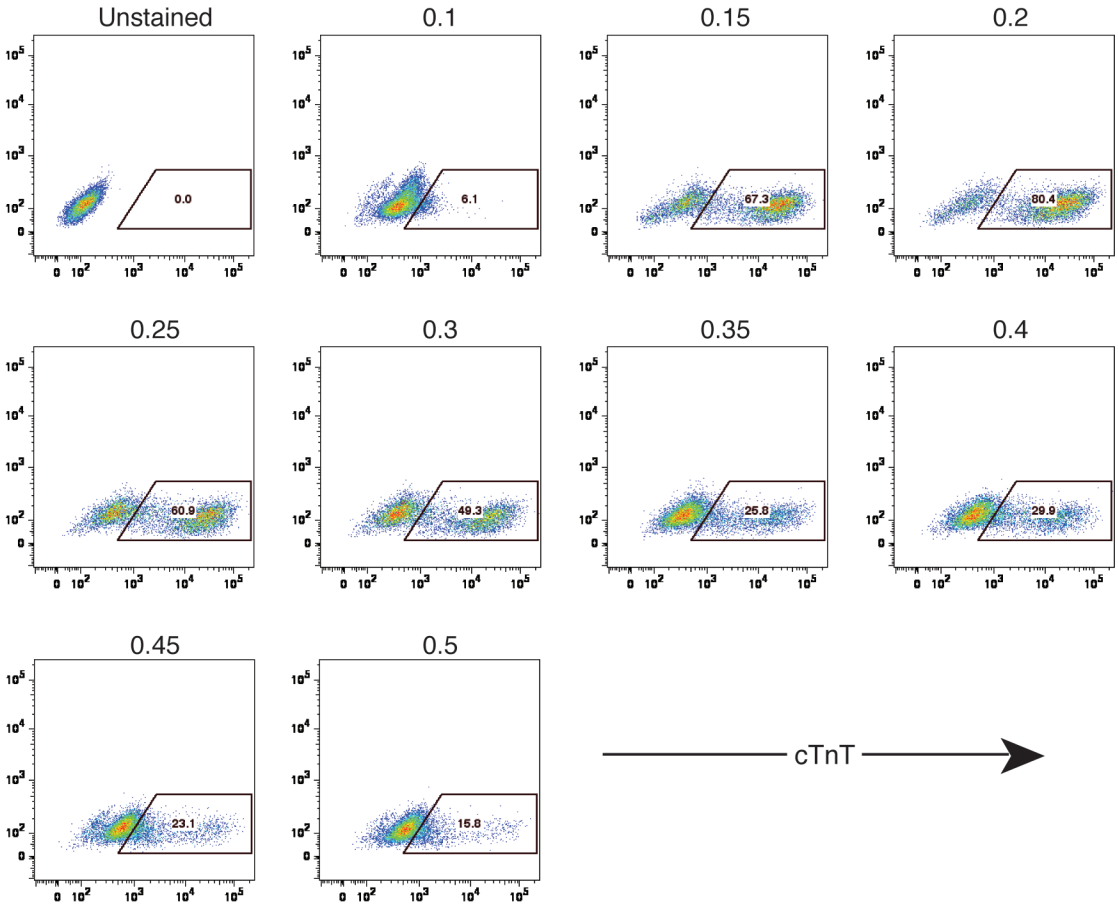


Figure 2.5. BMP4 concentration is critical for cardiomyocyte yield. Intracellular flow cytometry analysis for cTnT on Day 11 of directed cardiac differentiations treated with increasing concentrations of BMP4 demonstrates a relationship between BMP4 and cardiomyocyte yield. Importantly, maximal cardiac differentiation occurs within a narrow concentration range of BMP4. BMP4 concentration was varied during the mesoderm induction stage (between Day 2 and Day4). VEGF and Activin A concentration was kept constant at 5 ng/mL. Numerical values above each plot indicate BMP4 concentrations. Unstained plot was not stained for cTnT and was used to determine background fluorescence. y-axis is a blank channel.

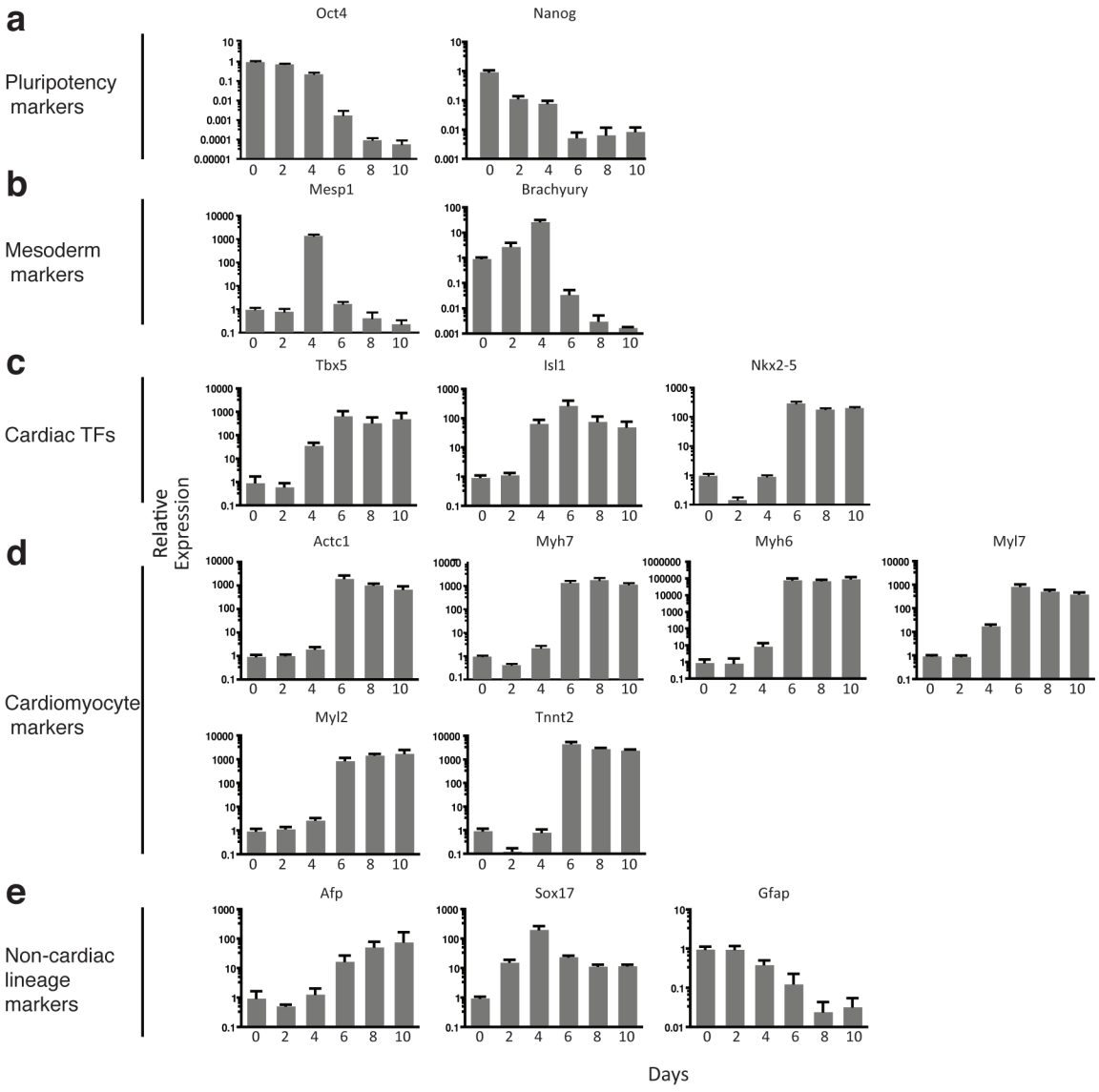


Figure 2.6. Gene expression analysis of directed cardiac differentiation. Relative gene expression changes during cardiac differentiation as measured by quantitative PCR. Marker genes analyzed include pluripotency markers (a), mesodermal markers (b), cardiac transcription factors (c), cardiac muscle genes (d), and markers of ectodermal and endodermal lineages (e).

represent key cell types during the transition of a pluripotent cell to a cardiomyocyte (Figure 2.1). We identified the **ESC** stage that included undifferentiated embryonic stem cells expressing pluripotency markers (*Oct4*, *Nanog*) collected prior to differentiation (Day 0). We identified the **MES** stage that included cultures at Day 4 of differentiation, which express numerous mesodermal markers (*T*, *Mesp1*). The cardiac precursor (**CP**) stage, which included cells transitioning from mesoderm to functional cardiomyocytes, was identified by expression of cardiac TFs, low expression of mature cardiomyocyte markers, and no spontaneous contraction (approximately Day 5). Finally, we included the cardiomyocyte (**CM**) stage, identified by visible contraction and high levels of cardiomyocyte makers (*Myh6*, *Myh7*). We focused on these four stages for our subsequent analysis in this chapter.

2.4.2 Expression and chromatin states in cardiac differentiation

We differentiated ES cells to enriched cultures of cardiomyocytes using our directed differentiation approach (Figure 2.7) and collected cellular material at each stage for subsequent analysis of chromatin modification and gene expression. We first analyzed global expression patterns of polyadenylated transcripts and microRNAs (miRNAs) for these stages by RNA sequencing (RNA-seq) and Nanostring, respectively. We identified over 13,500 genes that were expressed (reads per kilobase exon per million reads (RPKM) >1) during differentiation by RNA-seq. To identify genes with similar expression patterns during differentiation, we performed unsupervised clustering using HOPACH (Pollard and van der Laan, 2005). This analysis revealed many distinct clusters (Figure 2.8a), including genes with stage-specific expression patterns (e.g. Clusters A, L, N, S). These clusters comprised genes specifically expressed at each respective stage (e.g. *Mesp1* and *Eomes* in MES and *Actc1* and *Ryr2* in CM). Furthermore, the stage-specific

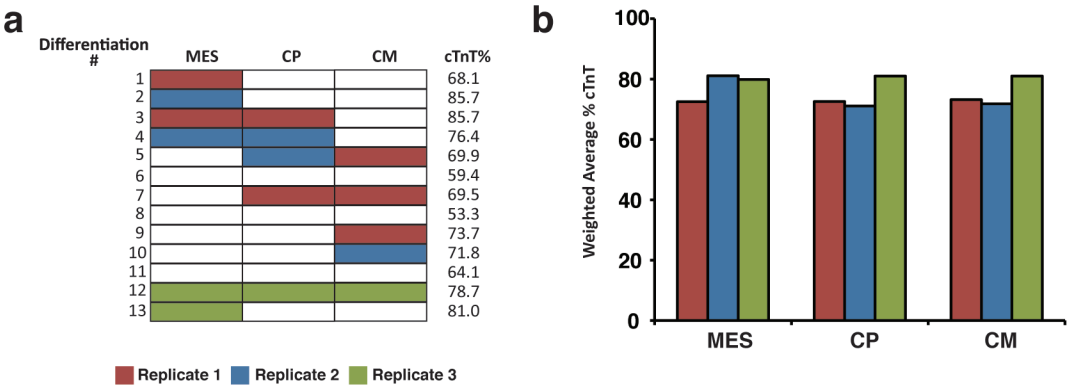
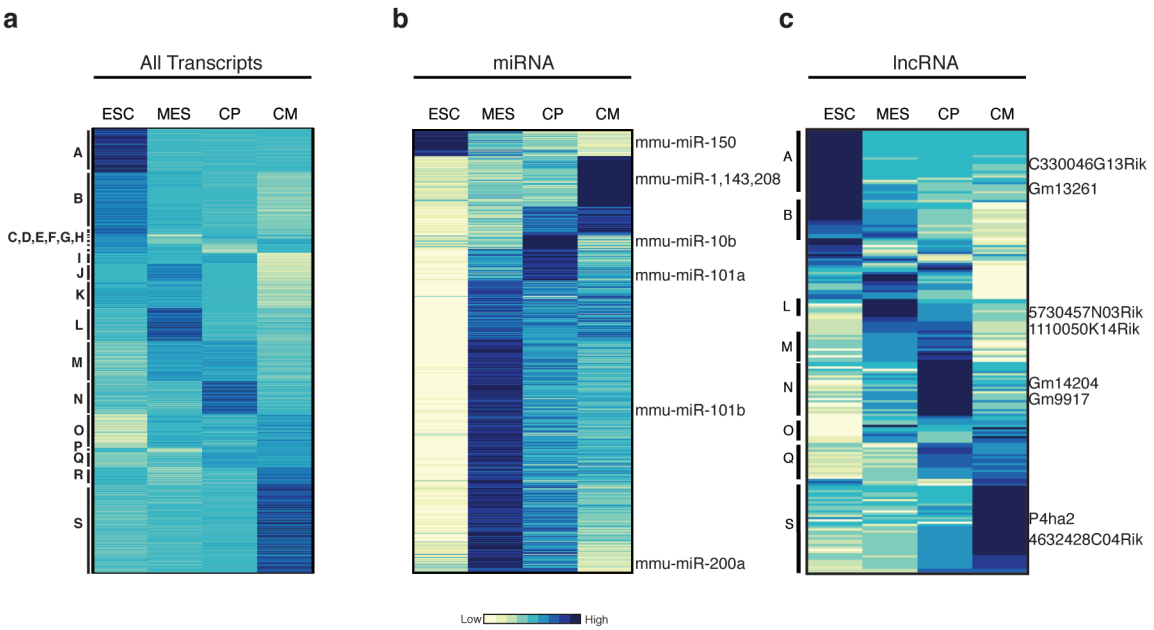


Figure 2.7. Analysis of cardiomyocyte differentiations used for global gene expression and chromatin analysis. a. Cardiomyocyte differentiations that were carried out and pooled for biological replicates at each stage of differentiation. Cells used for Replicates 1-3 are depicted by red, blue, and green shading respectively, with Day 10 cardiomyocyte purity measured by intracellular flow cytometry shown to the right of each row. b. Weighted average of cardiomyocyte purity for each biological replicate at each stage. Weighted average was calculated taking into account the cardiomyocyte yield for each differentiation and its relative contribution to a biological replicate.



d

Cluster	Description	Enriched Gene Ontology (Z-score)	Example Genes
Cluster A	ESC-specific expression	Stem cell maintenance (5.79), Prophase (5.44), Chromosome organization involved in meiosis (4.98)	Nanog, Essrb, Lif, Syce2
Cluster L	MES-specific expression	Embryonic morphogenesis (9.49), Multicellular organismal development (7.29), Cell fate commitment (6.89)	Mesp1, Wnt2b, Cer1, Eomes
Cluster N	CP-specific expression	Homophilic cell adhesion (6.28), Multicellular organismal development (5.37), Definitive hemopoiesis (5.16)	Ephb1, Slit2, Meis1, Isl1
Cluster S	CM-specific expression	Generationof precursor metabolites and energy (7.92), Oxidation-reduction process (7.56), Muscle system process (7.22)	Ndufa1, Actc1, Tnni1, Ryr2

Figure 2.8. Transcriptional analysis of cardiac differentiation. a. Heatmap displaying hierarchical clustering of coding and non-coding polyA+ gene expression across the four cell types. b. Hierarchical clustering of miRNA expression as determined by Nanostring quantification (565 miRNAs included in probe set). c. Hierarchical clustering of lncRNA expression that includes 196 lncRNAs expressed at >1 RPKM in at least one time point. d. Table of example gene expression clusters as well as enriched gene ontology and select genes for each cluster. Letters to the left of heatmaps indicates gene expression cluster.

gene clusters were enriched for Gene Ontology (GO) terms consistent with each stage of differentiation (Figure 2.8d, Supplementary Table 1). Using the Nanostring platform, we measured the expression of over 600 miRNAs and found that these transcripts were also expressed in a dynamic and stage-specific manner (Figure 2.8b). We confirmed the expression of key miRNAs in ESCs, such as the miR290 cluster, as well as known cardiac miRNAs including *miR-1*, *miR-208*, and *miR-143* that were detected at the CM stage. These data also revealed several other developmental stage-specific miRNAs as potential new players in cardiac differentiation. Lastly, long ncRNAs (lncRNAs) comprise a newly identified class of non-coding polyadenylated transcripts with emerging roles in gene regulation (Pauli et al., 2011). Because lncRNAs appear differentially regulated in mammalian cell types, implying key roles in lineage commitment (Cabili et al., 2011), we further analyzed our RNA-seq data to define their expression patterns. We observed striking stage specific expression for lncRNAs during differentiation as well (Figure 2.8c). Together, our transcriptome analysis has identified patterns consistent with cardiomyocyte differentiation as well as novel patterns of protein-coding genes, miRNAs, and lncRNAs that may help identify additional regulators of cardiac differentiation and heart development.

2.4.3 Chromatin state dynamics during cardiac differentiation

Chromatin structure is a major determinant of transcriptional regulation, yet little is known about how chromatin states are coordinated with gene expression during differentiation. To this end, we analyzed chromatin state patterns across the genome during cardiac differentiation. We performed chromatin immunoprecipitation (ChIP) and massively parallel sequencing (ChIP-seq) from a minimum of two biological replicates for several histone modifications at the same time points examined for gene expression

(ESC, MES, CP, CM). The modifications included H3K27me3 and H3K4me3, marks associated with inactive and active promoters, respectively, and H3K4me1 and H3K27ac, modifications associated with promoters and enhancers (Barski et al., 2007; Cui et al., 2009; Ernst et al., 2011; Guenther et al., 2007; Mikkelsen et al., 2007; Zhou et al., 2011). We also determined binding of RNA polymerase II phosphorylated at serine 5 (RNAP), which is enriched at transcriptional start sites (TSS).

Given the dynamic nature of gene expression observed during cardiomyocyte differentiation, we initially focused on elucidating histone modification patterns at transcription start sites (TSSs). To identify gene promoters with similar patterns, we performed unsupervised clustering of ChIP signal derived from a quantitative measure of read counts 2kb around the TSS of each gene. We identified 34 distinct gene clusters (Figure 2.9a). Previous studies observed that chromatin patterns at promoter regions are largely invariant across cell types (Heintzman et al., 2009). Consistent with this, the largest cluster, Cluster 1, displayed high levels of modifications associated with active chromatin and transcription, such as H3K4me3, H3K27ac, and RNAP, across the time course. GO analysis showed that these genes are involved in fundamental cellular functions, such cell metabolism and cell-cycle regulation (Supplementary Table 2). Other clusters revealed that similar to gene expression patterns, chromatin states are highly dynamic suggesting a fundamental role for chromatin regulation in cardiac gene expression and differentiation.

2.4.4 Dynamic chromatin states correlate with distinct expression patterns

We considered that genes with similar temporal expression patterns would share a

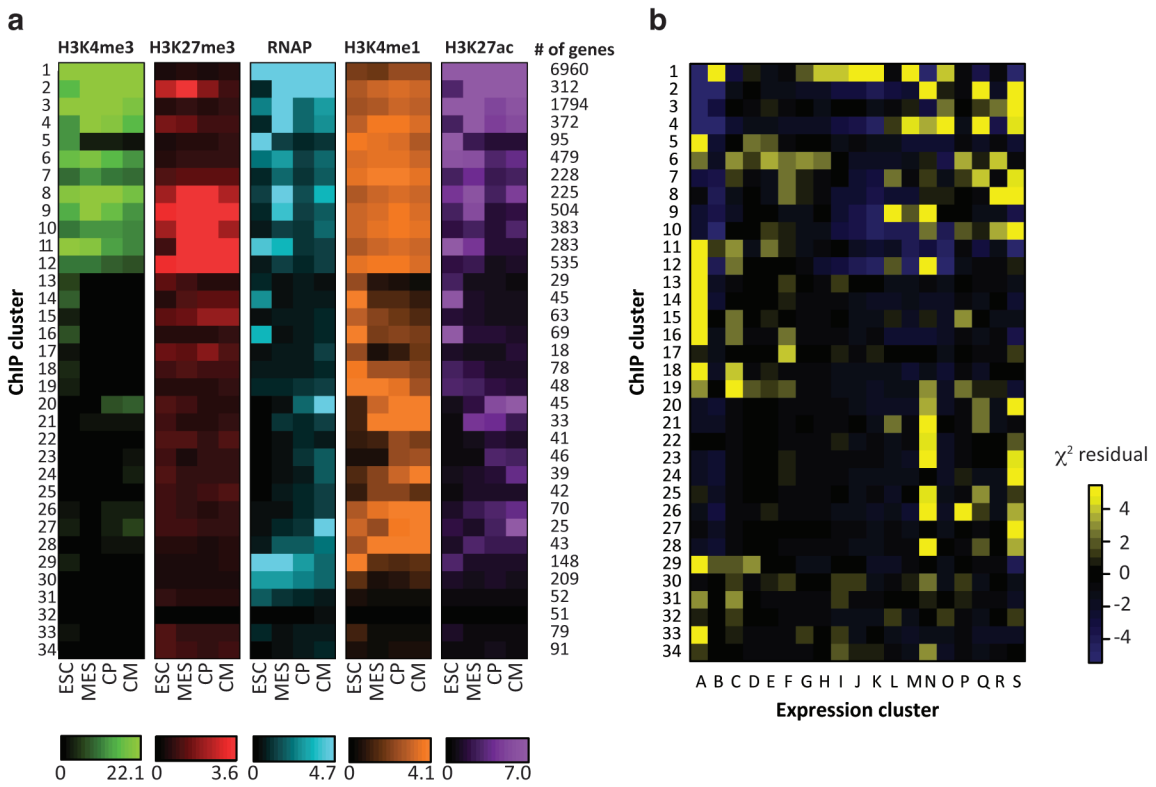


Figure 2.9. Chromatin state transitions during cardiac differentiation. a. Heatmap of hierarchal clustering of genes based on enrichment of histone modifications and RNA Polymerase (serine 5 phosphorylated) within 2 kb of the TSS. Color represents median enrichment value for each cluster of genes. Number of genes within each cluster shown on the right. b. The overlap of genes between chromatin clusters (vertical axis) and expression clusters (horizontal axis). Colors represent the Pearson residuals. Yellow color represents significantly greater overlap between the genes within chromatin cluster and expression cluster than expected by chance.

common chromatin pattern. Conversely, common expression patterns may be represented by multiple different chromatin patterns. To test this, we tabulated the number of genes shared between each chromatin cluster and each expression cluster and tested for statistical enrichment (Figure 2.9b). We discovered evidence for both scenarios. For example, the mesoderm-specific expression cluster L is largely associated with chromatin cluster 9. Conversely, the ESC-specific expression cluster A, which comprises genes that are rapidly silenced upon differentiation, correlated with several distinct chromatin patterns, including chromatin cluster 5 where active marks are lost without gain of additional marks, and cluster 11 in which a gradual loss of active marks is concomitant with enrichment of the repressive H3K27me3 modification (Figure 2.9a). Notably, genes in co-cluster A11 include many genes involved in maintenance of stem cell state, including *Oct4*. *Nanog*, which is also a regulator of the pluripotent state, is found in co-cluster A5, suggesting that the expression of pluripotency regulators is controlled by different epigenetic mechanisms.

We next examined whether similar chromatin and expression patterns in the co-clusters were coordinated at each stage of differentiation. We find that genes with similar expression patterns showed considerable variation in chromatin states during differentiation. For instance, co-cluster A11 includes active genes with highly correlated chromatin states at the ESC stage (Figure 2.10a). Upon differentiation, these genes were downregulated at the MES stage; however, this initial change in expression was not coordinated at the chromatin level until later in differentiation at the CP stage. Conversely, the mesoderm-specific expression Cluster L9, (Figure 2.10b) exhibited a strong correlation between active chromatin marks and gene expression at the MES

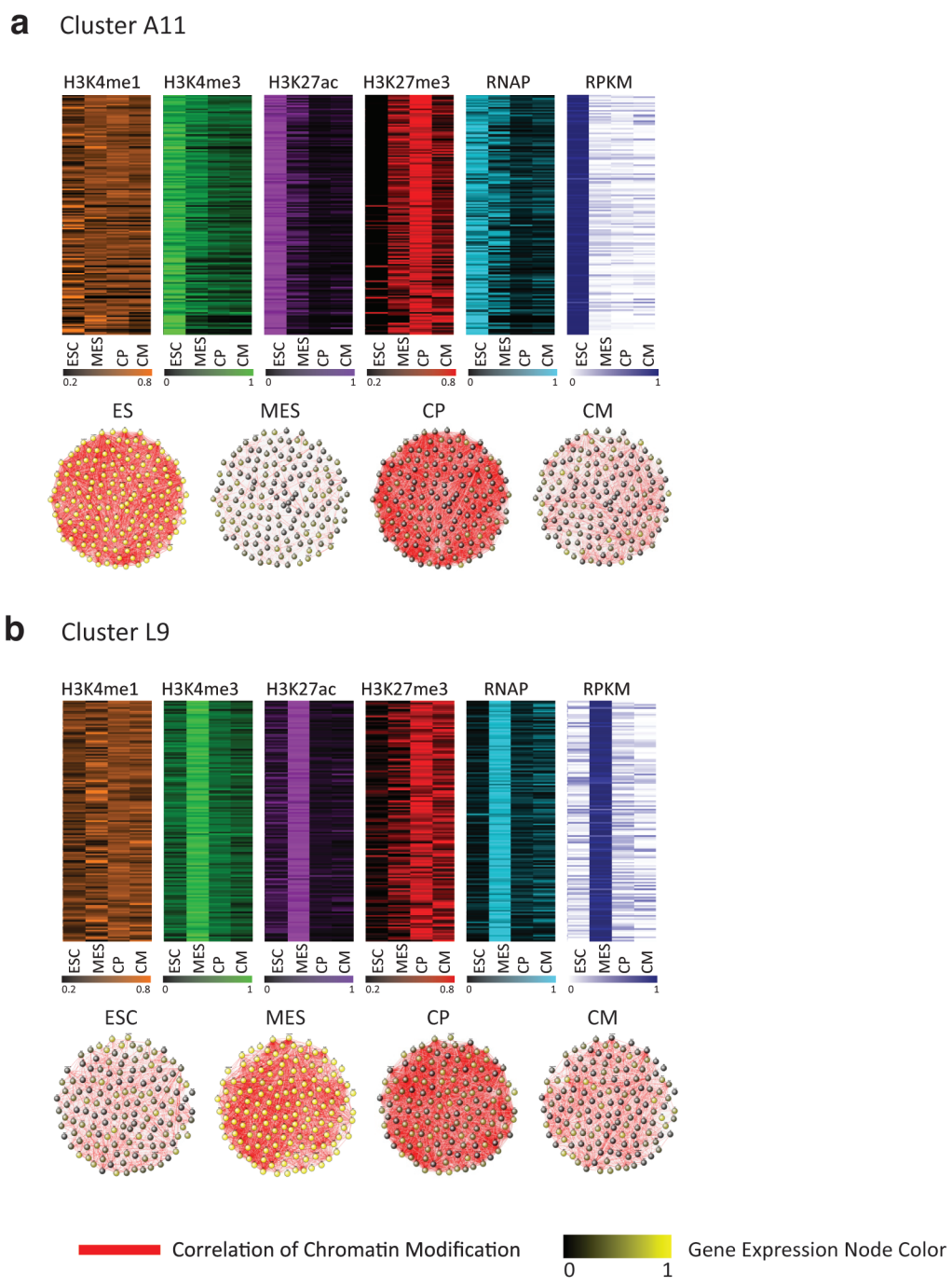


Figure 2.10. Dynamic and highly correlated chromatin and gene expression patterns during cardiomyocyte differentiation. a. Heat maps (top) of magnitude transformed, chromatin fold enrichment values and gene expression values, for co-cluster A11. Co-cluster A11 correlation network (bottom), where nodes represent genes in each module and edges (red lines) represent Pearson correlations of chromatin marks, calculated based on the magnitude transformed values with a threshold of 0.9.

Node color corresponds to gene expression state, where yellow indicates expression and black indicates down regulation. b. Co-cluster L9, analyzed in the same manner as co-cluster A11.

stage followed by a sharp transition to silent chromatin and gene repression at the CP stage.

Our analysis also revealed that functionally distinct genes with a similar expression pattern were distinguished by chromatin state patterns (Figure 2.11). Expression cluster S is comprised of genes expressed at the CM stage yet these genes are associated with diverse chromatin patterns. For example, we find genes with high levels of H3K4me3 and no H3K27me3 enrichment (S3), genes with high H3K4me3 and enrichment for H3K27me3 (S8 and S10), and genes that gain low levels of H3K4me3 during differentiation without H3K27me3 enrichment (S20, S23, S24, S26, S27, and S28). Notably, each subgroup includes genes involved in distinct processes, including metabolism (S3), signaling (S8 and S10), and muscle contraction (S20, S23, S24, S26, S27, and S28). These findings indicate that despite similar patterns of gene expression, functionally related groups of genes can be uniquely coordinated at the chromatin level. Collectively, our data illustrate a complex relationship between chromatin regulation and the gene expression patterns that govern cardiac differentiation and will help reveal the diverse epigenetic regulatory mechanisms that coordinate heart development.

2.4.5 A novel chromatin state transition during CM differentiation

Our promoter clustering analysis revealed an interesting group of genes that showed enrichment for H3K4me1 prior to the onset of transcriptional activation and enrichment of H3K4me3 and RNA Pol II. H3K4me1 has been largely associated with transcriptionally competent chromatin at distal enhancers (Creyghton et al., 2010; Ernst et al., 2011; Heintzman et al., 2009; Rada-Iglesias et al., 2011; Zentner et al., 2011), so we

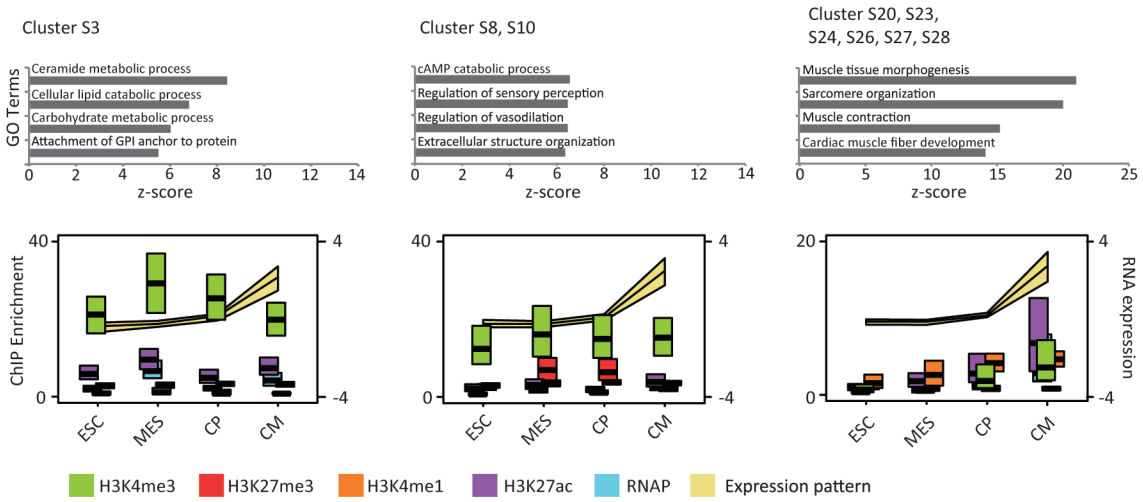


Figure 2.11. Chromatin pattern distinguishes functional gene classes within genes specifically expressed at the cardiomyocyte stage. Distinct groups of genes based on chromatin-expression co-cluster are shown. Chromatin pattern is depicted using boxplots, with median shown as a black line and interquartile range among all genes graphed represented by the extent of each box. RNA expression represents interquartile range normalized gene expression values. Middle line is median expression of all genes considered with upper and lower lines indicating first and third quartiles of this gene group. In each case, enriched gene ontology terms are shown above chromatin pattern.

hypothesized that H3K4me1 may be marking a similar promoter state. Consistent with this, H3K4me3 and H3K4me1 were often enriched at the same TSSs, however we found that a minor fraction (15-20%) of genes marked by H3K4me1 were not enriched in H3K4me3 (Figure 2.12a). This class of promoters was associated with genes with low expression levels (Figure 2.12b). Interestingly, these genes included many contractile protein genes, such as *Actc1* (Figure 2.12c), for which H3K4me1 was present at the MES stage, prior to transcriptional activation at the CP and CM stages. Moreover, H3K27me3 was not enriched at these promoters suggesting that this class of genes lacks Polycomb repression during cardiomyoctye differentiation.

To gain a broader view of this phenomenon, we classified genes based on the pattern of H3K4me1 and H3K4me3 at their TSS. We identified three groups of interest that showed varied patterns of H3K4me3 and H3K4me1 (Figure 2.12d,e). First, we identified genes that gained H3K4me1 before H3K4me3 enrichment and gene expression (Group I). This subset was enriched exclusively for genes in the cardiovascular lineage and those that code for contractile proteins associated with terminal differentiation and cardiomyocyte function. We also identified a group that gained H3K4me1 over time, but failed to gain H3K4me3 or robust expression (Group II). These included muscle lineage genes such as *Ckm*, *Ckmt2*, and *Tcap*, whose expression is associated with cardiomyocyte maturation. Finally, we found genes that transiently gained H3K4me1 at specific stages, but showed no H3K4me3 enrichment during differentiation (Group III). These genes were not expressed above background levels throughout differentiation and have been shown to function in non-cardiac lineages. Thus, H3K4me1 enrichment precedes activation of many cardiac-specific genes during cardiac differentiation. We suggest H3K4me1 may be an important

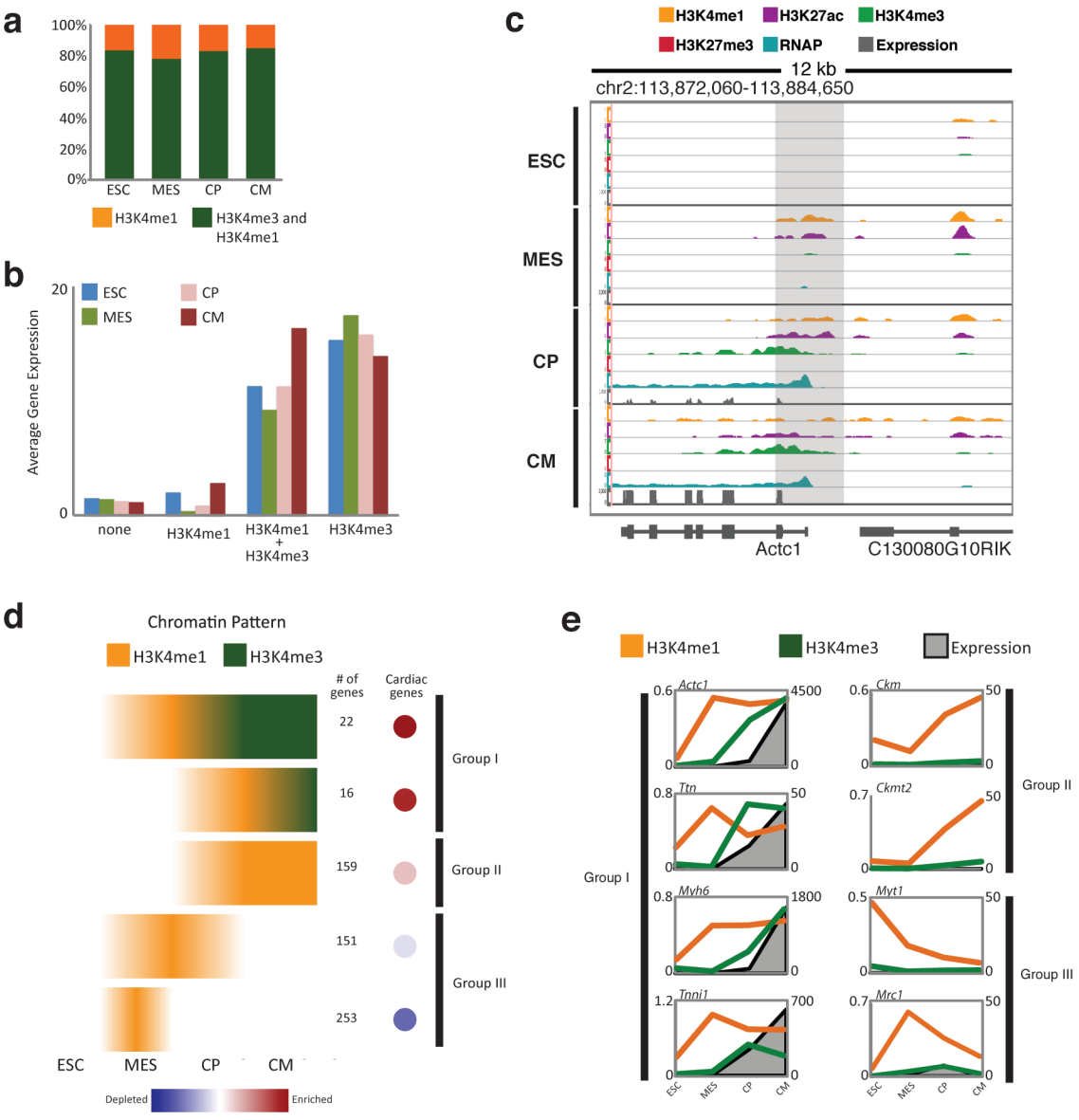


Figure 2.12. H3K4me1 marks cardiac contractile genes prior to gene activation.

a. Fraction of H3K4me1-marked genes that overlap with H3K4me3. An enrichment value at the TSS of 3 was used as the threshold to distinguish marked from unmarked genes. b. Average expression (RPKM) of genes marked with H3K4me1, H3K4me3, both H3K4me1 and H3K4me3, or neither modification for each stage of differentiation. c. Example of a pre-activated gene, Actc1. ChIP-Seq (H3K4me1, H3K27ac, H3K4me3, y-axis reads/million unique mapped reads) and RNA-seq (RPKM) genome tracks (mm9) are shown. Scale for each modification is constant throughout the time course. d. Classification of genes based on gain of H3K4me1 and H3K4me3 enrichment at the TSS. Enrichment value for genes with MGI cardiovascular expression was calculated

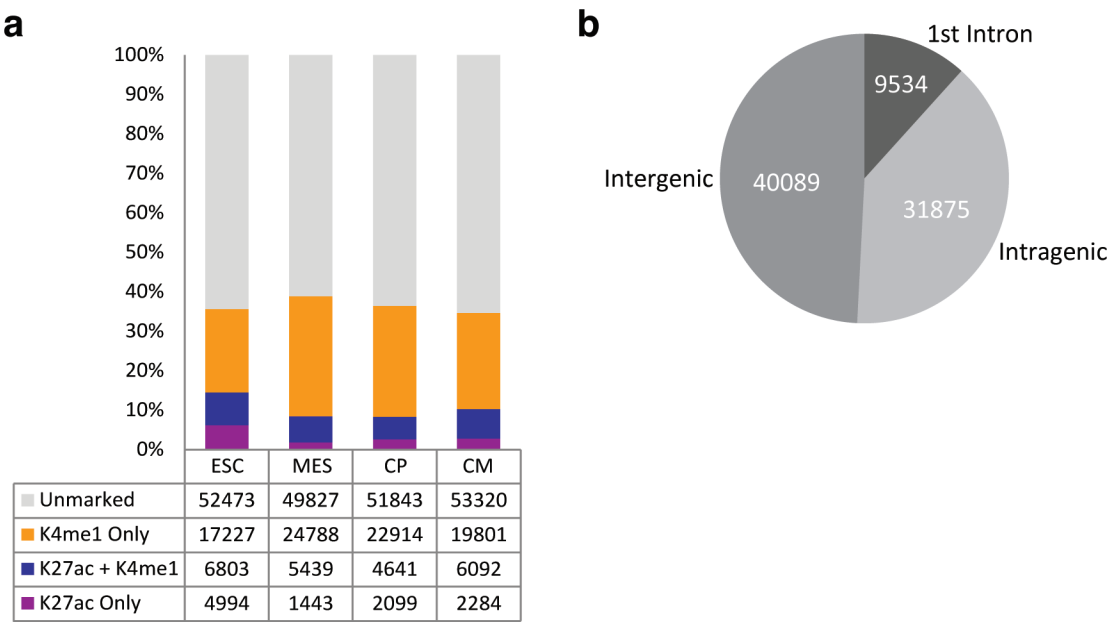
using a Pearson residual. e. Example genes for each group. Left axis represents mean normalized chromatin enrichment values at the TSS. Right axis represents RPKM expression value.

regulatory step for activation of differentiation-associated genes. Collectively, our data show that analysis of chromatin state transitions during a defined differentiation time course provides an unprecedented opportunity to uncover novel patterns of regulation during lineage commitment.

2.4.6 Enhancer activity correlates with cardiac specific programs

While regulation at promoters is important to enact gene activation or repression, distal enhancers play key roles in specifying tissue specific gene expression patterns during lineage commitment. In addition to their enrichment at TSSs, H3K4me1 and H3K27ac demarcate enhancer elements in a wide range of cell types (Creyghton et al., 2010; Ernst et al., 2011; Rada-Iglesias et al., 2011; Zentner et al., 2011). We analyzed our ChIP-seq data for these modifications and identified 81,497 putative distal enhancer regions during cardiac differentiation (Figure 2.13, Figure 2.14, Figure 2.15). H3K4me1 marks most elements at each stage, whereas H3K27ac is enriched at a subset of these regions, and at distal sites independently of H3K4me1 (Figure 2.13a). The broad enrichment of H3K4me1 is consistent with the idea that it represents a general mark of enhancers and open chromatin (Creyghton et al., 2010; Ernst et al., 2011; Heintzman et al., 2009; Rada-Iglesias et al., 2011; Zentner et al., 2011). Comparing H3K4me1 and H3K27ac profiles from neural precursors, liver, and pro B cells (Creyghton et al., 2010) demonstrates that our enhancers are largely unique to the cardiac lineage (Figure 2.16a).

Our set of H3K4me1 and H3K27ac enhancers significantly overlap the smaller subset identified by binding of the p300/CBP co-activators in fetal mouse heart, fetal or



	ESC	MES	CP	CM
Unmarked	52473	49827	51843	53320
K4me1 Only	17227	24788	22914	19801
K27ac + K4me1	6803	5439	4641	6092
K27ac Only	4994	1443	2099	2284

Figure 2.13. Identification of enhancer elements during cardiac differentiation. a. Total number of distal enhancers identified in ESC, MES, CP, and CM categorized by H3K27ac and H3K4me1 status at each developmental stage. b. Distribution of enhancer elements across the genome.

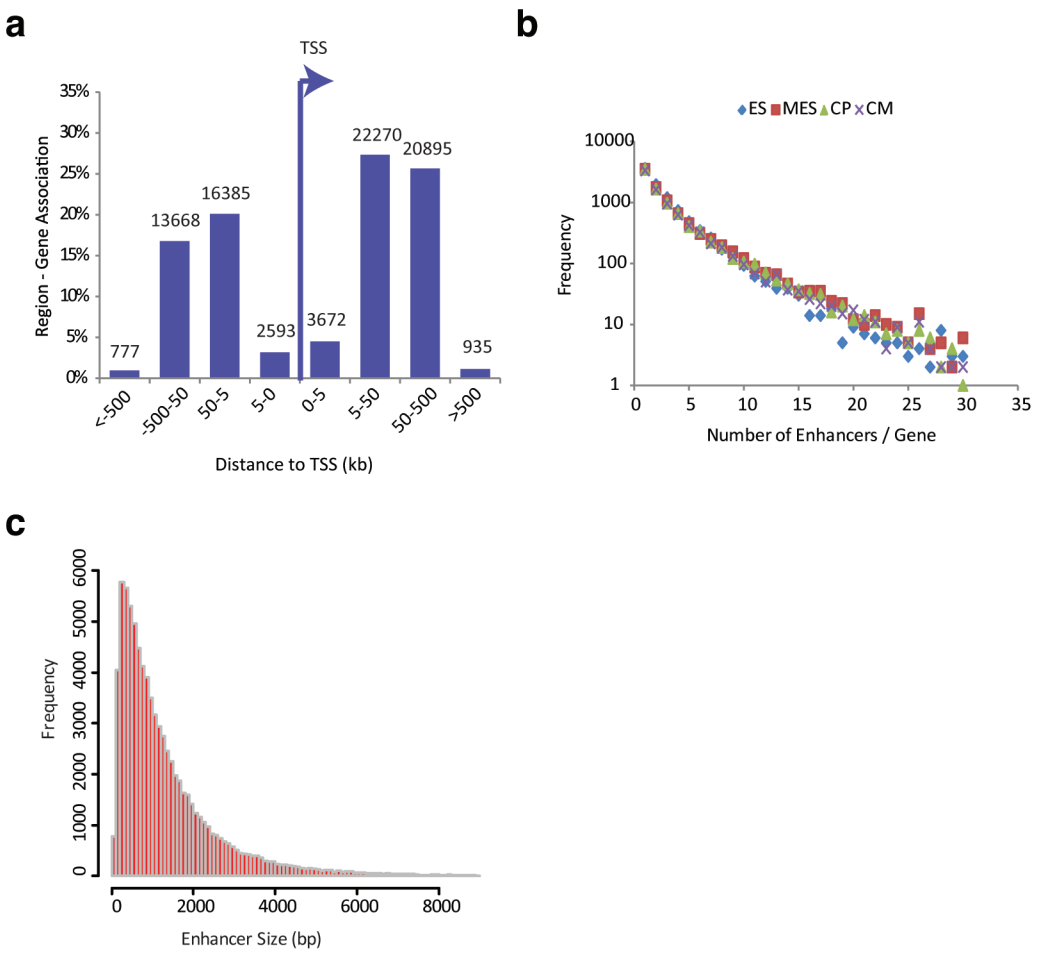


Figure 2.14. Characterization of enhancers identified by chromatin signature during cardiac differentiation. a. The frequency of total enhancers, including active and poised across all time points were plotted relative to transcriptional start sites (TSS) using the GREAT algorithm. The majority of enhancer regions fall within 5-500 Kb up or downstream of an annotated TSS. b. Of genes associated with enhancer regions, the distribution ranges from genes associated with one enhancer to genes associated with up to 30 or more enhancer regions. Genes with approximately 1-3 enhancer regions make up the vast majority of gene/enhancer associations. c. Enhancer regions range in size between 25 to 4000 bp with the majority falling in the 500 bp range. Enhancer region size is a combination of overlapping H3K4me1 and H3K27ac enriched regions.

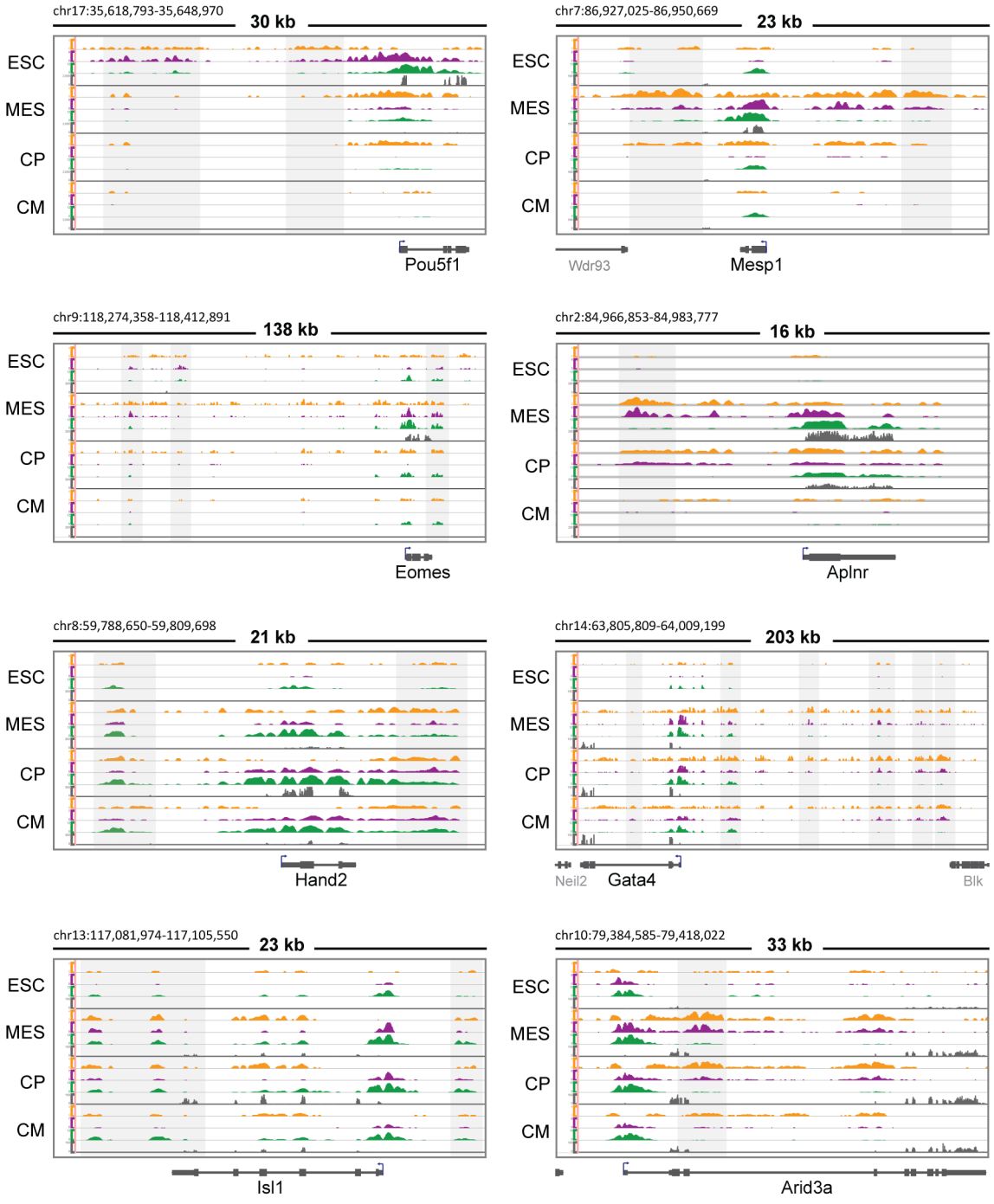


Figure 2.15. Example enhancer regions identified at genes expressed throughout cardiac development. H3K4me1 (orange), H3K27ac (purple), H3K4me3 (green) and RNA-Seq (grey) tracks are shown for the genomic regulatory region for select genes expressed during cardiac differentiation, including *Pou5f1* (*Oct4*), *Mesp1*, *Isl1*, *Hand2*,

Gata4, *Eomes*, *Aplnr* and *Arid3a*. Enhancer regions active in at least one stage of differentiation are vertically highlighted (grey). Enhancer regulatory regions displayed range in size from 16 kb (*Aplnr*) to 203 kb (*Gata4*) and include examples of intergenic, intragenic and 1st intron enhancers. We observe that H3K27ac enrichment at enhancer regions correlates with an increase in expression of the most proximal gene.

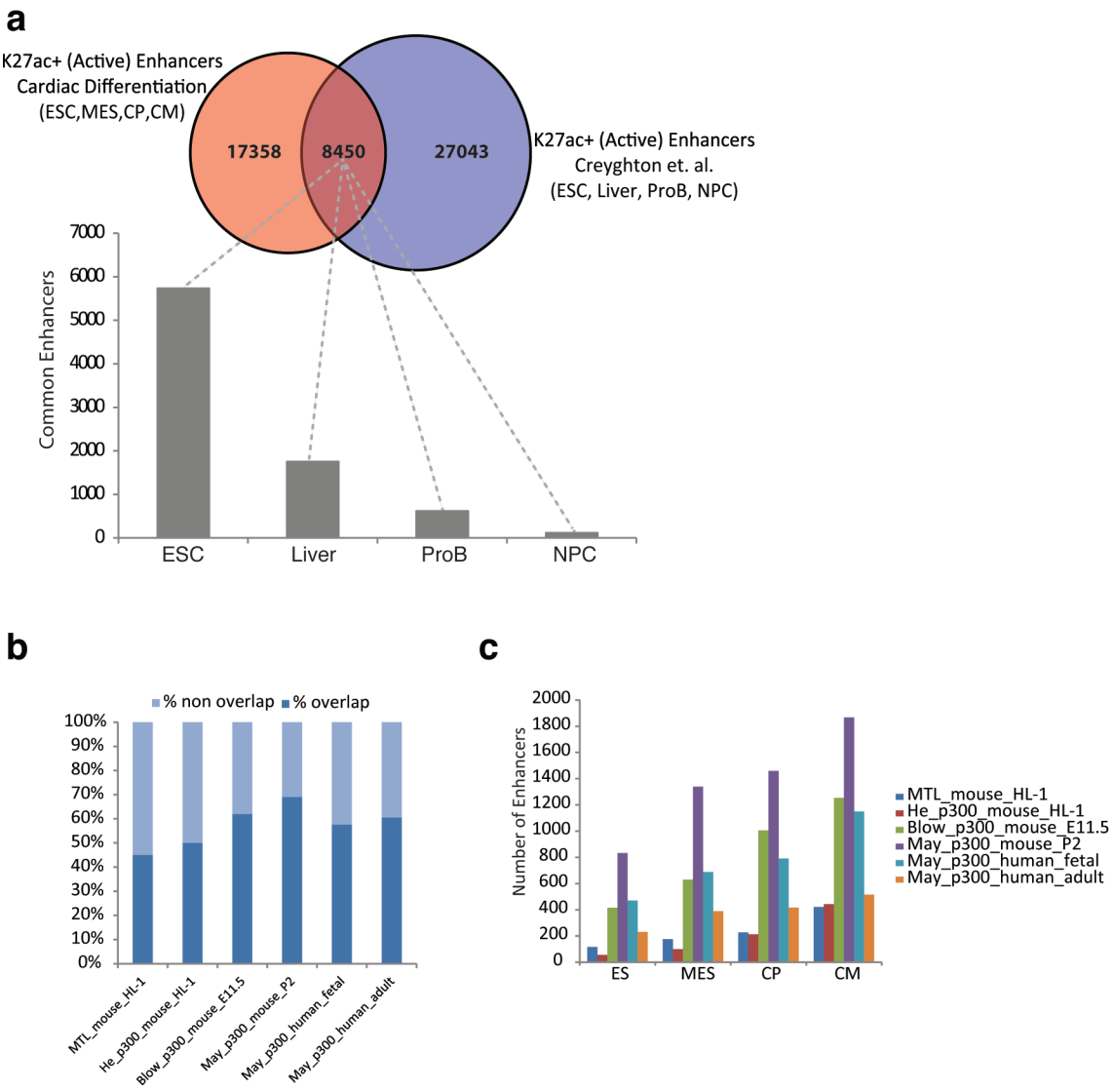


Figure 2.16. Comparison of cardiac enhancers with previous studies reveals significant overlap in predicted cardiac enhancers. a. Venn diagram displaying the overlap in active enhancer regions between our cardiac differentiation time course and the active enhancer regions identified in 4 distinct cell types, ES cells, Liver, Pro B cells and Neural Precursor cells (Creyghton et al., 2010). ES cells, the only cell type common to both data sets, show the largest number of overlapping enhancer regions. b. Our total enhancer list overlaps between ~45% – 70% of the regulatory elements identified in three cardiac related genome wide enhancer screens (He et al., 2011, Blow et al., 2010, May et al., 2011). MTL indicates enhancers identified by multiple transcription factor binding. Otherwise, enhancers were identified by p300 occupancy. c. Overlap analysis of our active enhancers, identified via H3K27ac, with May (p300), Blow (p300), He

(p300) and He (MTL) identified enhancers revealed that the overlap increases as cells progress toward differentiated cardiomyocytes, consistent with the later developmental stage of the cell types used in these prior studies. Similar to the total list of union enhancer list, the strongest overlap at each stage is with p300 marked enhancers in E11.5 mouse hearts and postnatal day 2 mouse hearts.

adult human hearts, or by the binding of multiple transcription factors in the HL1 cardiomyocyte cell line (Blow et al., 2010; He et al., 2011; May et al., 2012), validating our predictions (Figure 2.16b,c). Current evidence indicates that cardiac enhancers are conserved over a limited phylogenetic distance (Blow et al., 2010; May et al., 2012). Consistent with this, we find that only 2-4% of active enhancers identified here overlap with highly conserved (>600) vertebrate Phastcons elements, and that this overlap increases considerably (6-12%) upon analysis of highly conserved elements in placental mammals (Figure 2.17). Thus, we find many more putative enhancer regions that likely function in heart development than were previously known, including enhancers that may regulate the transition from pluripotency to a functionally differentiated state.

Histone modification patterns can distinguish active and poised enhancers, which correlate with tissue-specific expression or the potential of a gene to be expressed later during development, respectively (Creyghton et al., 2010, Ernst et al., 2011, Rada-Iglesias et al., 2011, Zentner et al., 2011). We next classified our enhancers as active (H3K27ac+, H3K4me1+/-) or poised (H3K4me1 only) at each stage of cardiomyocyte differentiation (Figure 2.18a). Most enhancers are poised at each stage, whereas a smaller subset is active. We then compared enrichment patterns of several other histone modifications and RNAP with these enhancer regions. We find that RNAP is highly enriched at active enhancers (Figure 2.18a), consistent with transcription initiation at these regulatory elements (Kim et al., 2010). Conversely, although H3K27me3 has been shown to demarcate poised enhancer elements (Rada-Iglesias et al., 2011; Zentner et al., 2011), this mark had a minimal overlap with our enhancer regions (Figure 2.19). Thus, while poised and active states can be broadly defined by a limited set of histone

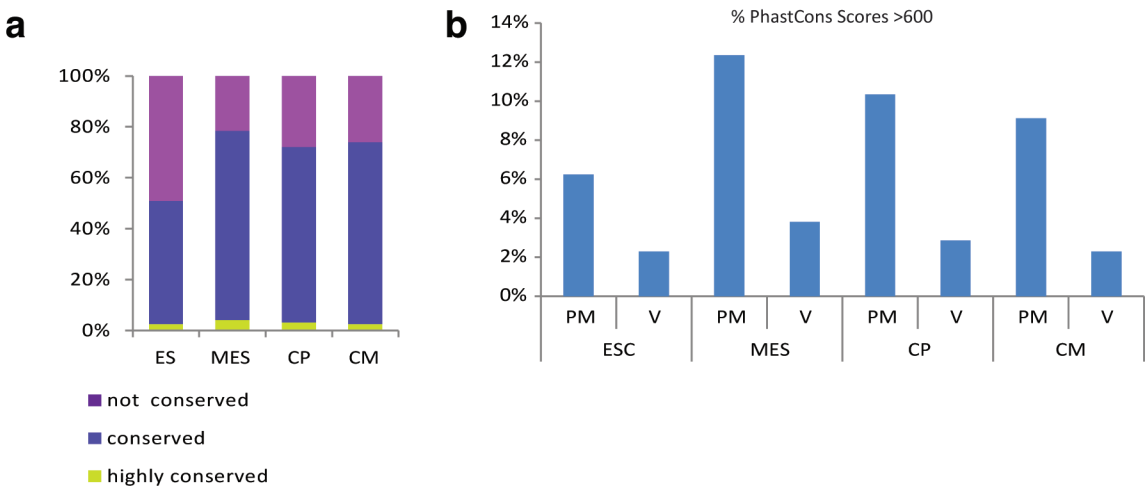


Figure 2.17. Cardiac enhancers display increased conservation within a limited phylogenetic distance. a. Between 50-75% of active enhancer regions at each stage during cardiac differentiation overlap a vertebrate PhastCons element. PhastCons elements (PHylogenetic Analysis with Space/Time models) are a hidden Markov model based measure of evolutionary constraint between species. The results indicate that the enhancer regions identified are conserved across vertebrate species. The percentage of enhancer regions overlapping highly conserved elements is similar to previous reports, hovering between 2-4% depending on the stage. b. Active enhancers display increased overlap with highly conserved (>600) Placental Mammal (PM) PhastCons scores in comparison to Vertebrate (V) PhastCons scores.

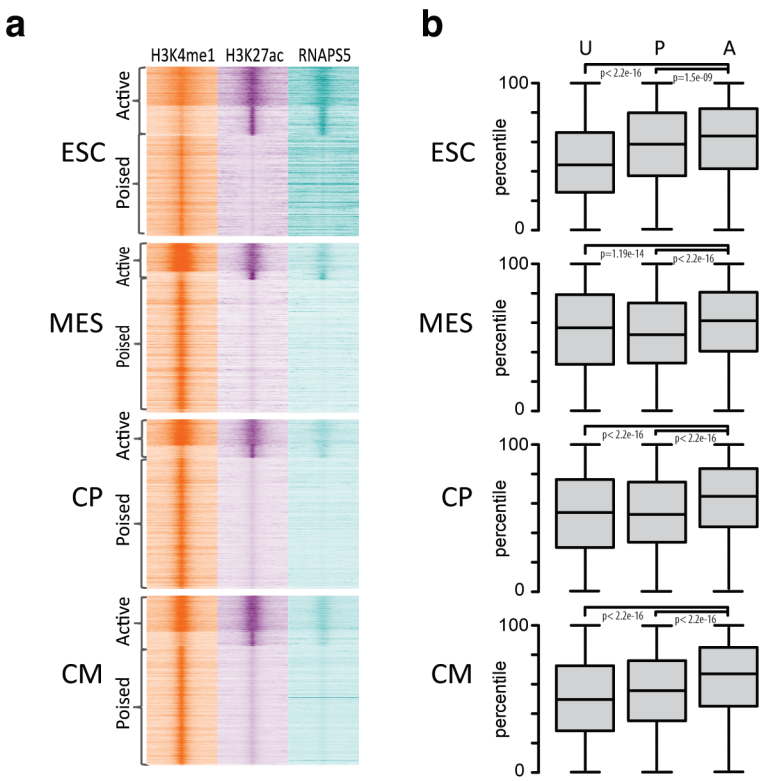


Figure 2.18. Chromatin features and activity of enhancer elements during cardiac differentiation. a. Density of ChIP-seq reads +/- 4kb relative to the midpoint of enriched regions for H3K4me1, H3K27ac and RNA Polymerase (serine 5 phosphorylated form). b. Boxplots of log2 transformed (RPKM) gene expression values for single nearest gene associated with unmarked (U), poised (P), and active (A) enhancer groups. p-values determined by Wilcoxon rank sum test with continuity correction.

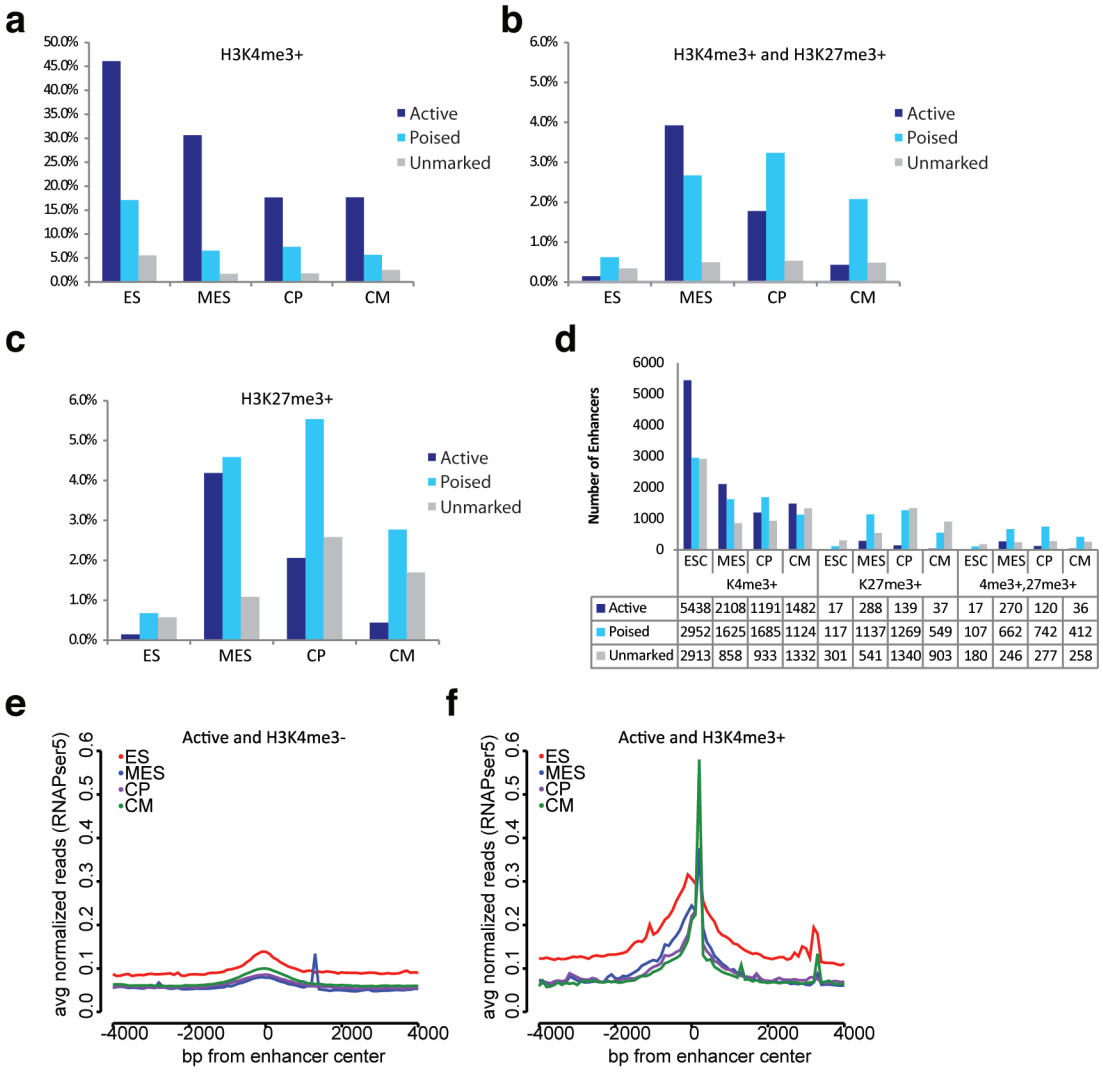


Figure 2.19. H3K4me3, H3K27me3 and RNAPser5P overlap distinct subsets of enhancers identified during cardiomyocyte differentiation. a. H3K4me3 preferentially overlaps active enhancers and displays the greatest overlap with identified enhancer regions in embryonic stem cells (ESC stage), proceeding to decline as cells progress toward cardiomyocytes. b. In contrast, H3K27me3 is present at less than 5% of active, poised and unmarked enhancer regions. The data indicates that H3K27me3 preferentially marks a subset of poised enhancers at each stage of the differentiation time course. As the current literature contains conflicting results as to the extent and presence of H3K27me3 at enhancer regions, it will be important in future studies to address where the variation in identification of H3K27me3 at enhancers resides. c. A

significant proportion of active and poised enhancers marked by H3K27me3 are also marked by H3K4me3, suggesting a potential role for bivalent domains at enhancer regions. d. Summary of enhancer regions marked with H3K4me3 and H3K27me3. e,f. Metagene analysis of RNAPser5 enrichment at H3K4me3 - (e) and H3K4me3 + (f) active enhancers over cardiac differentiation.

modifications, there likely exist many other sub-states that comprise functionally distinct enhancer states.

We hypothesized that active enhancers would correlate with stage-specific gene expression. To test this, we assigned active enhancers to their single nearest gene. As expected, expression of genes associated with active enhancers is significantly higher than genes associated with unmarked or poised enhancer regions at each stage of differentiation (Figure 2.18b). We then determined the enriched GO categories associated with these genes to define their biological functions (Figure 2.20). While active enhancers in ESCs correlate with genes that function in self-renewal and pluripotency, enriched categories progressively become more representative of cardiomyocyte-specific gene functions at later stages. For example, enhancer-associated genes at the MES stage function in mesoderm and embryonic pattern specification. At the CP stage, enhancer-associated genes function in heart morphogenesis and cardiac tissue development. In CMs, we observe a transition toward genes involved in cardiomyocyte structure and function. Many of these genes have important roles in heart development and their dysregulation is associated with heart defects and cardiovascular disease (Bruneau, 2008; Srivastava, 2006). Thus, we identified many new putative enhancers that correlate with genes involved in cardiac specification during embryonic development.

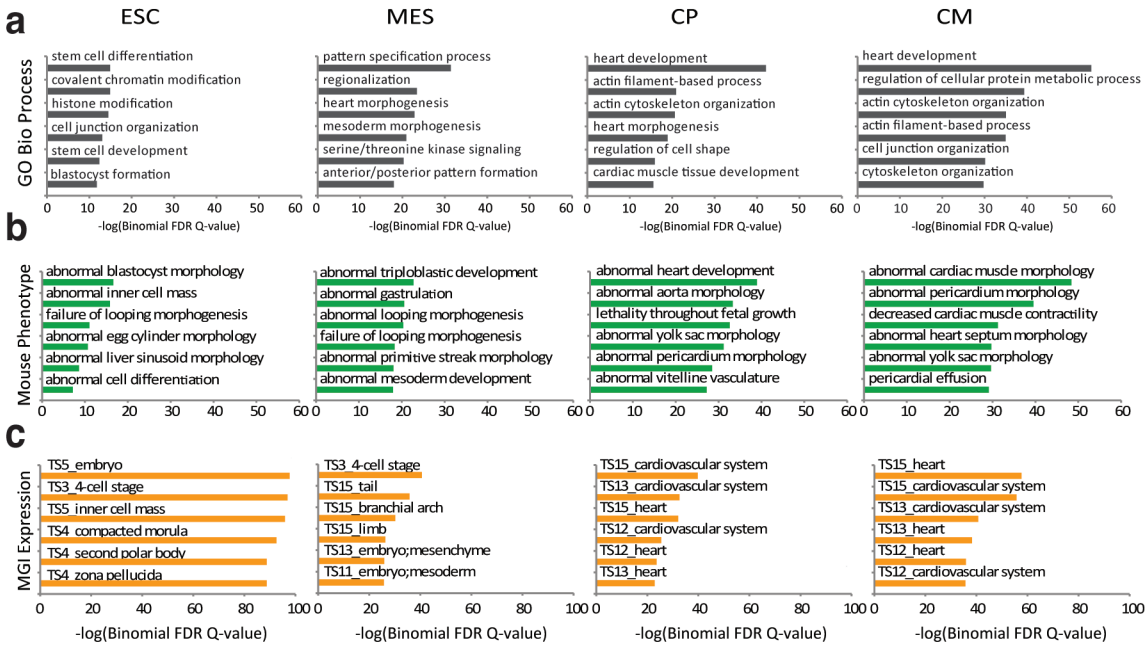


Figure 2.20. Active enhancer elements strongly associate with genes characteristic of each stage of differentiation. a-c. Log(Binomial FDR Q-value) scores for GO Biological Process (a), Mouse Phenotype (b), and MGI Expression (c) terms enriched in genes associated with active enhancers as determined by single nearest gene algorithm using GREAT algorithm.

2.4.7 Enhancer transitions during CM differentiation

How enhancer states change during cell-fate transitions is unknown, yet this is key to dissecting mechanisms that control the balance between cell identity and lineage commitment. To dissect enhancer state transitions that govern cardiac gene expression programs, we clustered enhancers according to their states (unmarked, poised, or active) at each stage (Figure 2.21a). The set of active and poised enhancers is largely unique at each stage indicating enhancer utilization is highly cell type-specific even between closely related cell types. A subset of enhancers showed poised-to-active state transitions concomitant with activation of transcription from the proximal gene (e.g. *Myh7 and Nkx2-5*) (Figure 2.21b). However, while many enhancers are poised, most fail to transition to an active state during cardiomyocyte differentiation. This may represent the initial poising of enhancers required to specify alternative cell lineages during early development and suggests cells retain significant plasticity during lineage commitment.

The dynamic nature and cell-type specificity of enhancer utilization suggests that transitions between poised and active enhancer states occur in a narrow temporal window. To test this, we compared the frequency of enhancer state transitions between each independent stage during cardiac differentiation. We find that the fraction of active enhancers that transition through a poised state is largest during the MES to CP and CP to CM transition (Figure 2.21c and Figure 2.22). Comparison of enhancer state transition frequency during cardiac differentiation relative to un-related cell types (Creyghton et al., 2010) further supports this trend. While less than 5% of enhancer transitions are active-to-poised or poised-to-active between ESCs and the differentiated cell types, upwards of 50% of these changes are observed between MES and CP stages during cardiac differentiation (Figure 2.22). The cell type specificity and rapid state transitions suggests

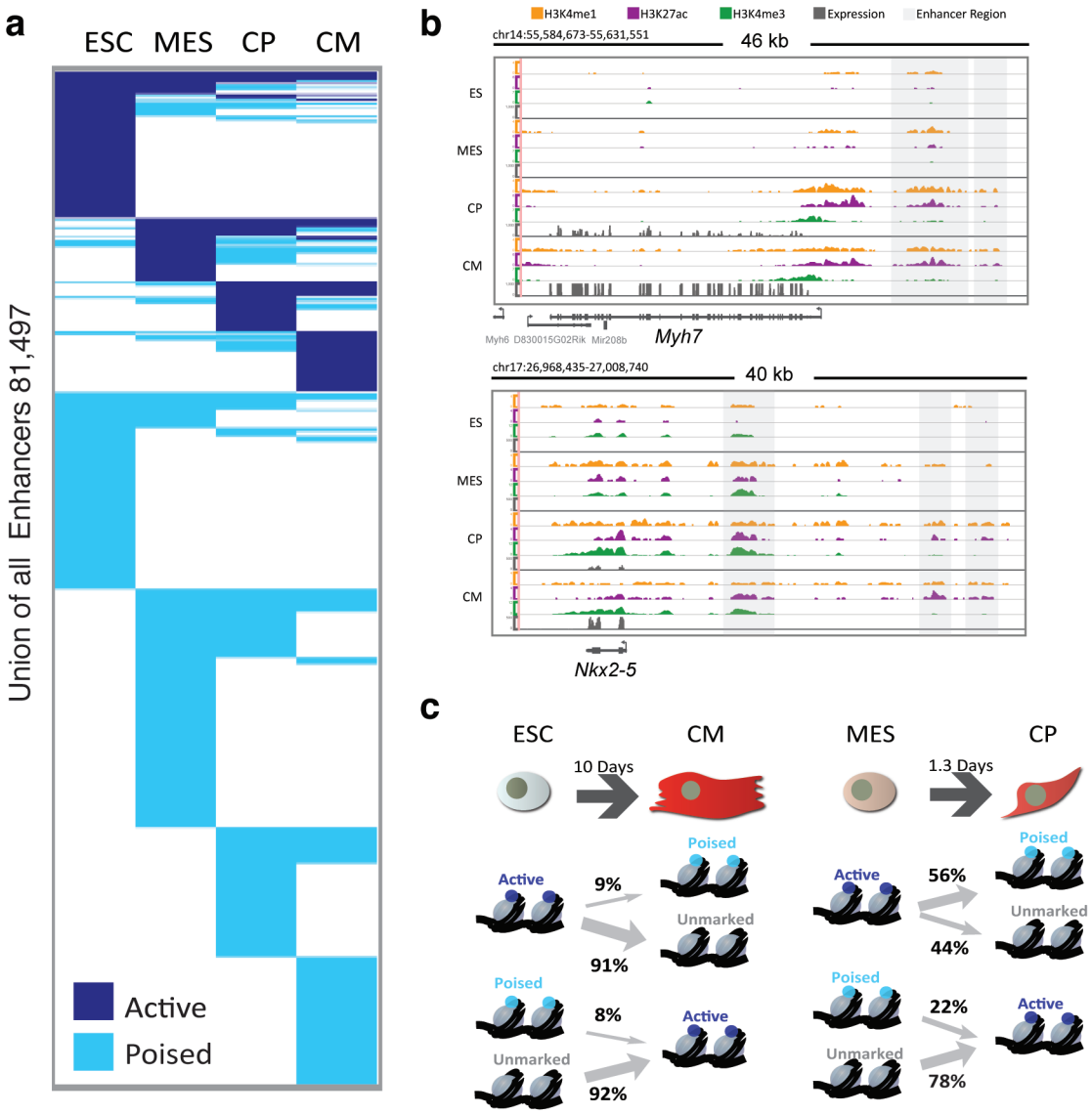


Figure 2.21. Transitioning enhancer states during cardiac differentiation. a. Union set of enhancers combined from all 4 time points during cardiomyocyte differentiation clustered based on enhancer status as Unmarked, Poised (H3K4me1+) or Active (H3K27ac+;H3K4me1+/-). b. Example ChIP-seq (H3K4me1, H3K27ac, H3K4me3, y-axis reads/million unique mapped reads) and RNA-seq (RPKM) genome tracks (mm9). Scale for each modification is constant throughout the time course. c. Percentage of poised-to-active and active-to-poised enhancer transitions during cardiac differentiation.

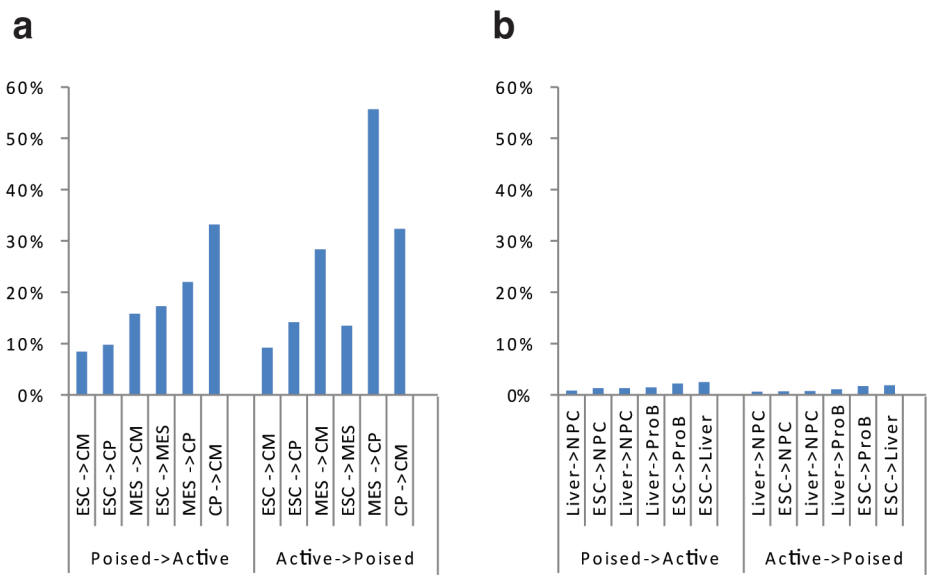


Figure 2.22. Active enhancer elements display rapid chromatin state transitions over the course of cardiac differentiation. a. Poised-to-active enhancer transitions occur most frequently between CP and CM stages, representing greater than 30% of transitions observed. Similarly, active-to-poised enhancer transitions occur most frequently between the MES and CP stages, representing greater than 50% of transitions observed. b. Comparison of active-to-poised and poised-to-active enhancer state transition frequency between embryonic stem cells (ESC), neural precursor cells (NPC), liver, and pro B cells (data from Creyghton et. al., 2010) further supports that the poised transition state occurs less often between independent cell types.

that this is an important mechanism to maintain tight control over tissue specific gene expression patterns.

2.4.8 Integrating enhancers into gene networks

While enhancers contribute to the regulation of global developmental gene expression patterns, a major challenge in the field has been integrating enhancers into the core transcriptional regulatory circuitry. Transcription factors (TFs) control gene expression programs by binding to specific motifs within *cis*-regulatory elements. Given the stage-specific expression of TFs in our time course (Figure 2.23a), we hypothesized that motifs for TFs that drive cardiac development would be enriched in active enhancers. Reasoning that TFs bind open chromatin regions (He et al., 2010; Verzi et al., 2010), we developed an algorithm to find depressions in the H3K27ac chromatin profile at active enhancer regions.

Among 124 highly conserved motifs in the TRANSFAC database, we found over-represented motifs at each stage, including those for TFs that are regulators of the ESC state (OCT4_01, 500seq_Marson, LRH1_Q5) and cardiac development (GATA_Q6, MEF_Q6_01, MEIS1BHOXA9_02, SRF_C) (Figure 2.23b). We then compiled a list of 264 TFs that are known to bind these highly conserved motifs and determined their expression patterns. We found a strong correlation between the expression of specific TFs and motif enrichment (Figure 2.23c). To address whether these predictions represented binding events, we analysed ChIP-seq data sets for Oct4 and Sox2 in ESCs (Marson et al., 2008) and Gata4 in the HL-1 cardiomyocyte cell line (He et al., 2011). Oct4 and Sox2 bound regions substantially overlapped the OCT4_01 and SOX9_B1

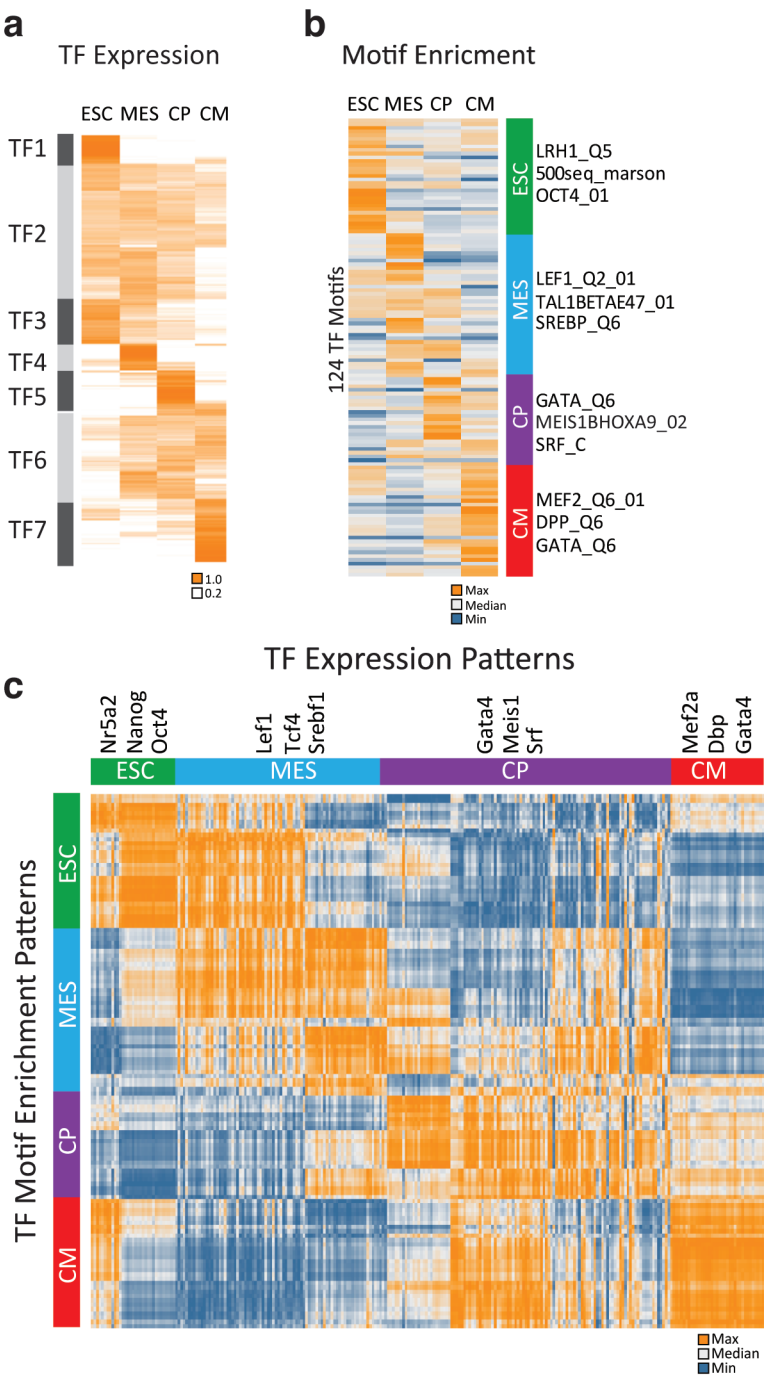


Figure 2.23. Identification of putative regulators of enhancer activity during cardiac differentiation. a. Hierarchical clustering of magnitude normalized RPKM values for transcription factors expressed during cardiac differentiation, subdivided into 7 groups (TF1-TF7). b. Clustering of magnitude normalized density based motif enrichment scores (-log(p-value)) shows stage specific enrichment of highly conserved transcription factor motifs. c. Pearson correlation matrix between enriched TF motifs and

the expression pattern of transcription factors known to bind the list of highly conserved motifs.

motifs at enhancers in ESCs (hypergeometric test p-values = $2e10^{-74}$ and $3e10^{-26}$) (Figure 2.24). At the CP and CM stage, Gata4 bound regions also strongly correlated with enhancers that have a GATA_Q6 motif (hypergeometric test p-value = $3e10^{-74}$ and $2e10^{-30}$ respectively).

To construct gene regulatory networks connected to specific TF-enhancer pairs, we selected target genes that displayed positive correlation (Pearson) with these pairs. Specific TF-enhancer correlations were associated with genes that function in common pathways, indicating enhancers regulate expression of specialized gene networks (Figure 2.25). For example, in ESCs, we find that genes connected to active enhancers with a *Nr5a2* motif, a TF known to substitute for *Oct4* during direct reprogramming (Heng et al., 2010), function in pathways for maintenance of stem-cell state, cell cycle, and proliferation and include genes such as *Sox2* and *CyclinE*. *Tcf4*, a target of β-catenin up-regulated in heart development (Cohen et al., 2007), correlates strongly with enhancers proximal to signaling genes, including those in the Notch and Wnt pathways (Figure 2.25). At the CP stage, *Meis1*, *Gata4*, and *Nfat4c* enhancer-associated genes comprise networks implicated in cardiovascular development and function (Figure 2.25 and Figure 2.26). While *Meis1* and *Meis2* have been implicated in heart development (Crowley et al., 2010; Pfeufer et al., 2010; Stankunas et al., 2008), their targets in cardiac differentiation are unknown. Our data suggest that *Meis1/2* may regulate a subset of genes that are important regulators of cardiac morphogenesis, including *Fos* and *NFAT* (Chang et al., 2004; King et al., 2011). Moreover, *Gata4* correlates with enhancers that may regulate important genes in cardiac development and disease, including *Nkx2-5*, *Mef2c*, and *Gata4* itself (Bruneau, 2008; Srivastava, 2006).

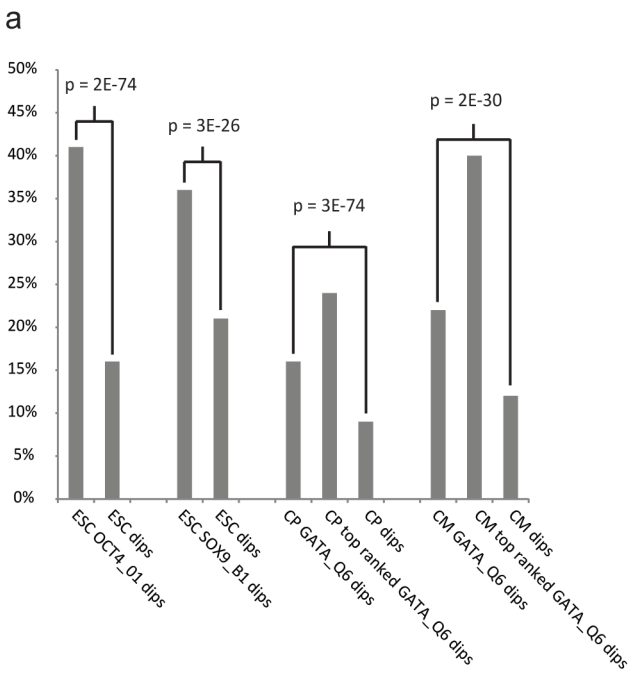


Figure 2.24. Predicted transcription factor binding events significantly overlap with genome-wide TF occupancy. Active enhancer identified dips, containing OCT4_01 and SOX9_B1 motifs, display increased overlap (0 bp tolerance) with *Oct4* and *Sox2* bound regions, respectively, as compared to total dips in active enhancers in ES cells (hypergeometric p-value = 2x10-74 and 3x10-26). Active enhancer identified dips, containing the GATA_Q6 motif, also display increased overlap (50 bp tolerance) with *Gata4* bound regions (hypergeometric p-value = 2x10-30) as compared to total dips in active enhancers in CP and CM.

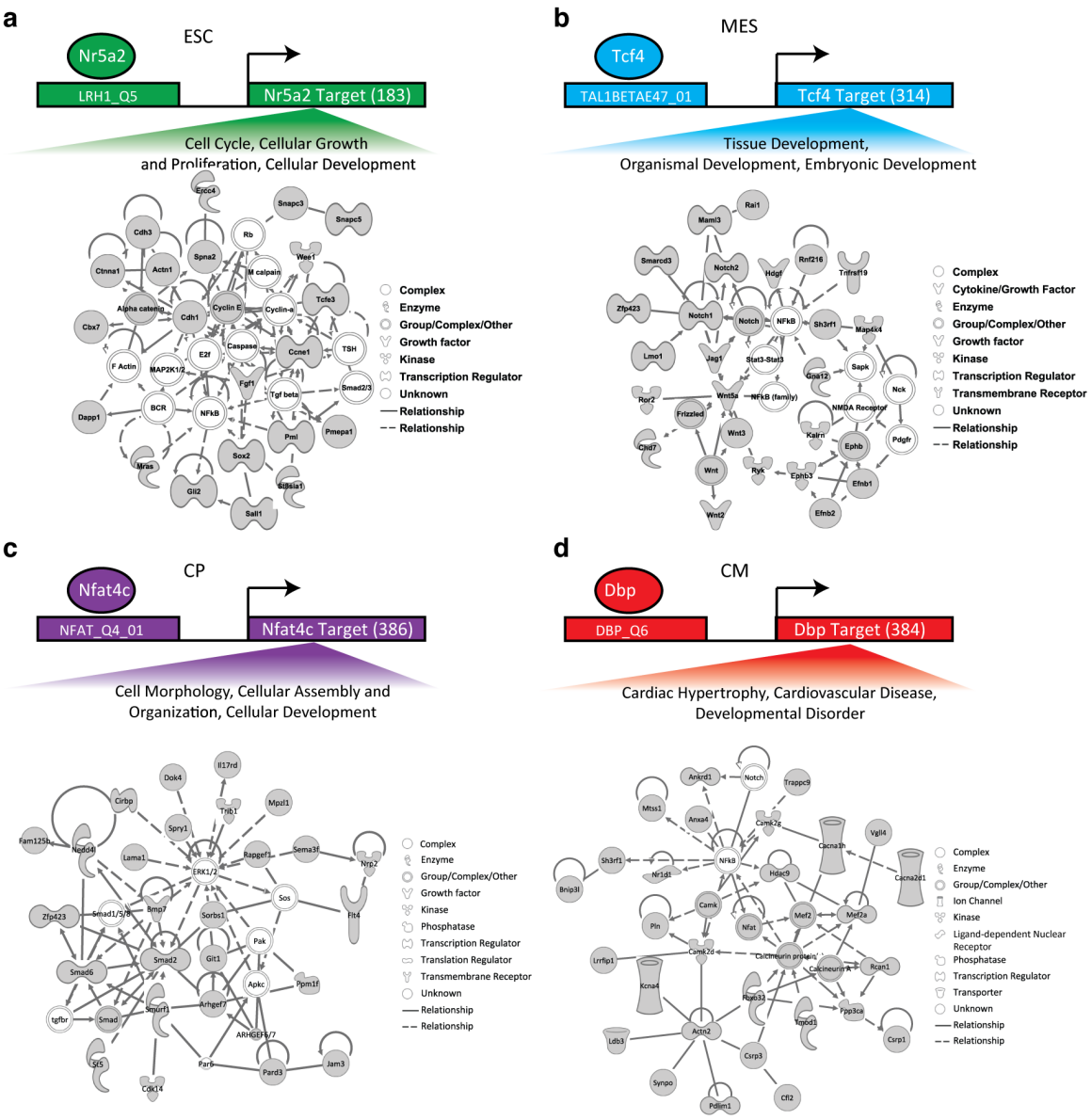


Figure 2.25. Transcription Factor/Motif identified target gene networks. Example target gene networks identified via active enhancers containing H3K27ac dips with indicated motifs at each stage during cardiac differentiation. Networks were generated using Ingenuity Pathway Analysis. Grey nodes indicate genes present in the input list used for network generation. a. LRH1_Q5, ESC. b. TAL1BETA47_01, MES. c. NFAT_Q4_01, CP. d. DBP_Q6, CM.

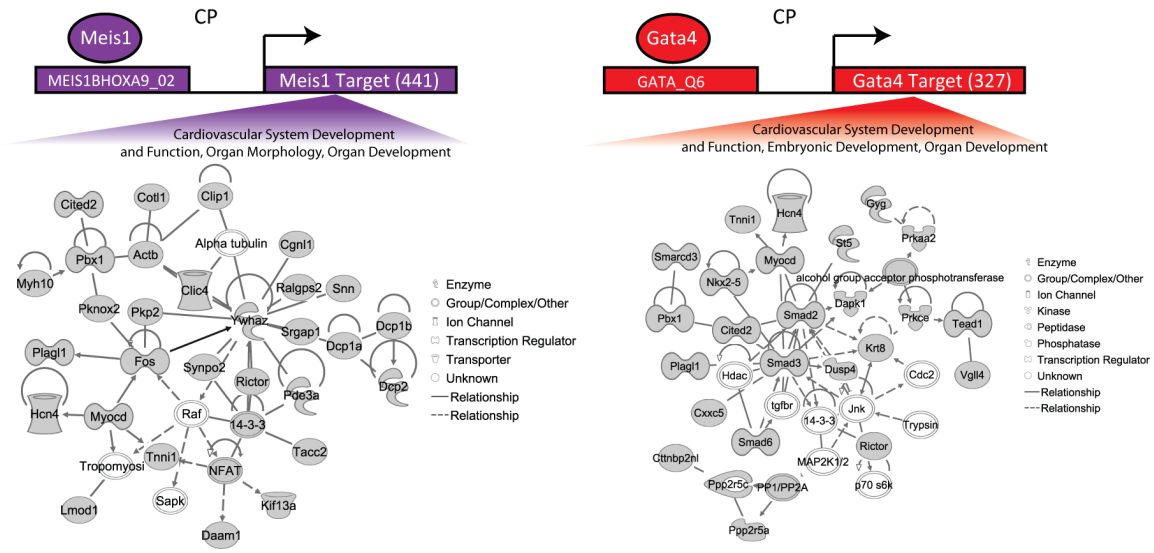


Figure 2.26. Putative enhancer gene networks in heart development. Examples of predicted target gene networks. Grey nodes represent genes identified via motif enrichment analysis.

We tested our ability to predict networks by analyzing the effects of loss of function of particular TFs on gene expression. *Oct4*-predicted motifs were highly correlated with genes affected by *Oct4* knockdown ($P=3e^{-44}$) (Loh et al., 2006). Despite considerable redundancy among GATA factors (Zhao et al., 2008), we also found a significant correlation between GATA motifs and the regulation of associated genes ($P=9e^{-12}$ at CP stage, $P=9e^{-15}$ at CM stage) by comparison with knockdown data in HL-1 cells (He et al., 2011). Collectively, these data show that we can reliably identify key TFs at enhancers that may regulate specialized gene expression networks during cardiomyocyte differentiation.

Combinatorial interactions among transcription factors can increase the diversity of regulatory modules governed by a particular factor. We observed a striking overlap among target genes associated with enhancers containing MEIS1BHOXA9_2 and GATA_Q6 motifs at the CP stage, suggesting co-regulation by Meis/Hox and GATA factors (Figure 2.27). We found distinct groups of developmentally important genes regulated by either GATA or Meis, or Meis/GATA together. For example, the GATA-only group contained several genes associated with sarcomere function, such as *Myl2*, *Myl7*, *Slc8a1*, and *Ryr2*, as well as well-known GATA targets like *Nkx2-5*. MEIS-only targets include *Irx3,* which functions in regulating the conduction system (Zhang et al., 2011), consistent with an association of *Meis1* with conduction parameters (Pfeufer et al., 2010). Enhancers co-enriched for MEIS/GATA motifs are associated with genes important for cardiac development, such as *Gata5*, *Irx4*, *Myocd*, and *Smarcd3*, and with genes known to influence conduction system function (*Hcn4*) and morphogenetic events that are disrupted in *Meis1*-deficient hearts (*Zfpm2*, *Wnt2*). Notably, we find MEIS/GATA motifs are more often co-enriched in the same enhancer (Figure 2.27b) suggesting a

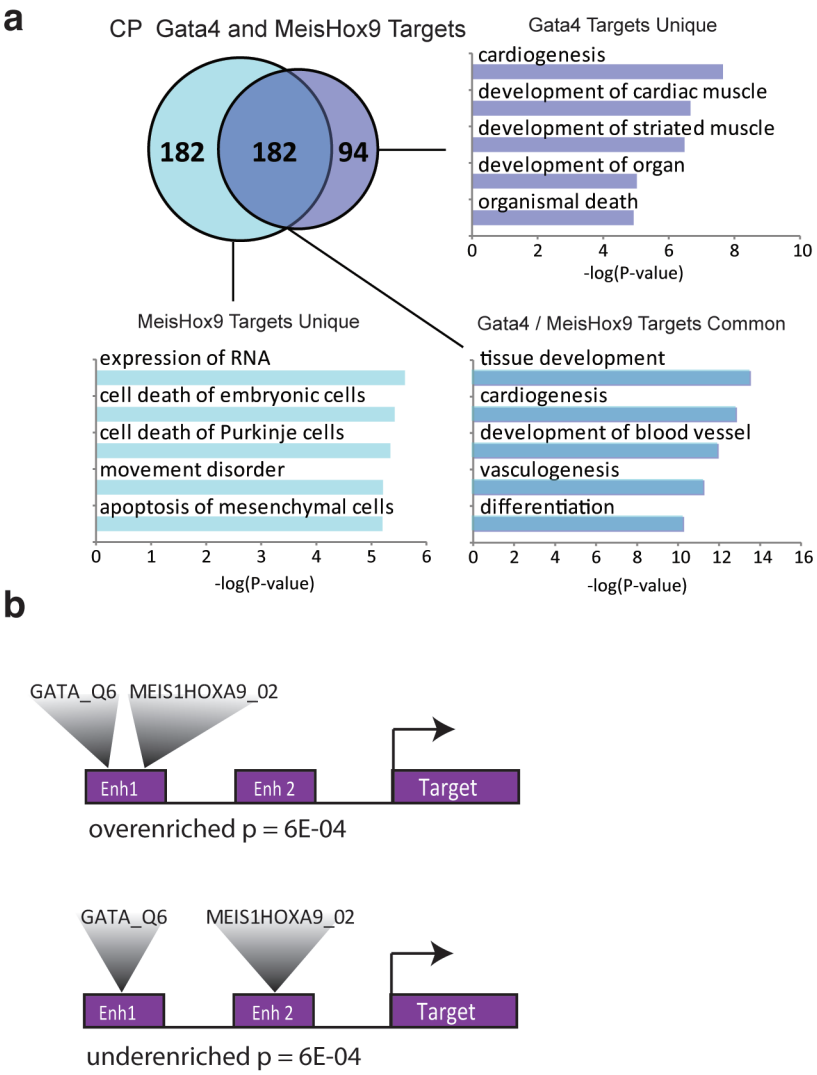


Figure 2.27. GATA and Meis factors are predicted to coregulate many enhancers during cardiac differentiation. a. Venn diagram shows overlap between MEIS1BHOXA9_2 and GATA_Q6 motif containing target genes, with associated GO terms for unique and common targets. b. Graphical representation of the preference for MEIS1BHOXA9_2 and GATA_Q6 motifs to occupy the same enhancer versus separate enhancers at common gene targets.

functional relationship between these two factors.

We further tested the potential co-regulatory function of Meis and GATA factors by luciferase reporter activation assays. We tested 5 enhancers with motifs for MEIS and GATA factors, including a *Myocd* enhancer that is active in the developing heart (Creemers et al., 2006). Co-transfection of the Myocd reporter with combinations of expression constructs for GATA4, MEIS1A, or MEIS2D, showed that this enhancer responded to GATA4 and importantly was synergistically activated by the combination of GATA4 and MEIS1A (Figure 2.28*)*. Interestingly, this appeared to be specific to MEIS1A, as co-transfection of MEIS2D with GATA4 did not lead to synergistic activation. Furthermore, most (4 of 5) enhancers tested were synergistically activated by the combination of MEIS1A and GATA4 (Figure 2.28). Thus GATA4 and MEIS1 function together to activate certain cardiac enhancers. Collectively, our work reveals a detailed picture of how gene expression programs may be coordinated during lineage commitment and provides novel insights into the key principles that underpin heart development and disease.

2.5 Discussion

We have defined chromatin state transitions during cardiac differentiation that provide new insights into the dynamic regulation of cellular differentiation and the coordinated regulation of gene expression programs. Our results show that there are complex but distinct chromatin patterns that accompany lineage decisions.

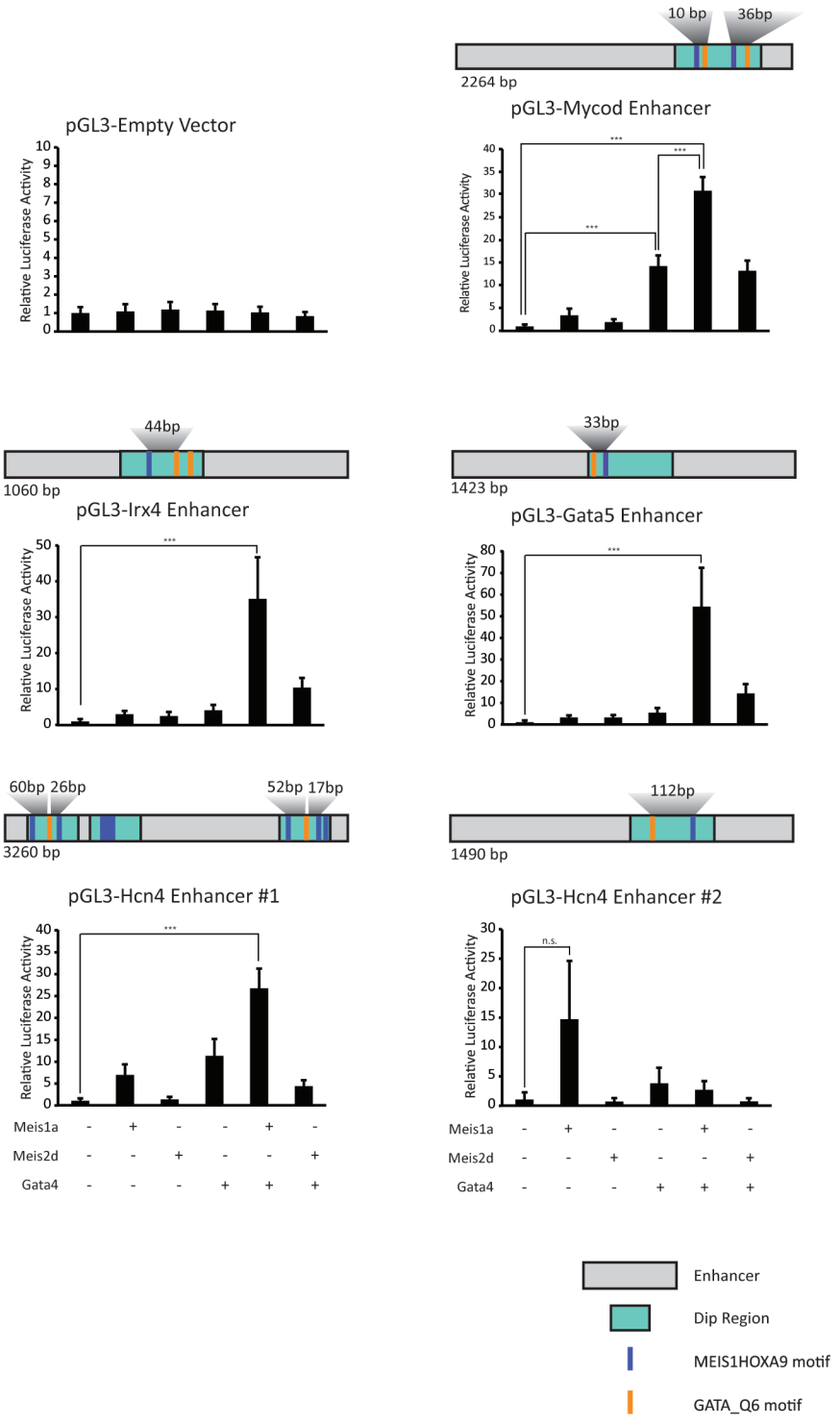


Figure 2.28. MEIS1A and GATA4 synergistically activate multiple putative target enhancer elements in a luciferase reporter activation assay. A graphical

representation of the candidate MEIS and GATA sites within the enhancer dip are shown. Distance between MEIS and GATA sites is shown in base pairs (bp). The graph shows relative luciferase activity over reporter construct alone. All enhancers except Hcn4 Enhancer #1 show synergistic activation by MEIS1A and GATA4 N=4; *** , $P<0.01$; n.s. , not significant.

2.5.1 Dynamic epigenetic transitions in differentiation

The rapid loss of expression of pluripotency-associated genes upon differentiation can be achieved by at least 9 different chromatin patterns. These patterns comprise broad groups that include loss of active marks (e.g. *Nanog*), or gradual loss of active marks with the simultaneous acquisition of repressive marks (e.g. *Oct4*). Conversely, during cardiac differentiation, we observed striking coherence among most mesoderm-specific genes, which share a specific chromatin pattern. On the other hand, genes expressed in later development can be classified by multiple distinct chromatin regulatory patterns that may precisely coordinate precursor and differentiated cardiomyocyte gene expression programs.

Our data also reveal that chromatin patterns can predict sets of functionally related genes. For example, genes associated with metabolic function share a similar chromatin pattern, while those involved in contractile function and sarcomere structure have a distinct pattern, although they share a similar expression profile. This implies that functionally-related co-expressed genes have specific modes of regulation. Distinct modes of epigenetic regulation may exist to ensure that functional gene modules are synchronized, thus ensuring robust and coordinated expression of key processes that may be critical for cell function and important for adaptation to stress.

2.5.2 A novel dynamic pattern of histone modifications

Analysis of chromatin states during cardiomyocyte differentiation has led to the identification of novel patterns that are highly informative to understand developmental regulatory programs. In particular, we have identified a pattern of H3K4me1 deposition at

the TSS that precedes the transcriptional activation and acquisition of H3K4me3 and recruitment of RNAP. This pre-activation pattern is consistent with the idea that molecular events in early lineage commitment mark genes for subsequent activation. This pre-activation pattern is likely important for genes that are not regulated by the concerted action of Polycomb and Trithorax. The presence of H3K4me1 in anticipation of transcriptional activation is reminiscent of its presence at poised enhancers, for which only a minority shows H3K27me3 enrichment. This suggests that diverse mechanisms can poise specific classes of TSSs and enhancers for subsequent activation.

2.5.3 Identification of transcriptional networks based on enhancer predictions

Chromatin marks at genomic regions distal to the TSS allows the identification of candidate enhancers (Creyghton et al., 2010; Rada-Iglesias et al., 2011). We have identified a large number of enhancers that show stage-specific activation. This rich data set has allowed us to discover transcription factor motifs that predict novel enhancer-driven transcriptional regulatory networks during cardiomyocyte differentiation.

Furthermore, we discovered potential new regulators of cardiac development. We identified enrichment for the MEIS1BHOXA9_02 motif, which predicts the binding of Meis factors, along with a partner Hox factor, at a subset of enhancers at the CP stage. *Meis1* has been implicated in heart development because *Meis1* null mice display congenital heart defects (Stankunas et al., 2008) and because *MEIS1* has been identified in GWAS studies of human arrhythmias (Pfeufer et al., 2010). Chromosomal deletions that include *MEIS2* have also been identified in patients that exhibit cardiac defects (Crowley et al., 2010). Our expression data show that both *Meis1* and *Meis2* are

robustly and transiently activated at the CP stage, consistent with a role in cardiac progenitors. Further, we find that only *Hoxb1*, *Hoxb2*, and *Hoxb3* are significantly expressed in cardiac differentiation, consistent with the expression of *Hoxb1* in a posterior subset of cells in the SHF that later contribute to the atria and outflow tract (Bertrand et al., 2011). Thus, we have identified a novel network potentially under control of Meis/Hox factors. Finally, we observed a striking overlap between GATA- and the MEISHOXA9-binding sites at enhancers, indicating that we have uncovered a previously unknown functional relationship between GATA and Meis/Hox TFs in heart development, reminiscent of their co-occurrence at genes during hematopoietic development (Wilson et al., 2010a; Wilson et al., 2010b).

Together, our study establishes a platform to understand the process of cardiomyocyte differentiation and provides an opportunity to identify mechanisms of complex disease loci by comparison with Genome Wide Association Studies (GWAS). Moreover, our data lay the foundation for understanding how the epigenetic landscape of cardiac differentiation integrates transcriptional inputs during normal development. These insights will be valuable to develop improved cardiac reprogramming strategies (Ieda et al., 2010; Takeuchi and Bruneau, 2009) and to elucidate how disruption of these diverse regulatory modules contributes to congenital heart disease.

2.6 Materials and methods

2.6.1 Cardiomyocyte differentiation

E14 Tg(Nkx2-5-EmGFP) mouse ES cells (Hsiao et al., 2008) were cultured in feeder-free conditions using standard techniques. For directed differentiations (Kattman et al.,

2011), mouse ES cells were aggregated into embryoid bodies (EB) and cultured at 75,000 cells/ml for two days in serum free media (3 parts IMDM (Cellgro #15-016-CV): 1 part Ham's F12 (Cellgro #10-080-CV), 0.05% BSA, 2 mM GlutaMax (Gibco), B27 supplement (Gibco #12587010), N2 supplement (Gibco #17502048)) supplemented with 50 ug/ml ascorbic acid and 4.5 x 10^{-4} M monothioglycerol. Embryoid bodies were dissociated and reaggregated for 40 hours in the presence of 5 ng/mL human VEGF (R&D #293-VE) and human Activin A (R&D #338-AC) and human BMP4 (R&D #314-BP) at concentrations empirically determined depending on lot. EBs were dissociated and plated at 470,000 cells/cm^2 in StemPro-34 (Gibco #10639011) supplemented with 5 ng/mL VEGF, 10 ng/mL human basic FGF (R&D #233-FB) and 25 ng/mL FGF10 (R&D #345-FG). 1-2 differentiations were pooled to generate independent replicates of 5 x 10^6 cells for RNA-seq and 4-8 x 10^7 cells for ChIP-seq at the following stages of differentiation: ESC stage – Day 0; MES stage – 40 hrs after Activin A, BMP4, and VEGF treatment; CP stage – 32 hrs after replating in StemPro-34; CM stage – Day 10 (Figure 2.1, Figure 2.7). Each differentiation was assessed after ten days for cardiomyocyte yield using troponin T intracellular FACS.

Differentiation efficacy and yield was variable for directed cardiomyocyte differentiations in our hands. We found it absolutely essential to determine the optimal concentration of Activin A and BMP4 in order to generate highly enriched cultures of cardiomyocytes. Activin A concentrations of 5-10 ng/mL typically yielded successful differentiations in our hands, but cultures were highly sensitive to BMP4 concentration. We found between 0.1-0.8 ng/mL BMP4 was generally optimal for cardiomyocyte differentiation, but deviation from the optimal dose by as little as 0.1 ng/mL was enough to significantly reduce cardiomyocyte yield in some cases. In general, we standardized

passage number, ES cell culture, and growth factor lots to ensure reproducibility. If we observe reduced cardiomyocyte yields, change to a new lot of Activin A or BMP4, or are using a new cell line, we retitrate growth factor concentrations to ensure maximal differentiation efficiency.

2.6.2 Immunofluorescence microscopy

Immunofluorescence staining was performed on differentiations cultured in 8 chamber slides. Cultures were fixed for 30 minutes at room temperature in 3.7% formaldehyde D-PBS, washed with D-PBS, blocked, and incubated with primary antibody. Slides were washed three times with 0.1% Triton X-100 D-PBS and incubated in secondary antibody. After staining, slides were washed three times, stained with Hoechst 33342 (10 ug/mL), and mounted with either polyvinylalcohol or Vectashield (Vector labs). Antibodies were Anti-cardiac isoform of Troponin T (cTnT) 1:100, anti-GFP 1:2000, anti-Isl1 1:100

2.6.3 Flow cytometry

For FLK-1/PDGFRα staining, EBs were trypsinized, quenched with serum, and washed four times with 4% FBS, D-PBS. Cells were then stained with a biotinylated anti-FLK-1 antibody for 30 minutes at 4°C (1:100 dilution). After incubation, cells were washed three times with 4% FBS D-PBS and then stained with an PE-conjugated anti-PDGFRα (1:400 dilution) and PE-Streptavidin (1:200 dilution) for 30 minutes at 4°C. Cells were washed an additional three times and then analyzed on an LSRII flow cytometer (BD).

Data was analyzed using FlowJo software (Treestar).

For intracellular FACS, cultures were trypsinized, quenched with serum, and fixed in D-PBS with 3.7% formaldehyde for 30 minutes at room temperature. Fixed samples were washed twice and stained with anti-cTnT for 30 minutes at room temperature. After staining, samples were washed twice, incubated with secondary antibody, and washed two additional times. All steps were performed in D-PBS with 0.5% saponin and 4% FBS. Samples were stained with Hoechst 33342 (10 ug/mL) in D-PBS with 4% FBS. Samples were analyzed on an LSRII flow cytometer (BD). Data was analyzed using FlowJo software (Treestar).

2.6.4 Quantitative PCR

RNA was extracted using TRIzol® and 650 ng RNA was reverse transcribed using High-Capacity cDNA Reverse Transcription kit (Applied). Quantitative PCR was performed using Taqman probes and expression was normalized to Gapdh. All reactions were performed in triplicate. The following probes were used: *Oct4* – Mm00658129_gH, *Nanog* – Mm02384862_g1, *Mesp1* – Mm00801883_g1, *T* – Mm00436877_m1, *Isl1* – Mm00627860_m1, *Nkx2-5* – Mm00657783_m1, *Tbx5* – Mm00803521_m1, *Actc1* – Mm01333821_m1, *Tnnt2* – Mm01290255_m1, *Myh6* – Mm00440354_m1, *Myh7* – Mm00600555_m1, *Myl7* – Mm00491655_m1, *Myl2* – Mm00440383_m1, *Gfap* – Mm01253033_m1, *Sox17* – Mm00488363_m1, *Afp* – Mm00431715_m1

2.6.5 RNA-seq

Total RNA was isolated from $5x10^6$ cells using *TRIzol®* Reagent according the manufacturer's instructions. 10 ug of total RNA was used as input material for the preparation of the RNA-seq libraries, as indicated by previous iterations of the Illumina RNA-seq protocol. Sequencing libraries were prepared according to Illumina RNA Seq library kit with minor modifications. Briefly, mRNA was isolated using Dynabeads® mRNA Purification Kit (Invitrogen) followed by fragmentation (Ambion) and ethanol precipitation. First and second strand synthesis were performed followed by end repair, A-tailing, paired end adapter ligation and size selection on a Beckman Coulter SPRI TE nucleic acid extractor. 200-400 bp dsDNA was enriched by 15 cycles of PCR with Phusion® High-Fidelity DNA Polymerase (NEB) followed by gel purification of ~250 bp fragments from the amplified material. Amplified libraries were sequenced on an Illumina GAIIx sequencer.

2.6.6 RNA-seq analysis pipeline

Paired-end RNA-seq 36 base pair reads were aligned to mm9 (*Mus musculus* assembly July 2007) using novoalign V2.07.00 (http://novocraft.com) with default options, except that no repeats were reported (-r None) and the fragment length and standard deviation were set at 176 and 11, respectively (-i PE 176,11). Raw read counts per gene were input into DESeq (Anders and Huber, 2010) for normalization and analysis of differential gene expression. Subsequent to determining which genes were differentially expressed, RPKM was used for filtering, clustering, and visualization purposes. The RNA-seq application in USeq 7.0 (Nix et al., 2008) was used to generate gene-level read counts

and estimate RPKM (reads per kilobase of exon per million reads mapped). Only genes with expression values >1 RPKM in at least one cell type were considered for subsequent analysis. Using this gene set, expression was normalized to the interquartile range across the time course to emphasize changes in gene expression levels. These interquartile numbers were used as input for clustering using a cosine angle distance metric and the Hopach clustering package (http://www.bioconductor.org/packages/2.1/bioc/html/hopach.html).

2.6.7 miRNA analysis

The same total RNA used for RNA-seq, isolated from each stage of cardiac differentiation in duplicate, was sent to Nanostring for miRNA expression analysis on the nCounter Digital Analyzer. To account for slight differences in hybridization and purification efficiency, data were background subtracted and normalized to the average count of all exogenously added controls in each sample. The housekeeping genes Actb, B2m, Gapdh and Rpl19 were used to normalize for RNA content in different samples.

2.6.8 ChIP-seq

Chromatin immunoprecipitation of histone modifications were performed according to the Young lab protocol (Lee et al., 2006) with minor modification. Briefly, frozen pellets of cross-linked cells ($10x10^6$) were thawed in cold lysis buffer 1 (50 mM HEPES-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, 1× protease inhibitors) and gently rocked at 4°C for 10 minutes in 14 mL conical tubes.

Cells were pelleted at 1350 x g at 4°C in a clinical centrifuge and resuspended in cold lysis buffer 2 (10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1× protease inhibitors) and gently rocked at 4°C for 10 minutes in 14 mL conical tubes. Cells were pelleted at 1350 x g at 4°C in a table top centrifuge and resuspended in 2 mL cold lysis buffer 3 (10 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate, 0.5% N-lauroylsarcosine, 1× protease inhibitors) and sonicated to 200-600 bp fragments using a Diagenode Bioruptor (3 x 10 minute cycles, 30 sec ON / 30 sec OFF at 4°C). Sonicated lysates were cleared by pelleting insoluble material at 20,000 x g at 4°C followed by incubation with antibody bound Protein A/G magnetic beads (2.5 ug Ab / 50uL beads / IP) in 1 mL of 0.5% BSA/PBS overnight at 4°C. Magnetic beads were washed 3 times with block (0.5% BSA/PBS), incubated for approximately 4 hrs at 4°C with antibody in block and then washed 3 times with block prior to addition of cleared cell lysates. Immunoprecipitated material was washed five times with cold wash buffer (RIPA: 50 mM HEPES-KOH, pKa 7.55, 500 mM LiCl, 1 mM EDTA, 1.0% NP-40, 0.7% Na-Deoxycholate) and one time with TE plus NaCl, followed by elution and uncrosslinking in 210 uL of 1% SDS in TE overnight at 65°C. 200 uL of uncrosslinked material was treated with RNAse A for 2 hours, proteinase K for 2 hours and extracted 2 times with phenol chloroform isoamyl alcohol, followed by ethanol precipitation with a glycogen coprecipitant, 80% ethanol wash and final resuspension in TE. Chromatin immunoprecipitation of RNAPser5 was performed as described above for histone modifications, with alterations to the sonication buffer and wash buffers as described in Rahl et al. 2010 (Rahl et al., 2010). Nucleic acid yield was determined via Quant IT fluorescence assay (Invitrogen) and histone modification enrichment was evaluated by qPCR. Illumina sequencing libraries were generated (Schmidt et al., 2009),

with minor modification. Briefly, 5-50 ng of immunoprecipitated nucleic acid was end repaired, a-tailed and ligated to Illumina single end adapters using an NEB Next kit (New England Biolabs). ~200-300 bp adapter ligated nucleic acid was gel purified and enriched via 19 cycles of PCR amplification with Phusion® High-Fidelity DNA Polymerase, followed by sequencing on an Illumina HiSeq 2000 system.

2.6.9 ChIP-seq analysis pipeline

Short reads were aligned to the mm9 mouse genome using Bowtie (2 bp mismatch). Sequences were extended +200 bp for histone marks and RNAPser5 and allocated in 25-bp bins. Biological replicate whole cell extracts were sequenced for each time point and combined for use as background model. A Poissonian model was used to determine statistically enriched bins with a P-value threshold set at $1x10^{-9}$ as described previously (Marson et al., 2008). Additionally, we required that genomic bins were at least 5 fold over input to be considered enriched peaks.

2.6.10 Analysis of chromatin marks at TSSs

Chromatin marks at promoters were evaluated by computing the ratio of ChIP to input within a 2kb region centered on the TSS of each Ensembl transcript. ChIP values were evaluated as the sum of the depth of reads of every base in the 2kb window, normalized to the total number of ChIP reads in the given replicate. Input reads taken from each cell type were not found to differ substantially. Thus, input data from ESC, MES, CP, and CM cells was pooled together and the input value for each transcript was calculated as the

sum of the depth of reads at every base in the 2kb window, normalized to the total number of input reads in the data set. Since expression data was computed at the gene level, the value of each chromatin promoter mark was defined as the highest value observed across all transcripts of a gene. Results were then pooled across replicates by using the median value across the 2-3 replicates available. The gene list was limited to those genes included in the expression analysis and that had at least one chromatin mark in one cell type. The genes were then clustered with Hopach (http://www.bioconductor.org/packages/2.1/bioc/html/hopach.html) as a function of the five chromatin marks across four stages. A spearman distance metric was calculated as one minus the spearman correlation, where the spearman correlation was calculated with the R function 'cor' and then the distance matrix was explicitly supplied to the Hopach algorithm. Other distance metrics, such as 1 minus the kendall's tau correlation and cosine angle gave similar results. A spearman distance was employed to emphasize the patterns of chromatin markings rather than the absolute values and resulted in clusters that were more easily interpreted.

In order to quantify the faction of genes which are marked with H3K4me3, H3K4me1, or H3K27me3 in various cell types a threshold was applied to the chromatin value; values above three were considered marked, values below three were considered unmarked. Applying a threshold of 2.75 or 3.25 led to comparable results. For grouping genes based on H3K4me3 and H3K4me1 pattern, H3K27me3 marked genes at any point throughout differentiation were excluded, given that H3K4me3 does not correlate well with transcriptional activity at H3K27me3 modified promoters. To assess the fraction of genes with a particular pattern of marks that were heart associated annotations from the Mouse Genome Informatics (MGI) database were used. Genes annotated as

present in the following anatomical structures were considered heart genes: aorta, aortic valve, atrio-ventricular canal, atrio-ventricular cushion tissue, atrium, atrium; cardiac muscle, bulbus cordis, cardiac muscle, cardiogenic plate, cardiovascular system, common atrial chamber, common atrial chamber; cardiac muscle, common atrial chamber; right part, endocardial cushion tissue, heart, heart atrium, heart; atrium; cardiac muscle, heart; endocardial lining, heart; endocardial tube, heart left ventricle, heart; outflow tract, heart; valve, heart ventricle, heart; ventricle, interventricular septum; cardiac muscle, left atrium; cardiac muscle, left ventricle, left ventricle; cardiac muscle, mitral valve, primitive heart tube, primitive ventricle, primitive ventricle; cardiac muscle, pulmonary artery, pulmonary valve, pulmonary vein, right atrium; cardiac muscle, right ventricle, right ventricle; cardiac muscle, tricuspid valve, ventricle, ventricle; cardiac muscle. The number of heart genes with a pattern of marks was compared to the number of genes with the same pattern that were annotated in the MGI database.

2.6.11 Evaluation of the correlation of chromatin and expression clusters

Chromatin-expression sub-clusters with a significantly different number of genes than expected were identified by computing a Pearson residual for every chromatin-expression sub-cluster. This corresponds to a χ^2 test with a Monte Carlo simulated p-value of <0.00001 (based on 100,000 replicates), indicating that the expression and chromatin clusters are not independent.

2.6.12 GO analysis of chromatin clusters

Chromatin clusters, expression clusters, and chromatin-expression sub clusters were analyzed for enrichment of gene ontology terms with GO Elite (http://www.genmapp.org/go_elite/) using the entire list of genes included in the respective clustering as the background.

2.6.13 Chromatin expression co-clusters networks

ChIP fold enrichment values and gene expression values were magnitude transformed according to the following equation, $m_t = \frac{x_t}{\sqrt{\sum_i x_i^2}}$ where *i*=ESC, MES, CP and CM. x_t is the RNA-seq RPKM value or ChIP-seq reads at a specific time point. Co-cluster correlation networks were generated using Cytoscape. In each time point module, genes are represented by nodes and Pearson correlation of chromatin marks are represented by edges (red lines). Pearson correlations are based on the magnitude transformed values with a cut off of 0.9. Node color corresponds to gene expression state, where yellow indicates up regulation and black indicates down regulation.

2.6.14 Enhancer identification

ChIP-seq for a particular histone modification was performed in biological replicate or triplicate. To identify a high confidence set of enriched regions for each antibody used, we selected intersecting regions that passed the threshold criteria from the biological

replicates and required the presence of an enriched region in 2 of 3 in the case of triplicate data sets. The total set of unique enhancer regions was generated by combining all H3K4me1 or H3K27ac enriched regions at each stage of the time course. The size of each overlapping region reflects the maximum length of the called interval in the replicate data sets. This interval list was then filtered by removing any interval that overlapped with regions +/-1 Kb of transcriptional start sites as annotated by the mm9 mouse UCSC and Ensembl genome builds. In cases where an extended region overlapped the +/-1 Kb window around a TSS, the region was trimmed +/-5 Kb of the TSS, leaving the distal portion in either direction as independent enhancer regions. The set of enhancer regions were then annotated as active (H3K27ac+, H3K4me1+/-), poised (H3K27ac -, H3K4me1+) and unmarked (H3K27ac -, H3K4me1-).

2.6.15 Enhancer analysis

Comparison to enhancers identified via p300 and multiple transcription factors (MTL) was performed by extending published genomic intervals (Blow et al., 2010, He et al., 2011) by 1000bp on each side and overlapping these extended intervals with our total enhancer list. The published genomic intervals were extended because shorter intervals identified using transcription factor binding or p300 can reside between peaks of H3K4me1 and H3K27ac regions. Enhancers annotated as active or poised for a specific stage were clustered by that stage's H3K27ac signal followed by H3K4me1 signal. A region +/- 4kb of each enhancer's center was then selected to find the read density (to a resolution of 100bp) for the H3K27ac, H3K4me1, H3K4me3 and RNAPser5 ChIPs at each stage, and the resulting cluster data table was visualized using Java TreeView. To

determine changes in global gene expression between genes associated with active, poised and unmarked enhancers, a gene list was generated consisting of genes that were the single nearest of some enhancer. Genes in this list were annotated as 'active', 'poised' or 'unmarked' at each stage according to the state of their associated enhancer. If a gene was associated with multiple enhancers, enhancers were given annotation priority in the order active>poised>unmarked. The RPKM values of these genes were ranked, and each RPKM was replaced with its percentile, calculated as 100*(1 – [gene's RPKM rank]/[# of genes in list]). Boxplots of RPKM percentile values for genes annotated as 'active', 'poised' or 'unmarked' were then generated in R, and the difference in RPKM values between groups was tested for significance using the wilcoxon rank sum test with continuity correction in R. Gene ontology scores were generated using GREAT single nearest gene setting with a whole genome background (McLean et al., 2010). FDR corrected binomial distribution scores are used in figures and scores were required to pass both the GREAT default binomial distribution and hypergeometric statistical tests. Genome tracks were generated using the Integrative genomics viewer (IGV) and Adobe Illustrator.

2.6.16 Conservation of cardiac enhancers

Enhancer conservation was determined by overlapping vertebrate PhastCons elements with active enhancer regions at each time point.

2.6.17 Dip finding algorithm

To identify H3K27ac depressions or “dips” in active enhancers, the summed reads per million (RPM) of the H3K27ac ChIP replicates were used as input for each stage at a resolution of 25 bp. All active enhancers on a given day were scanned for dips. If enhancers were within 1kb of each other, the enhancers were merged and the combined “enhancer regions”, including the intervening bases, were scanned for dips. The algorithm conceptually works by 'walking' along the chromatin profile and testing whether the parameters that define a 'dip' have been satisfied. The parameters are as follows:

Slope length

Dips consist of a downward slope to the left of the minimum and an upward slope to the right of the minimum. If the slope length is set to N, then all passing dips must have a steady downward slope at least N bins long and a steady upward slope at least N bins long. A "steady" downward slope only requires that once a downward slope has been detected, an upward slope must not be detected for at least N bins; the slope profile is allowed to level-off temporarily. Similarly, a "steady" upward slope requires that once an upward slope has been detected, a downward slope must not be detected for at least N bins. A slope length of 5 was used.

Dip size filter

The dip size was measured as the height from the minimum of the dip to the lowest height at which the dip is interrupted. The dip is interrupted where the downward slope to the left of the minimum or the upward slope to the right of the minimum ceases to exist. A minimum dip size of 0.8 was required.

Returned dip region height

The region returned as the location of the dip was a region centered around the minimum of the dip such that the height from the minimum to the height at both ends of the returned region was at least 0.3. If this region was longer than 200bp, a 200bp region centered at the minimum was returned instead.

Zero Count Limit

If in the process of searching for a dip, a stretch at least 250bp long (10 bins) for which the chromatin profile was zero was detected, the dip did not pass.

Dip ranking

The dips were ranked by the height of the minimum of the dip, with the lowest minima ranking first. The total number of dips found in a given stage roughly correlated with the total number of active enhancers from that stage. The top 3500 dips of each stage were then selected for differential analysis between the stages. Selecting a fixed number of dips from each stage prevents biases in the p-value resulting from certain stages having more dips (the larger the input data size, the stronger the p-values).

2.6.18 Identification of enriched motifs

Explain from BioBase (BioBase Biological Databases, Beverly, MA) was used to identify motifs in the dips from each stage. The profile of motif position-weight-matrices used to scan the dips was "vertebrate non-redundant", and only high specificity matrices were

used. This amounted to a set of 196 position-weight-matrices. Motif search was performed using the Match algorithm (default cutoff p-value 0.5). The background set used for each stage consisted of randomized sequences having the same length distribution and base distribution as the input set for that stage. A hypergeometric test was used to determine if the overlap observed between active enhancer dips near published ChIP-seq identified transcription factor bound regions and active enhancer dips containing an enriched motif was significant relative to chance. The top 3500 dips was used as the universal set of regions for the test.

2.6.19 Transcription factor expression analysis

To determine transcription factor expression clusters over the time course, transcription factors expressed at values of >1 RPKM in at least one cell type (973 of 1675, as annotated by the Riken Center (Kanamori et. al., 2004) were assigned a pseudo count of one. These values were then magnitude transformed and clustered via complete linkage using a cosine angle correlation distance metric in Spotfire (TIBCO Spotfire, Somerville, MA).

2.6.20 Transcription factor and motif enrichment correlation

To determine the correlation between transcription factor expression and motif enrichment, Pearson correlation analysis was performed on magnitude transformed values. The transformation was performed on the RPKM values of the 264 transcription

factors known to bind the 124 enriched motifs, as annotated by Explain (BioBase Biological Databases,Beverly, MA), and the signed –log10 of the p-value for motifs (124/196) over and under-enriched relative to the background in a least one stage of the time course. The correlation matrix was visualized by clustering the expression data as described above against the enriched motifs clustered via complete linkage of the Pearson correlation values using a squared Euclidean distance metric in Spotfire (TIBCO Spotfire, Somerville, MA). Ordering of the enriched motifs from the correlation matrix was maintained for the independent visualization of the motif enrichment over the time course.

2.6.21 Gene network analysis

Top candidate enriched motifs displaying high correlation with expression of a transcription factor known to bind the motif (Pearson correlation >0.9, FDR < 0.25) were selected for further analysis. The nearest gene associated with the genomic interval, corresponding to the top 3500 H3K27ac dips containing an enriched motif, were identified. Target genes displaying a positive expression correlation (Pearson) with motif associated TF expression were used to compile a non-redundant target gene list for discovery of target gene networks using Ingenuity Pathway Analysis (IPA, Ingenuity Systems, Inc., Redwood City, CA). Direct and indirect node connections were allowed in network generation. Example networks displayed in Figure 2.25 and Figure 2.26 received an Ingenuity network score of equal to or greater than 38. Overlap and GO analysis of MEIS1BHOXA9_02 and GATA_Q6 networks was performed in IPA. A hypergeometric test was used to determine the significance of MEIS1BHOXA9_02 and

GATA_Q6 motifs preference for residing in the same enhancer at common target genes. For this test, the universal set consisted of all dips associated with the same gene as a dip containing the GATA_Q6 motif; Set 1 consisted of all dips that shared an enhancer region with a dip that contains the GATA_Q6 motif; Set 2 consisted of all dips that both contained a MEIS1BHOXA9_02 motif and were associated with the same gene as a dip containing GATA_Q6. The p-value indicates the significance of the overlap of Set 1 and Set 2 (where the overlap represents MEIS1BHOXA9_02 and GATA_Q6 being present in the same enhancer region at a shared target gene versus being in different enhancer regions at a shared target gene; enhancer regions are defined in the explanation of the dip-finding algorithm).

Chapter 3

Brg1 Is Required for Repression of Developmental Regulators by Polycomb Repressive Complexes in Mesoderm Differentiation

3.1 Contributions

Studies described in this chapter were performed by the author. Data for Figure 3.2 was generated with support from John Wylie. $Brg1^{fl/fl}$; Actin-CreER mES cells were made by Dr. Lena Ho and are a kind gift from Dr. Gerald Crabtree.

3.2 Abstract

Cell differentiation requires large-scale changes in chromatin structure that modulate gene expression. Cells employ numerous factors to regulate chromatin structure, including chromatin remodeling complexes. However, we still do not understand the many essential roles for chromatin remodelers in development and how they might impact broad chromatin states during this process. Here, we investigated the function of the chromatin remodeling factor, *Brg1*, during cardiomyocyte differentiation of embryonic stem cells. We find that while *Brg1* is dispensable in cardiomyocytes, it is essential for cardiomyocyte differentiation. Upon further investigation, we found loss of *Brg1* leads to misregulation of gene expression during mesoderm differentiation. These genes included a large number of critical developmental regulators, a class of gene known to

be regulated by Polycomb repressive complexes. Interestingly, *Brg1*-dependent genes were enriched for H3K27me3 and known Polycomb targets. Furthermore, genes upregulated by loss of *Brg1* demonstrated reduced levels of H3K27me3 in *Brg1*-depleted cultures. Collectively, these data point to a novel and essential function for *Brg1* in cardiac differentiation and suggest that Polycomb-mediated silencing requires *Brg1* during mesoderm induction.

3.3 Introduction

Embryonic development necessitates precise spatial and temporal patterns of gene expression to produce a viable organism. This is achieved through a complex network of DNA-binding developmental regulators and chromatin-modifying machinery that must activate appropriate gene expression while keeping inappropriate genes off. This is achieved in part by ATP-dependent chromatin remodeling complexes, which mediate chromatin accessibility by unwinding or displacing nucleosomes. *Brg1*/*Brahma*-associated factor complexes represent an essential chromatin remodeler in mammals (Ho and Crabtree, 2010). Composed of 10-12 subunits organized in a ~ 1 MDa complex, BAF complexes are thought to exist in all mammalian cell types and have been implicated in a broad spectrum of biological processes, including DNA repair (Zhao et al., 2009, Lee et al., 2010), HIV transcription (Agbottah et al., 2006, Ariumi et al., 2006, Mahmoudi et al., 2006, Treand et al., 2006), immune response (Chi et al., 2004), and cancer (Reisman et al., 2009).

BAF complexes contain either BRM or BRG1 as the enzymatic subunit; however, only *Brg1* is essential for embryonic development arguing that most developmental

processes utilize BRG1-containing BAF complexes (Bultman et al., 2000, Reyes et al., 1998). Tissue-specific inactivation of *Brg1* has demonstrated required roles in skin differentiation (Indra et al., 2005), erythropoiesis (Griffin et al., 2008, Stankunas et al., 2008a), T cell maturation (Gebuhr et al., 2003), neurogenesis (Lessard et al., 2007), and gametogenesis (Kim et al., 2012). Furthermore, multiple studies have demonstrated an important function for *Brg1* in the cardiovascular lineage (Stankunas et al., 2008, Hang et al., 2010, Takeuchi et al., 2011). Interestingly, a small number of genes appear to be *Brg1*-dependent in the developing heart, suggesting that BRG1-containing complexes play highly specialized roles in differentiating cardiovascular cell types. However, the function of *Brg1* in early stages of cardiac differentiation has not been investigated.

In contrast, Polycomb repressive complexes regulate gene expression by depositing repressive post-translational histone modifications and reinforcing compact, silent chromatin (Surface et al., 2010). PRC2 functions through catalysis of H3K27 methylation (H3K27me), which can silence genes through recruitment of additional repressive complexes and by antagonizing histone acetylation (Cao et al., 2002, Lee et al., 2007, Pasini et al., 2010). PRC2 targets many developmental TFs and signaling molecules, such as Hox genes, for repression and by doing so has a widespread importance in regulating cell differentiation (Boyer et al., 2006, Lee et al., 2006, Lien et al., 2011). Classic studies from *Drosophila* have suggested that BAF and Polycomb complexes have antagonistic function (Kennison et al., 1988, Tamkun et al., 1992). However, recent studies suggest a more complex relationship in mammals, raising the possibility that BAF and Polycomb complex may cooperate to regulate gene expression in some contexts (Ho et al., 2011).

Using mouse ES cells, we studied the function of *Brg1* in the differentiation of

pluripotent cells to cardiomyocytes. We find *Brg1* is essential for cardiomyocyte differentiation of mouse ES cells. While *Brg1* is required for normal expression of a few cardiac genes in differentiated cardiomyocytes, *Brg1* loss leads to widespread misregulation of developmental regulators during mesoderm differentiation. Interestingly, the majority of these genes were upregulated by *Brg1*-depletion. *Brg1*-dependent genes were enriched for Polycomb targets, including all four Hox loci, and showed enrichment for H3K27me3. Furthermore, *Brg1*-dependent upregulated genes showed reduced levels of H3K27me3 upon loss of *Brg1*, demonstrating a requirement for *Brg1* in PRC2-mediated silencing during mesoderm differentiation. Our results reveal that *Brg1* regulates expression of developmental regulators during cardiac differentiation and suggest cooperativity between BAF and PRC2 complexes in gene silencing.

3.4 Results

3.4.1 BAF complex subunits are downregulated with cardiac differentiation

In our studies of directed cardiomyocyte differentiation of mouse ES cells, we profiled global gene expression patterns using RNA sequencing (see Chapter 2). To better understand the role of chromatin regulators in this process, we analyzed our data to look for gene expression patterns of known chromatin regulators. Interestingly, many genes encoding chromatin modifying proteins had high expression levels in undifferentiated cultures, showing a trend towards downregulation with cardiomyocyte differentiation. Among these were multiple genes encoding subunits of BAF complexes, including *Smarcb1* (*Baf47*), *Smarcc1* (*Baf155*), *Smarca4* (*Brg1*), *Smarcd1* (*Baf60a*), *Smarcd2* (*Baf60b*), *Smarce1* (*Baf57*), *Actl6a* (*Baf53a*), and *Arid1a* (*Baf250a*) (Figure

3.1). Many of these genes (5 of 8) showed highest expression in cultures at the mesodermal stage of differentiation, including the enzymatic subunit *Smarca4* (referred to as *Brg1* for simplicity). To confirm this expression pattern, we measured expression of *Brg1* by quantitative PCR. Consistent with the RNA-seq analysis, we observed enrichment for *Brg1* mRNA at mesodermal stages of differentiation followed by downregulation with cardiomyocyte differentiation (Figure 3.2a). Furthermore, BRG1 protein showed a similar pattern during differentiation (Figure 3.2b). These findings suggested an uncharacterized role for BAF complexes in the early stages of cardiac differentiation.

3.4.2 Brg1 is required for cardiomyocyte differentiation

In order to understand the function of BAF complexes in cardiac differentiation, we focused on *Brg1* given its known importance in embryonic development and differentiation. To determine the requirement of *Brg1* in cardiomyocyte differentiation, we sought to inactivate *Brg1* at distinct points during the differentiation process. To this end, we utilized an ES cell line that allows for inducible deletion of *Brg1* using 4-hydroxytamoxifen (4-OHT) (Brg1 fl/fl; Actin-CreER, Ho et al., 2009). Treatment of these cells with 4-OHT led to near complete removal of the *Brg1* flox allele by 24 hours (Figure 3.3a). In addition, 4-OHT treatment led to a substantial reduction in BRG1 protein levels by 48 hours after treatment in cells cultured under ES cell conditions, as well as cells aggregated to form embryoid bodies (Figure 3.3b,c). By 72 hours, BRG1 protein was completely undetectable.

We next differentiated these cells to cardiomyocytes and depleted these cultures

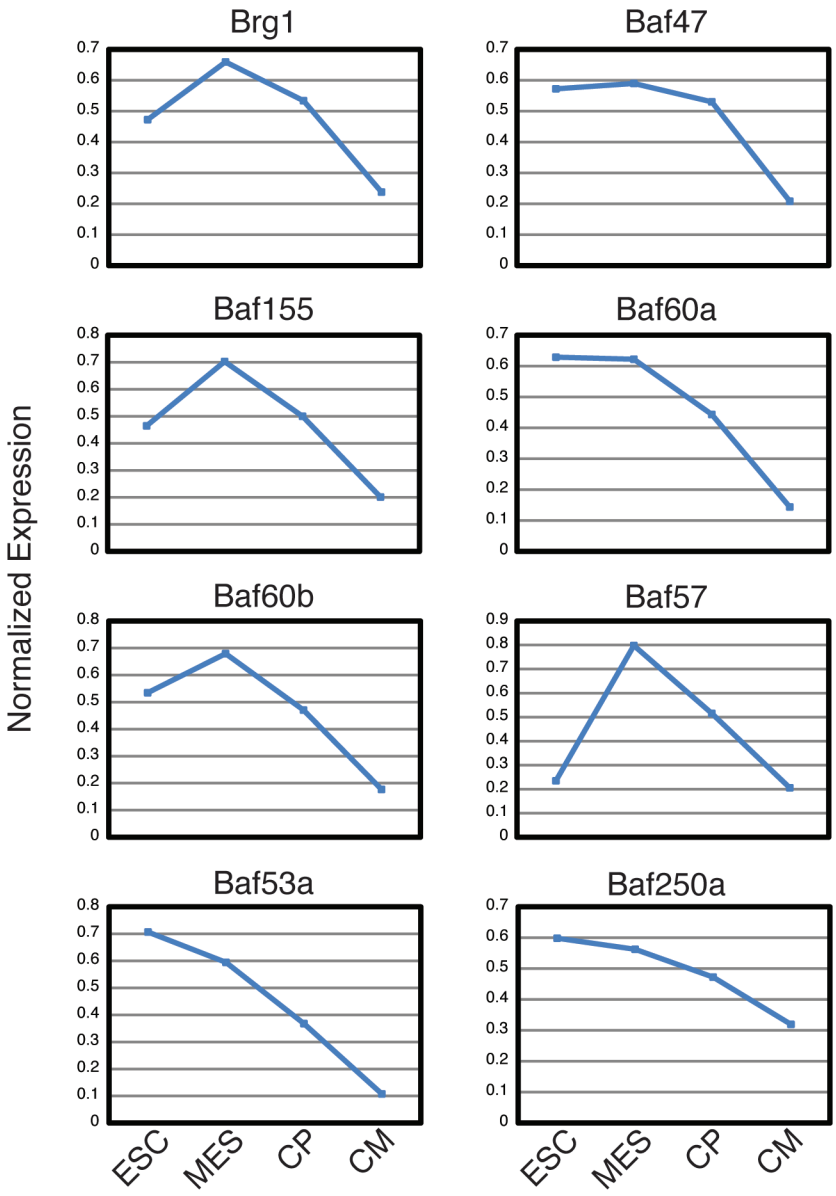


Figure 3.1. BAF complex subunits are downregulated with cardiac differentiation. RNA expression measured by RNA-seq demonstrates that many BAF complex subunits, including the ATPase *Brg1*, are downregulated during directed cardiomyocyte differentiation. Expression values are magnitude normalized as described in Section 2.5.14.

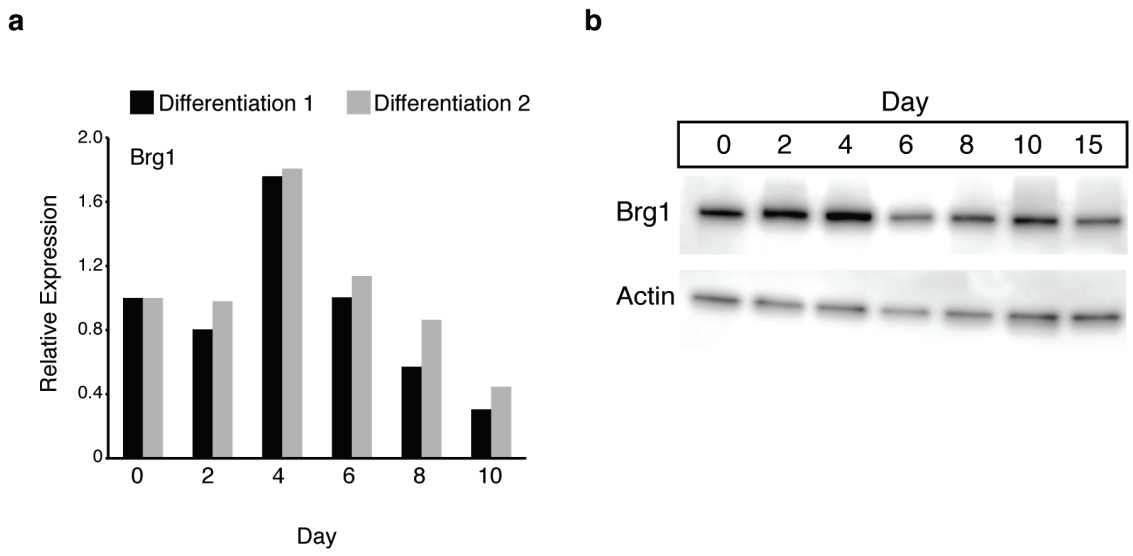


Figure 3.2. *Brg1* is downregulated during cardiomyocyte differentiation. a. qPCR analysis of *Brg1* expression for two separate cardiomyocyte differentiations demonstrates that *Brg1* expression peaks at Day 4 of differentiation and is subsequently downregulated. b. Western blot analysis of BRG1 protein levels during cardiomyocyte differentiation demonstrates high abundance of BRG1 during early stages of differentiation. Actin is used as a loading control.

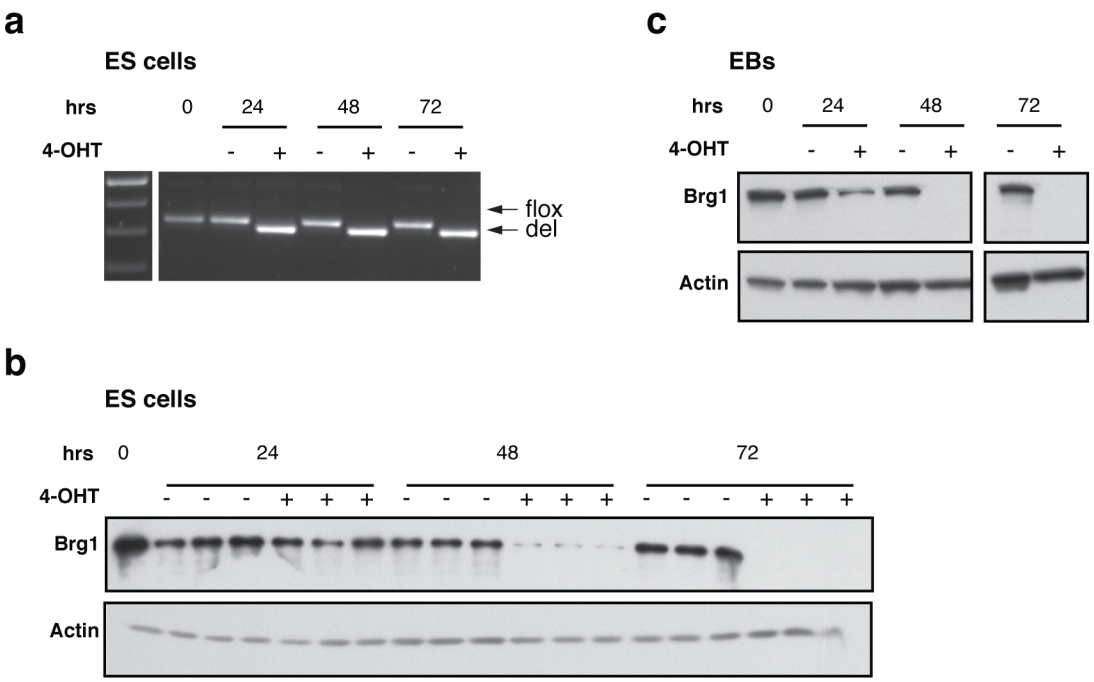


Figure 3.3. Treatment of *Brg1* fl/fl; Actin-CreER ES cells with 4-OHT leads to efficient deletion of *Brg1*. a. PCR genotyping demonstrates that treatment of *Brg1* fl/fl; Actin-CreER ES cells with 200 nM 4-OHT leads to complete loss of floxed allele (top band) by 24 hours. Floxed allele is converted into del allele (bottom allele) by Cre-mediate recombination. b. Western blotting of *Brg1* fl/fl; Actin-CreER ES cells treated with 4-OHT shows loss of BRG1 protein by 48 hours. Actin used as a loading control. c. *Brg1* fl/fl; Actin-CreER ES cells treated with 4-OHT while differentiated in serum-free media as embryoid bodies show similar loss in BRG1 protein.

for *Brg1* by addition of 4-OHT at three distinct timepoints of differentiation. We added 4-OHT on Day 8, which led to *Brg1* deletion after these cultures had differentiated to cardiomyocytes. In addition, we added 4-OHT at Day 2 and Day 4 of differentiation, which coincided with before and after mesoderm induction, respectively (Figure 3.4a). Treatment of cultures with 4-OHT at Day 8 resulted in little difference in the number of $cTnT^+$ cardiomyocytes seen by immunofluorescence when compared to control (Figure 3.4b). In contrast, 4-OHT treatment at Day 2 and Day 4 led to a significant reduction in the number of differentiated cardiomyocytes in end stage cultures (Figure 3.4b). We further quantified cardiomyocyte yield by intracellular flow cytometry and calculated the ratio of $cTnT^+$ cells detected in 4-OHT-treated cultures compared to control-treated cultures. Day 8 treated cultures showed similar numbers of $cTnT^+$ cells between 4-OHT and control treatments (ratio - 0.89; Figure 3.4c). However, Day 4 treatment led to fewer $cTnT^+$ cells in the 4-OHT treatment (ratio – 0.35 $p < 0.05$), and this was more pronounced in Day 2-treated cultures (ratio – 0.12 , $p < 0.001$; Figure 3.4c). These data demonstrate that *Brg1* is required for differentiation of cardiomyocytes from embryonic stem cells.

Despite similarities in cardiomyocyte yield, we measured gene expression changes in Day 8 treated cultures in order to identify *Brg1*-dependent cardiac gene expression. We profiled a panel of genes, including master cardiac TFs *Isl1*, *Gata4*, *Nkx2-5*, and *Tbx5*, as well as *Irx4* (ventricular myocardium), *Myh6*, *Myh7*, *Myl2*, *Myl7* (sarcomeric proteins), *Nppa*, *Nppb* (cardiac hormones), *Sox17* (endodermal marker), and *Bmp10* (cardiomyocyte proliferation) (Figure 3.5). We observed modest changes in gene expression, as only *Nppa* and *Myh7* were changed by greater than 2-fold versus control. However, changes in *Bmp10*, *Nppa*, and *Myh7* expression were

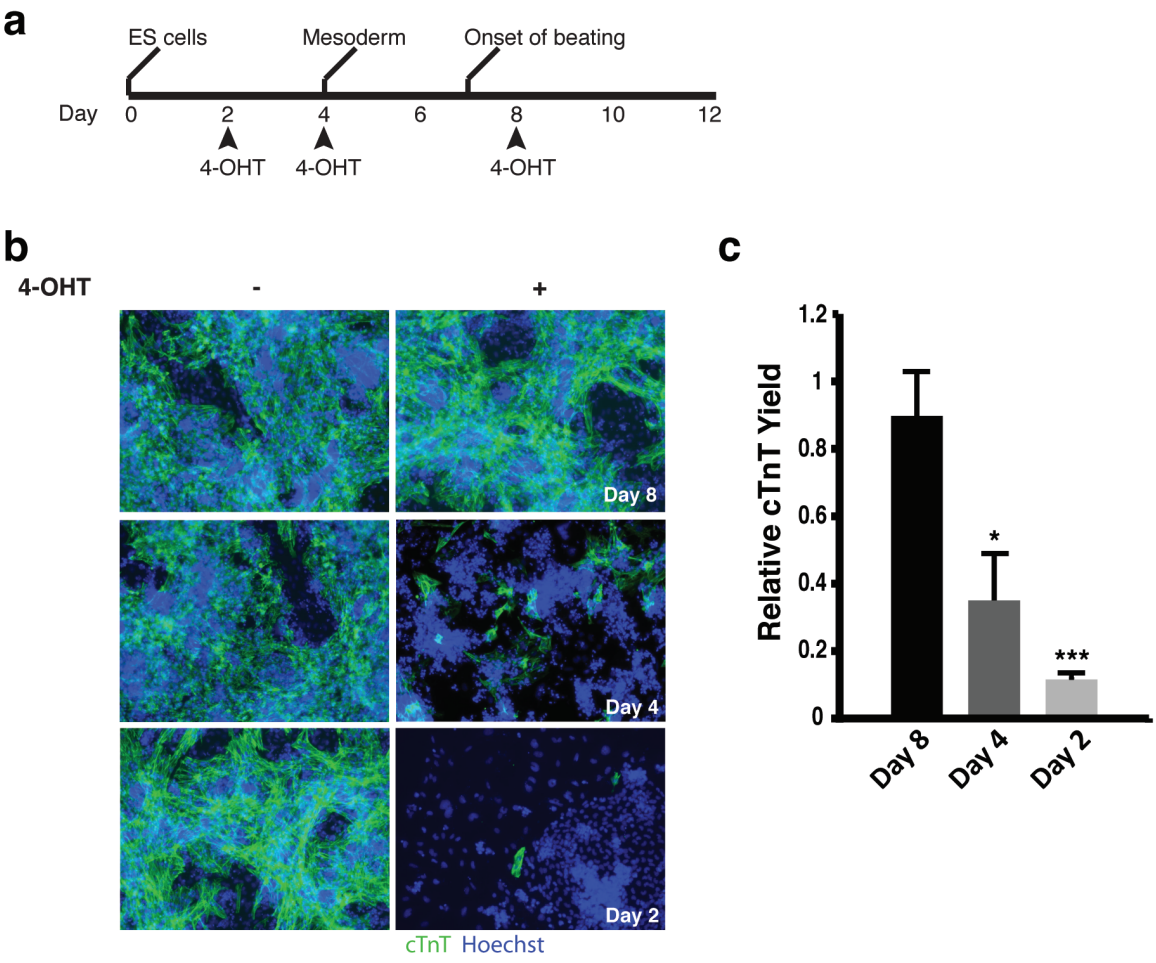


Figure 3.4. *Brg1* is required for cardiomyocyte differentiation from ES cells. a. Timeline depicting experimental design. 4-OHT treatment was initiated at three timepoints: Day 2, Day 4, and Day 8. Hallmarks of differentiation are indicated above. b. Immunofluorescence for cTnT at Day 12 of differentiation. DNA is stained using Hoechst. A reduction in the number of cardiomyocytes is seen in cultures treated with 4-OHT at Day 2 and Day 4. Control conditions are cells from the same differentiation treated with THF only. c. Quantification of cardiomyocyte purity by intracellular flow cytometry for cTnT demonstrates significant reductions in cardiomyocytes in Day 4 and Day 2 treatments.

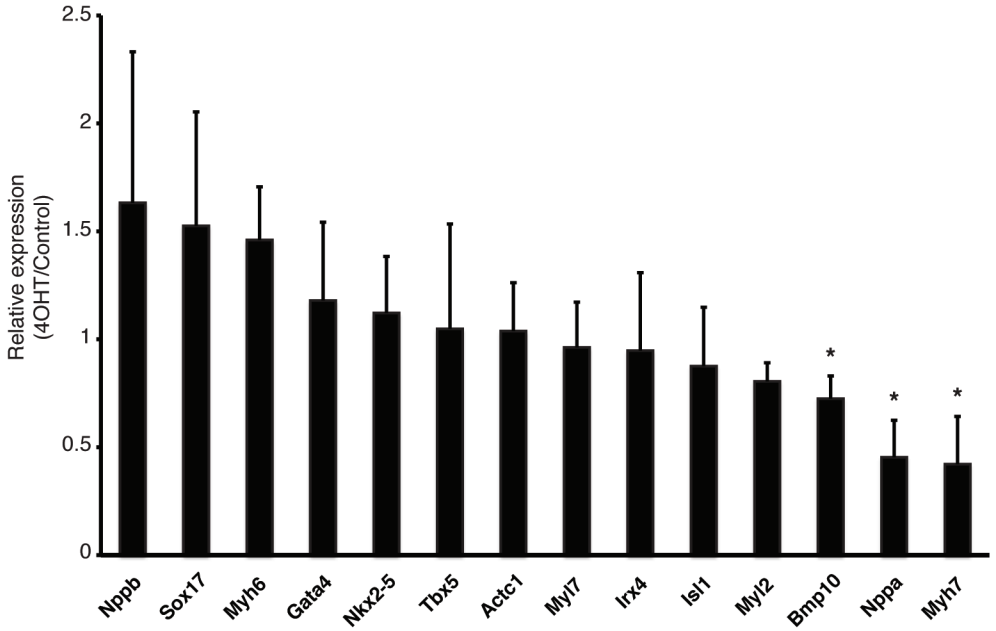


Figure 3.5. Loss of *Brg1* leads to misregulation of a small subset of cardiac genes. qPCR analysis of differentiated cultures treated at Day 8 with either control (THF) or 4-OHT. Values graphed are 4-OHT treated expression levels divided by control expression levels, such that no change gives a value of 1. Error bars represent standard deviation. * , $p < 0.05$, one sample t-test.

statistically significant ($p < 0.05$, one sample t-test). *Bmp10*, *Nppa*, and *Myh7* are known targets of *Brg1* in the developing heart (Hang et al., 2010, Takeuchi et al., 2011). Thus, our analysis confirms a modest and highly specific role for *Brg1* in cardiomyocytes.

We next focused on the function of *Brg1* during cardiac differentiation and focused on the early depletion of *Brg1* by addition of 4-OHT at Day 2 of differentiation, given the severity of the differentiation defect. We observed that many cells in these cultures appeared to undergo apoptosis rather than differentiate. To test this possibility, we used 7-AAD, a DNA-binding dye excluded from live cells. We observed increased numbers of 7-AAD$^+$ cells 40 hrs after 4-OHT treatment, confirming an increase in cell death in *Brg1*-depleted cultures (Figure 3.6). This difference became more pronounced by Day 12 of differentiation, as nearly all remaining cells were permeable to 7-AAD. These data demonstrate that *Brg1* is essential for the survival of differentiating cardiac precursors *in vitro*.

3.4.3 Loss of Brg1 causes dysregulation of developmental regulators

The differentiation defect seen in cultures depleted for *Brg1* during mesoderm differentiation suggested an important role for *Brg1* in mesodermal precursors. To better understand the function of *Brg1* in these cells, we profiled gene expression changes in normal and *Brg1*-depleted cultures during mesoderm differentiation. We collected RNA from EBs prior to mesoderm induction (Day 2), as well as EBs after mesoderm induction treated with either 4-OHT or control (Day 4) (Figure 3.7). Based on this analysis, we identified 350 genes that were downregulated and 502 genes that were upregulated in *Brg1*-depleted cultures (fold change > 2, FDR = 0.01; Figure 3.8). Among the genes

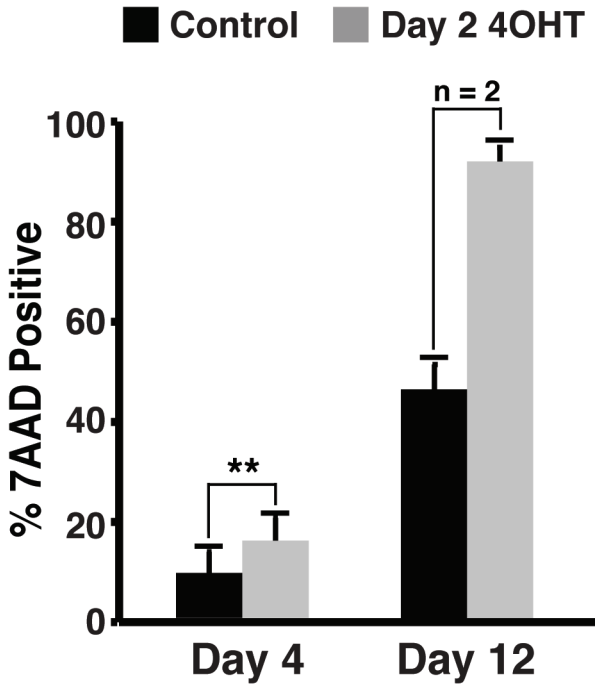


Figure 3.6. Loss of *Brg1* leads to an increase in cell death. Analysis of 7AAD staining by flow cytometry demonstrates an increase in stained cells in 4-OHT-treated cultures. **, $p < 0.01$. $n \geq 3$ unless noted.

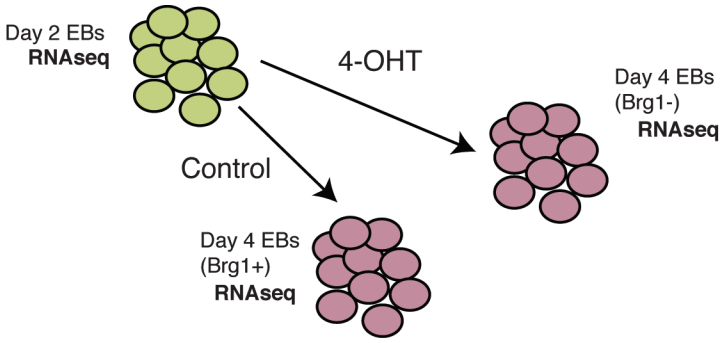


Figure 3.7. Cartoon representation of RNA-seq experimental design. Directed cardiac differentiations were collected for analysis of gene expression by RNA-seq at multiple timepoints and treatment groups. Cultures were collected at Day 2 of differentiation, which is prior to mesoderm induction. Cultures were also collected at Day 4 of differentiation (after mesoderm induction) when treated with either a control (THF) or 200 nM 4-OHT. This experimental design allowed for not only identification of differentially expressed genes comparing control and *Brg1*-deficient mesoderm, but also determination of how the expression of these genes changes during mesoderm induction.

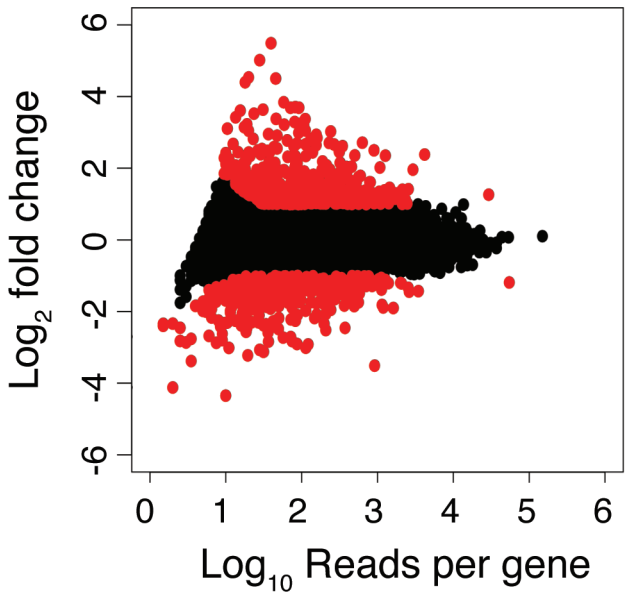


Figure 3.8. Loss of *Brg1* leads to significantly upregulated and downregulated gene expression. Scatterplot of global gene expression measured by RNA-seq. Reads per gene (log10) measured in control-treated cultures is plotted vs fold change (log2) between 4-OHT-treated and control-treated cultures. Red dots indicate genes that are significantly differentially expressed using an fold change and FDR cutoff of 2.0 and 0.01, respectively.

downregulated with *Brg1*-depletion were important regulators of mesoderm differentiation such as *Mesp1* and *Eomes*. These changes are consistent with a defect in the efficacy of cardiogenic mesoderm differentiation. Surprisingly, we also identified known transcription factors of the cardiac lineage, such as *Nkx2-5* and *Tbx5,* as being upregulated in *Brg1*-deficient mesodermal cultures.

We next separated genes based on whether they were upregulated or downregulated by loss of *Brg1* and analyzed their expression levels compared to uninduced cultures (Day 2), reasoning that these two groups of genes may be differentially regulated during mesoderm differentiation. This analysis revealed two major trends. Genes downregulated in *Brg1*-deficient cultures were generally activated during mesoderm differentiation (Figure 3.9a, left bar). Reduced levels of *Brg1* led to lower levels of induction for these genes during mesoderm differentiation (Figure 3.9a, right bar). Thus, this demonstrates a requirement for *Brg1* in the activation of gene expression during mesoderm induction. On the contrary, genes upregulated by loss of *Brg1* did not show a trend towards activation or repression during mesoderm differentiation. Instead, these genes remain unchanged compared to undifferentiated cultures (Figure 3.9b, left bar). Furthermore, many of these genes are expressed at very low levels in normal mesodermal cultures, but were ectopically-activated in *Brg1*-deficient cultures (Figure 3.9b, right bar). Taken together, our findings demonstrate a requirement for *Brg1* in gene activation, but also gene silencing, during the differentiation of mesoderm from ES cells.

We next performed gene ontology (GO) analysis to identify the functional classification of genes affected by *Brg1* depletion during differentiation. Genes

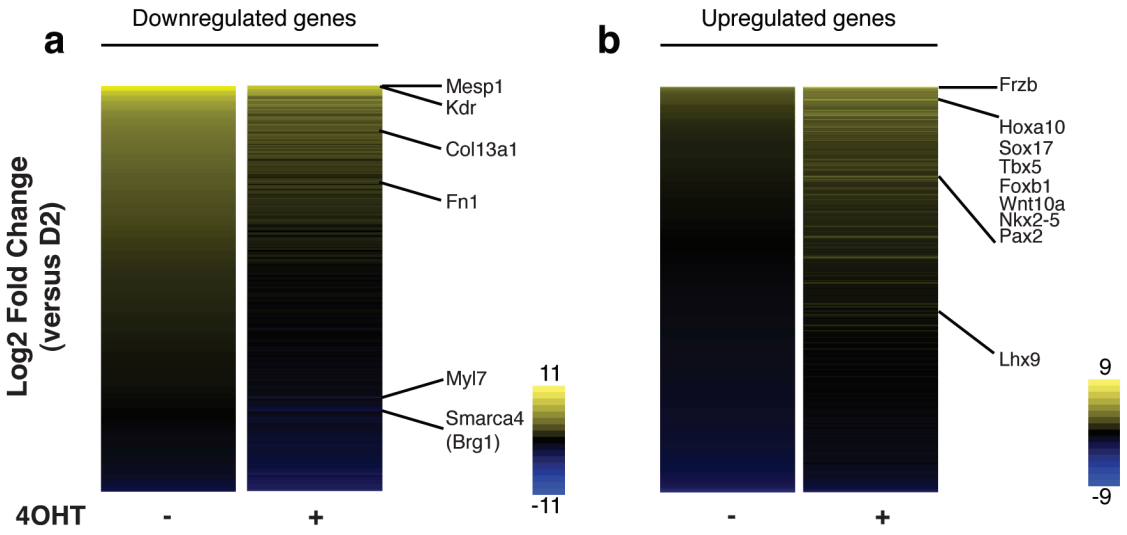


Figure 3.9. Genes downregulated and upregulated by loss of *Brg1* show distinct regulation during mesoderm differentiation. a-b. Heatmap representation of fold changes (log_2) during mesoderm differentiation for genes downregulated (a) and upregulated (b). In each case, control treated cultures (left bar) and 4-OHT treated cultures (right bar) are compared to Day 2 cultures to calculate fold change with mesoderm induction. Yellow and blue color indicates activated or repressed expression during mesoderm induction, respectively.

downregulated by loss of *Brg1* showed broad enrichment for genes involved in diverse biological processes. Furthermore, these genes encoded components of extracellular matrix as well as signaling molecules (Figure 3.10a). Strikingly, upregulated genes were highly enriched for genes that function in development and organogenesis (Figure 3.10b). These genes were also classified as DNA-binding factors. Indeed, many lineage-specific transcription factors were upregulated by reduced levels of *Brg1*, including those expressed in derivatives of all three germ layers. This included cardiac TFs, which appeared to be derepressed prior to their normal expression pattern.

3.4.4 Brg1-dependent genes are Polycomb target genes

Brg1 functions through the regulation of chromatin. Hence, we sought to identify commonalities in the chromatin regulation of *Brg1*-dependent genes. To address this, we used our previously generated datasets for genome-wide localization of histone modifications and RNA polymerase in mesodermal stage differentiations (see Chapter 2) and looked at the average enrichment of each mark around the TSS of *Brg1*-dependent genes. Interestingly, this analysis showed that both genes downregulated and upregulated by loss of *Brg1* were enriched for H3K27me3 when compared to all genes (Figure 3.11a). H3K27me3 is deposited by PRC2, which mediates the repression of many developmental genes in ES cells. Consistent with this, upregulated genes, which showed the most robust levels of H3K27me3, are highly enriched for many developmental TFs (Figure 3.10). Hallmarks of active chromatin such as H3K4me3, H3K27ac, and RNAP were also found at upregulated genes. H3K4me3, and in some cases RNA polymerase, have been detected at H3K27me3 marked genes in pluripotent

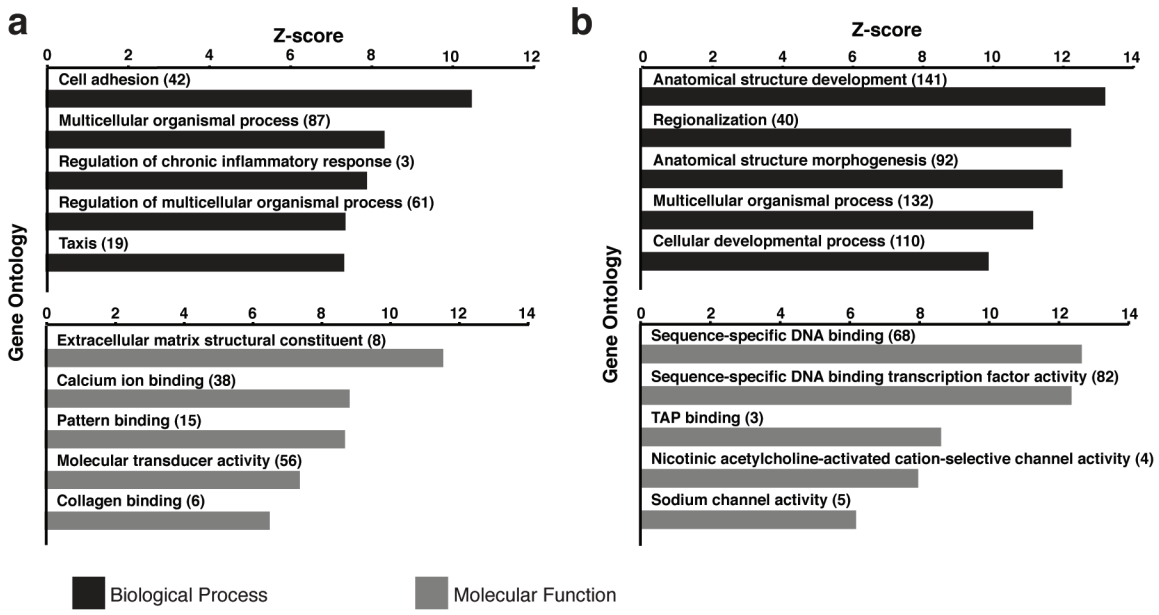


Figure 3.10. *Brg1* is required for the regulation of developmental genes during mesoderm induction. Gene ontology over-representation analysis of genes downregulated (a) and upregulated (b) by loss of *Brg1*. Black indicates ontologies related to biological processes; grey indicates ontologies related to molecular functions. Genes upregulated by loss of *Brg1* are particularly enriched for transcription factor genes that are involved in development.

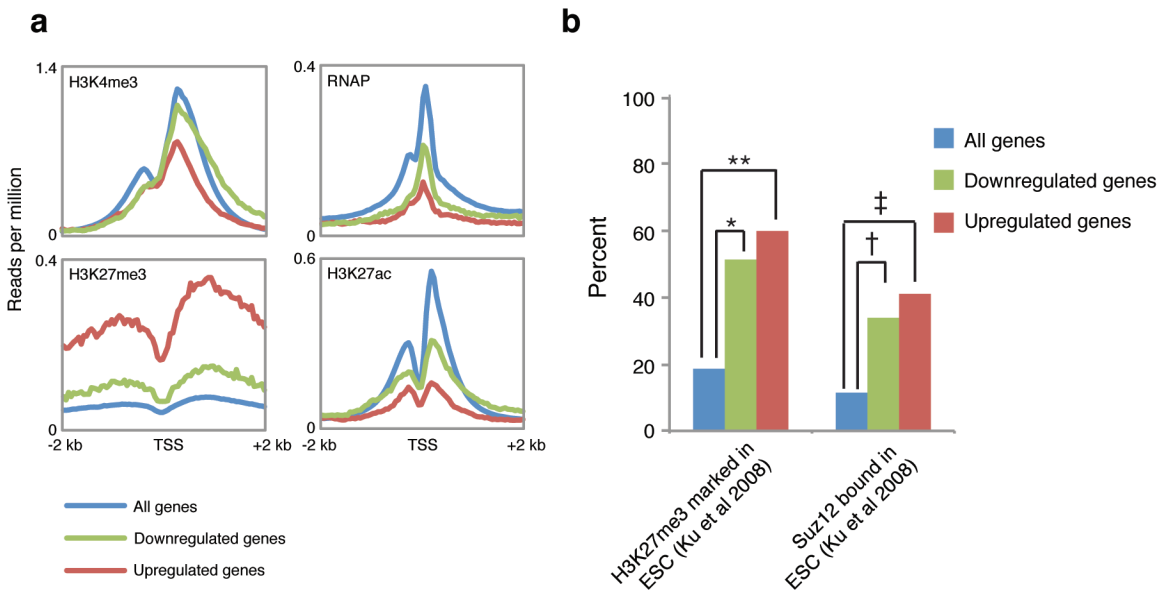


Figure 3.11. *Brg1*-dependent genes are targets of Polycomb repressive complexes. a. Metagene analysis demonstrates enrichment profiles for chromatin features relative to the TSS in mesodermal cultures. Both downregulated genes and upregulated genes have higher levels of H3K27me3 than all genes. b. Overlap between downregulated or upregulated genes and genes identified to be marked by H3K27me3 or bound by SUZ12 in ES cells (based on data from Ku et al., 2008). Downregulated genes and upregulated genes have greater relative overlap with H3K27me3/Suz12 than expected by chance. Upregulated genes include those genes significantly upregulated upon loss of *Brg1*. Downregulated genes include those genes significantly downregulated upon loss of *Brg1*. * - $p = 1.76 \times 10^{-28}$; ** - $p = 9.03 \times 10^{-80}$; † - $p = 9.05 \times 10^{-26}$; ‡ - $p = 5.15 \times 10^{-82}$.

ES cells in regions termed bivalent domains (Bernstein et al., 2006, Pan et al., 2007, Mikkelsen et al., 2007, Stock et al., 2007). It is speculated that this unique chromatin configuration poises genes for subsequent activation with differentiation. Despite being expressed at low levels in mesodermal cultures, our analysis supports the possibility that genes upregulated by loss of *Brg1* may exist in a similar poised state during mesoderm induction. This might explain their potential for transcriptional activation upon loss of *Brg1*.

To further establish the link between *Brg1* and Polycomb regulation, we utilized published genome-wide datasets for H3K27me3 and *Suz12* occupancy in undifferentiated ES cells to determine the overlap with *Brg1*-dependent genes. We found that many *Brg1*-dependent genes are marked by H3K27me3 in ES cells (Figure 3.11b). Indeed, 51% of downregulated genes and 60% of upregulated genes are marked by H3K27me3 in ES cells, which is much higher than the incidence observed for all genes (19%). Moreover, Brg1-dependent genes were similarly enriched for SUZ12 occupancy (Figure 3.11b). These data strongly demonstrate that Brg1-dependent genes are targets of Polycomb repressive complexes.

3.4.5 Brg1 is required for H3K27me3 levels at upregulated genes

Recent studies have implicated *Brg1* in regulating H3K27me3 levels in ES cells (Ho et al 2011). Thus, we hypothesized that *Brg1* might be required for appropriate levels of PRC2-mediated silencing at target genes. We looked at genome-wide levels of H3K27me3 in normal and Brg1-depleted mesodermal differentiations by ChIP-seq. This analysis revealed a clear reduction in H3K27me3 at genes derepressed upon loss of

Brg1 (Figure 3.12a). Interestingly, this was not observed for genes downregulated by loss of *Brg1*, which supports a specific requirement for *Brg1* in H3K27me3 enrichment at derepressed genes. Genes with reduced levels of H3K27me3 included developmental regulators such as *Sox21*, *Pax2*, and *Nkx2-9* (Figure 3.12b). We also observed reduced H3K27me3 at Hox loci, such as the Hoxb cluster (Figure 3.12b). This is consistent with the observed gene expression changes seen in *Brg1*-depleted cultures, which expressed members of each Hox cluster at elevated levels. *Brg1* did not appear to be regulating H3K27me3 levels indirectly through regulation of PRC2 subunits. *Ezh1*, *Ezh2*, *Suz12*, and *Eed* were all expressed at normal levels in Brg1-deficient cultures (data not shown). Moreover, H3K27me3 demethylases *Jmjd3* and *Kdm6a* were not significantly altered by loss of *Brg1*. These findings support an important role for *Brg1* in H3K27me3 enrichment at developmental regulators. Taken together, our study provides novel insights into the transcriptional control of cardiac differentiation and suggest changing roles of *Brg1*-containing complexes in this process. Furthermore, our findings provide additional support for a cooperativity between *Brg1* and Polycomb complexes.

3.5 Discussion

3.5.1 An expanded role for Brg1 in cardiac differentiation

An essential role for *Brg1* in cardiac development has now been robustly demonstrated (Stankunas et al., 2008, Hang et al., 2010, Takeuchi et al., 2011). However, this essential role stems from the dysregulation of a surprisingly small set of essential genes in *Brg1*-deficient hearts. Misregulation of *Bmp10*, α-MHC (*Myh6*), and β-MHC (*Myh7*) correlate with proliferation and maturation defects in developing myocardium without

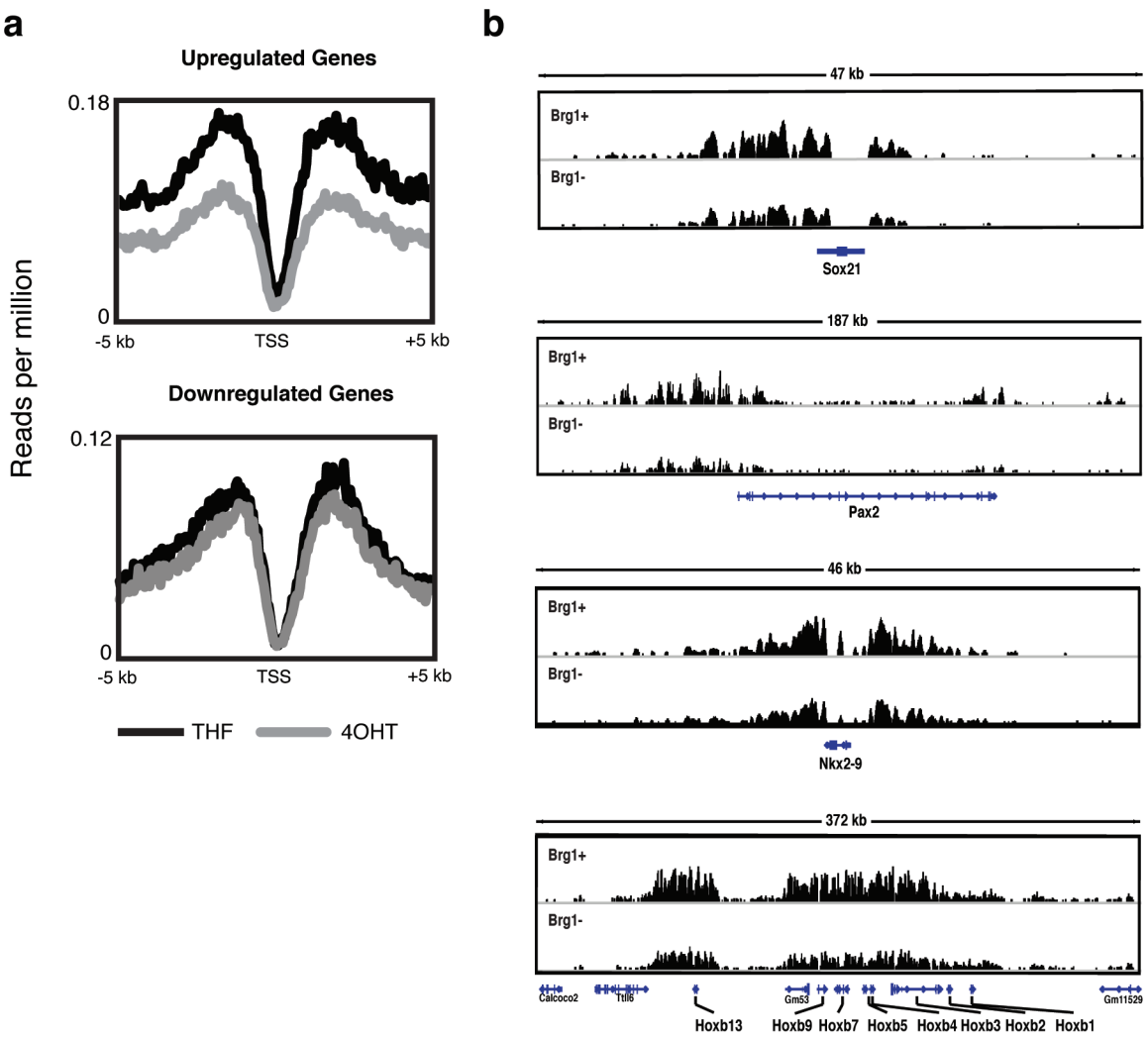


Figure 3.12. *Brg1* is required for H3K27me3 levels at genes upregulated by loss of *Brg1*. a. Metagene analysis demonstrates the enrichment profile of H3K27me3 relative to the TSS in mesodermal cultures treated with either control or 4-OHT. Genes upregulated by loss of *Brg1* show a clear reduction in H3K27me3 levels in the *Brg1*-depleted cultures. b. Example genomic regions showing reductions in H3K27me3 in -depleted cultures.

Brg1 in the mouse (Hang et al., 2010). Ventricle specific deletion of *Brg1* using an *Nkx2-5::Cre* leads to abnormal chamber morphology and reduced levels of *Nppa*, *Tbx5*, and *Bmp10* (Takeuchi et al., 2011). Furthermore, trabeculation defects due to loss of *Brg1* in the developing endocardium derive from derepression of a single gene, *Adamts1*, and can be rescued with chemical inhibition of this enzyme (Stankunas et al., 2008). This is consistent with studies in other lineages, which have uncovered regulation of essential tissue-specific genes by *Brg1*. Our study focused on ES cell differentiation as a model system to study *Brg1* function specifically during cardiomyocyte differentiation and allowed for distinction between early and late roles for *Brg1* in this process. Consistent with the aforementioned literature, our data support a highly specific role for *Brg1* in differentiated cardiomyocytes. From the panel of genes measured in Day 8 treated cultures, only *Bmp10*, α-MHC, and *Nppa* were significantly changed. Although β-*MHC* was not significantly altered in our analysis, we observed a trend for upregulation in *Brg1*-deficient cultures that is consistent with previous studies in embryonic hearts. Thus, known targets of *Brg1* were affected by *Brg1* deletion in cardiomyocytes, while instructive factors, such as key cardiac TFs *Nkx2-5*, *Tbx5*, and *Isl1,* were unaffected.

In addition, our studies have uncovered a requirement for *Brg1* in the differentiation of cardiomyocytes, as loss of *Brg1* during directed differentiation of ES cells leads to severe reductions in the number of $cTnT^+$ cells. This is consistent with other studies implicating BAF complex subunits with mesoderm differentiation (Gao et al., 2008). Analysis of gene expression during mesoderm differentiation demonstrated that *Brg1* is required for the proper regulation of a diverse group of key developmental regulators. Master regulators of mesoderm differentiation *Mesp1* and *Eomes* showed reduced

expression in *Brg1*-deficient cultures. This might be due to a reduction in the number of transcripts per cell, indicating a role for *Brg1* in the activation of these genes. Reduced numbers of *Mesp1*/*Eomes*-expressing cells resulting from a differentiation defect could also explain these findings. Analysis of FLK-1, which is also downregulated at the level of RNA, suggests both possibilities may occur. *Brg1*-deficient cultures have fewer $FLK1^{+}$ cells as well as less receptor per cell (data not shown). While it is unclear whether *Brg1* is directly regulating the expression of these genes, direct regulation is compatible with these findings. *Brg1* may be essential for efficient, synchronized gene activation (leading to reduced numbers of expressing cells), as well as maximal expression levels (leading to reduced transcripts per cell). BAF complexes and homologous complexes in other organisms have been shown to be critical mediators of gene activation in a variety of contexts (Fry et al.,2001, Trotter and Archer., 2007).

Our studies also demonstrate a requirement for *Brg1* in gene repression. Our RNA-seq analysis of differentiating cultures with reduced levels of *Brg1* showed multiple developmental regulators were expressed at higher levels when compared to control cultures. Furthermore, we observed that many of these genes were normally expressed at very low levels during mesoderm induction. Together, these data suggest that many of the genes upregulated by loss of *Brg1* are normally kept off during mesoderm differentiation but are ectopically activated with reduced *Brg1* levels. This group of genes included many lineage-specific transcription factors such as members of the Pax, Hox, Sox, and Tbx TF families. *Nkx2-5*, a cardiac TF, was also upregulated in *Brg1*-deficient mesoderm. This likely represents the premature derepression of *Nkx2-5* prior to its normal expression pattern.

Our studies suggest that these gene expression changes may be specific to

mesoderm differentiation. Analysis of *Nkx2-5* and *Sox17* (also misregulated in mesoderm) demonstrated that these genes were not upregulated in *Brg1*-deficient cardiomyocytes. One possible explanation is that both *Nkx2-5* and *Sox17* have detectable expression levels of cardiomyocytes such that these genes may no longer be targets of active repression. However, many genes upregulated in mesoderm do not show similar levels of derepression when *Brg1* is depleted in undifferentiated ES cells or EB cultures not induced to form mesoderm (data not shown). Moreover, our analysis clearly shows that loss of *Brg1* leads to dramatically different effects depending on when during differentiation it is depleted. Together, these results suggest changing functions for *Brg1* during cardiac differentiation.

There are many potential sources for specificity of *Brg1* function in cardiac differentiation. First, genes may recruit additional layers of gene repression, such as DNA methylation, upon terminal differentiation. This would make gene silencing more robust and limit the effect of disrupting any one aspect of repression. Secondly, gene misregulation could depend on the cellular context such as expression of additional transcriptional regulators or activation of particular signaling pathways. In this case, perhaps gene activation requires relief of gene repression along with activity of cell type specific activators. Lastly, the function of *Brg1* might be directly modulated during differentiation. This possibility is perhaps most intriguing. Vertebrate BAF complexes are modulator and polymorphic, which set them apart from homologous complexes in *D. melanogaster* and *S. cerevisae*. Thus, incorporation of different subunits may fine-tune complex function during cardiac differentiation. These subunit "switches" occur often in embryonic development and play instructive roles in cell commitment and differentiation (Lessard et al., 2007, Lamba et al., 2008, Forcales et al., 2012). Incorporation of *Baf60c*

(also known as *Smarcd3*) is an appealing choice for such a switch in the cardiac lineage. Baf60c is restricted in the crescent-stage embryo to the differentiating cardiac tissue and is essential for cardiac morphogenesis (Lickert et al., 2004). Moreover, Baf60c promotes cardiac transdifferentiation of non-cardiac tissue (Takeuchi and Bruneau, 2009). *Baf60c* expression is initiated between mesodermal induction and cardiomyocyte differentiation. Thus, onset of *Baf60c* expression correlates with apparent changes in *Brg1* function and may mediate a specialization of cardiac BAF complexes towards regulating a subset of critical cardiac genes.

3.5.2 Brg1 and Polycomb-mediated silencing

Polycomb repressive complexes are well-established repressors of developmental regulators through chromatin compaction (Surface et al., 2010). We find that *Brg1*-dependent genes are enriched for Polycomb targets in ES cells. Interestingly, downregulated genes show lower levels of H3K27me3 (although still enriched when compared to all genes) in mesodermal cultures. Low H3K27me3 is consistent with gene activation and indeed many genes downregulated by loss of *Brg1* are typically activated with mesoderm induction. In addition, H3K27me3 levels at these genes are not modulated by loss of *Brg1*. This suggests that *Brg1* is required for gene activation in a manner other than antagonizing Polycomb repression, which has been observed at STAT3 target genes in undifferentiated ES cells (Ho et al., 2011).

In contrast, H3K27me3 levels are dependent on *Brg1* at upregulated genes. These genes demonstrate robust levels H3K27me3 in mesodermal cultures and are not transcriptionally active. Loss of *Brg1* leads to the ectopic expression of these genes

during mesoderm induction. These findings are consistent with previously published data for undifferentiated ESCs, which showed *Brg1*-dependent H3K27me3 and gene silencing of Hox loci (Ho et al., 2011). However, our work suggests that, at least during mesoderm differentiation of ES cells, *Brg1* plays a role in PRC2-mediated silencing of not only Hox loci, but also a diverse group of developmental regulators, including cardiac TFs and other TFs kept silent in the developing heart. For example, *Six1* is upregulated in *Brg1*-deficient cultures and has recently been identified as misexpressed in *Ezh2*-deficient hearts (Delgado-Olguin et al., 2012, He et al., 2012). While current data does not support widespread colocalization of *Brg1* and PRC2 complexes (Ho et al., 2009), occupancy data for chromatin remodeling complexes suffer from poor enrichment and low signal to noise (Gelbart et al., 2005). These difficulties have limited our ability to test for direct *Brg1* regulation of *Brg1*-dependent genes. Therefore, we cannot exclude the possibility that *Brg1* affects PRC2-silencing indirectly via an unknown factor.

Developmental regulators posses promoter regions in a bivalent chromatin state, marked with H3K4me3 as well as H3K27me3, which is thought to poise these genes for later activation during development and differentiation. Furthermore, these genes have other features of transcriptional poising, such as a stalled RNA polymerase (Guenther et al., 2007, Stock et al., 2007, Zeitlinger et al., 2007). Indeed, many genes appear to transition through a poised transcriptional state during cardiac differentiation (Golob et al., 2011). Our analysis revealed that genes upregulated by *Brg1*-depletion shared many features with a poised state, including clear enrichment for H3K4me3 and H3K27me3 and low but detectable enrichment for RNAP and histone acetylation. This might explain the sensitivity of these genes to even partial reductions in H3K27me3 as seen in *Brg1*-deficient cultures. In addition, if the establishment of poising at these

genes is a regulated step during mesoderm differentiation, this could help to explain the stage-specificity of *Brg1*-dependent gene expression described in our studies.

3.5.3 Brg1 classification as a Trithorax-group gene

The Drosophila homolog of *Brg1*, *brahma* (*brm*), was originally discovered for its regulation of homeotic genes, which are master transcriptional regulators in metazoa that specify segment identity. Specifically, reduced *brm* levels abolished homeotic transformation found in *polycomb* mutants, leading to the model that *brm* antagonizes *polycomb* function as a Trithorax-group gene by activating homeotic genes (Tamkun et al., 1992). This supported work in yeast, which had mainly uncovered activating roles for yeast SWI/SNF complexes. However, recent insights have suggested mammalian BAF complexes may have evolved distinct functions. For instance, BAF subunits, including *Brg1*, play clear roles in gene repression and physically-associate with transcriptional repressors like HDACs (Underhill et al., 2000, Sif et al., 2001, Pal et al., 2003, Bilodeau et al., 2006). In addition, recent studies in ES cells demonstrate that loss of Brg1 leads to derepression of Hox genes, suggesting cooperation rather than antagonism with PRCs (Ho et al., 2011). Our data expand this and demonstrate that during differentiation PRC2-mediated repression is *Brg1*-dependent at many genes, including the four Hox loci. It is not clear if this reflects direct repression of Hox loci by *Brg1*-containing complexes. However, these data indicate that, at least in some contexts, *Brg1* is not a Trithorax-group gene in mammals. Instead, *Brg1* is essential for the repression of many Polycomb targets and may play changing roles in the activation and repression of these genes through development.

These findings join other examples of a unique relationship between classic Trithorax- and Polycomb-group members in mammals. For instance, bivalent domains are prevalent in mammals, demonstrating recruitment for both H3K4me3-depositing Trithorax complexes as well as H3K27me3-depositing Polycomb complexes to these regions (Bernstein et al., 2006). In Drosophila, however, these domains are rare, and the localization of Polycomb and Trithorax is largely exclusive (Armstrong et al., 2002, Kharchenko et al., 2011). Thus, it appears that the relationship between these two fundamental chromatin regulatory families has been altered during the course of vertebrate evolution to adapt conserved regulators to the need for increased complexity in gene regulation.

3.6 Materials and methods

3.6.1 Cardiomyocyte differentiation

Mouse ES cells were cultured in feeder-free conditions using standard techniques. For directed differentiations (Kattman et al., 2011) mouse ES cells were aggregated into embryoid bodies (EB) and cultured at 75,000 cells/ml for two days in serum free media (3 parts IMDM (Cellgro #15-016-CV): 1 part Ham's F12 (Cellgro #10-080-CV), 0.05% BSA, 2 mM GlutaMax (Gibco), B27 supplement (Gibco #12587010), N2 supplement (Gibco #17502048)) supplemented with 50 ug/ml ascorbic acid and 4.5×10^{-4} M monothioglycerol. Embryoid bodies were dissociated and reaggregated for 40 hours in the presence of 5 ng/mL human VEGF (R&D #293-VE) and human Activin A (R&D #338-AC) and human BMP4 (R&D #314-BP) at concentrations empirically determined depending on lot. EBs were dissociated and plated at 470,000 cells/cm^2 in StemPro-34 (Gibco #10639011) supplemented with 5 ng/mL VEGF, 10 ng/mL human basic FGF

(R&D #233-FB) and 25 ng/mL FGF10 (R&D #345-FG).

For Brg1 deletion studies, Brg1fl/fl; Actin-CreER ES cells (Ho et al., 2009, Ho et al., 2011) were cultured and differentiated as described above. Cultures were treated with 200 nM 4-hydroxytamoxifen (4-OHT) diluted from a 5 mg/mL stock solution in tetrahydrofuran (THF). Control cultures were treated with only THF. Media and 4OHT were replaced everyday until the end of the experiment, except between Day 2 and Day 4 and between Day 4 and Day 6. In these cases, treated media was left longer than 24 hours as replacement of media can affect differentiation efficacy.

3.6.2 Immunofluorescence microscopy

Immunofluorescence staining was performed on differentiations cultured in 96 well format. Cultures were fixed for 30 minutes at room temperature in 3.7% formaldehyde D-PBS and washed once with D-PBS. Wells were then blocked in 2% bovine serum albumin 0.1% Triton-X-100 D-PBS for 30 minutes at RT. After blocking, cultures were incubated with primary antibody at 4°C overnight. Slides were washed three times with 0.1% Triton X-100 D-PBS and incubated in secondary antibody at room temperature for 1 hour. After staining, slides were washed three times with 0.1% Triton X-100 D-PBS, stained with Hoechst 33342 (10 ug/mL) in D-PBS, and immediately imaged in 50 uL D-PBS. Antibodies were anti-cardiac isoform of Troponin T (cTnT) 1:100 (Thermo Scientific #MS295, Clone 13-11).

3.6.3 Flow cytometry

Cultures were trypsinized, quenched with serum, and fixed in D-PBS with 3.7% formaldehyde for 30 minutes at room temperature. Fixed samples were washed twice and stained with anti-cTnT for 30 minutes at room temperature. After staining, samples were washed twice, incubated with secondary antibody, and washed two additional times. All steps performed in D-PBS with 0.5% saponin and 4% FBS. Samples were stained with Hoechst 33342 (10 ug/mL) in D-PBS with 4% FBS. Samples were analyzed on an LSRII flow cytometer (BD). Data was analyzed using FlowJo software (Treestar).

For cell death analysis, cells were trypsinized and resuspended in 4% FBS in D-PBS. After one wash in 4% FBS in D-PBS, cells were resuspended in 4% FBS D-PBS with 20 ug/mL 7-AAD and incubated at 4°C for 30 minutes. Samples were analyzed on an LSRII flow cytometer (BD). Data was analyzed using FlowJo software (Treestar).

3.6.4 Quantitative PCR

RNA was extracted using *TRIzol®* and 500-60 ng RNA was reverse transcribed using High-Capacity cDNA Reverse Transcription kit (Applied). Quantitative PCR was performed using Taqman probes and expression was normalized to Gapdh. All reactions were performed in triplicate. The following probes were used: *Bmp10* – Mm01963768_s1, *Nppa* – Mm01255747_g1, *Nppb* – Mm00435304_g1, *Actc1* – Mm01333821_m1, *Myl2* – Mm00440383_m1, *Myl7* – Mm00491655_m1, *Gata4* – Mm00484689_m1, *Irx4* – Mm00502170_m1, *Myh6* – Mm00440354_m1, *Myh7* – Mm00600555_m1, *Nkx2-5* – Mm00657783_m1, *Sox17* – Mm00488363_m1, *Tbx5* –

Mm00803521_m1, *Isl1* – Mm00627860_m1, *Brg1* – Mm01151944, *Gapdh* – 4352932-0812026

3.6.5 Western blotting

Western blotting was performed using standard techniques. Briefly, protein lysate was sonicated (4 pulses for 30 seconds) and cleared by centrifugation at 13,000 RPM. Supernatant was diluted 1:1 with 2x Laemmli Buffer and 100 mM DTT and boiled for 10 minutes at 95°C. Following electrophoresis, protein was transferred to a PVDF membrane. Membranes were blocked for 1 hour at room temperature with 5% milk Tris-buffered saline Tween (TBST). Following blocking, membranes were incubated with desired antibody in 5% milk TBST overnight at 4°C. Membranes were washed 4 times for 15 minutes at room temperature in TBST and then stained with secondary antibody in 5% milk TBST for 1 hour at room temperature. After antibody staining, membranes were washed as after primary incubation, incubated in SuperSignal chemiluminescence substrate (Thermo Scientific) and visualized.

3.6.6 RNA-seq

Total RNA was isolated from 1.5-2x10^6 cells using *TRIzol®* Reagent according the manufacturer's instructions. 8 ug of total RNA was used as input material for the preparation of the RNA-Seq libraries, as indicated by previous iterations of the Illumina RNA-Seq protocol. Sequencing libraries were prepared according to Illumina RNA Seq library kit with minor modifications. Briefly, mRNA was isolated using Dynabeads®

mRNA Purification Kit (Invitrogen) followed by fragmentation (Ambion) and ethanol precipitation. First and second strand synthesis were performed followed by end repair, A-tailing, adapter ligation and size selection on a Beckman Coulter SPRI TE nucleic acid extractor. 200-400 bp dsDNA was enriched by 13 cycles of PCR with Phusion® High-Fidelity DNA Polymerase (NEB). Amplified libraries were sequenced on an Illumina HiSeq 2000 system.

3.6.7 RNA-seq analysis

Single end 40 bp reads were aligned to the mouse genome (mm9) using Bowtie (Langmead et al., 2009). Differential gene expression between conditions was determined using the USeq package (Nix et al., 2008) with minimum FDR threshold set to 20 and considering all Refseq genes. USeq was also used to calculate RPKM and fold change values between conditions. Repeat sequence reads greater than 50 were excluded from the analysis as recommend by USeq developers and genes expressed below 0.5 RPKM in all conditions were excluded from subsequent analysis. Gene ontology analysis was performed using Go Elite (http://www.genmapp.org/go_elite/) using all genes with an RPKM of greater than 0.5 in at least one condition as the gene universe. Graphical representation of upregulated and downregulated genes was performed in R.

3.6.8 Metagene analysis

Average enrichment of histone modifications and RNA polymerase were compared for all Refseq genes and genes corresponding to those significantly upregulated or

downregulated based on the RNA-seq analysis. This analysis was performed using the scripts found in biotoolbox (http://code.google.com/p/biotoolbox/). Briefly, aligned sequence reads within 4 kb of transcriptional start sites were grouped into 50 bp bins and read counts per bin were then normalized based on read number per library (reads per million) and the number of genes considered.

3.6.9 Polycomb target analysis

TSSs were extended 1 kb in each direction and compared to genome-wide data sets for H3K4me3, H3K27me3, and Polycomb subunits in ES cells (Ku et al 2008) to identify overlap. Transcriptional start sites for all Refseq genes and genes corresponding to these significantly upregulated or downregulated based on the RNA-seq analysis were considered. A hypergeometric test was used to determine statistical significance and calculate p-values for this comparison.

3.6.10 ChIP-seq

Chromatin immunoprecipitation of histone modifications were performed according to the Young lab protocol (Lee et al., 2006) with minor modification. Briefly, frozen pellets of cross-linked cells ($10x10^6$) were thawed in cold lysis buffer 1 (50 mM HEPES-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, 1× protease inhibitors) and gently rocked at 4°C for 10 minutes in 15 mL conical tubes. Cells were pelleted at 1350 x g at 4°C and resuspended in cold lysis buffer 2 (10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1× protease inhibitors) and

gently rocked at 4°C for 10 minutes in 15 mL conical tubes. Cells were pelleted at 1350 x g at 4°C in a table top centrifuge and resuspended in 0.5 mL cold ChIP lysis buffer (50 mM HEPES-NaOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% sodium deoxycholate) and sonicated to 200-1000 bp fragments using a VirSonic sonicator. Sonicated lysates were cleared by pelleting insoluble material at 13,000 RPM at 4°C followed by incubation with 5 ug antibody overnight. Next, Protein A magnetic beads (45 uL) were added to the lysate and incubated at 4°C for 7 hrs. Prior to addition, magnetic beads were washed 3 times with block (0.5% BSA/PBS). Immunoprecipitated material was washed 2 times each with ChIP lysis buffer, high salt lysis buffer (50 mM HEPES-NaOH pH 7.5, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% sodium deoxycholate), and LiCl wash buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 1 mM EDTA, 0.5% NP-40, 0.5% sodium deoxycholate) and one time with TE plus NaCl, followed by elution and reverse crosslinking in 210 uL of 1% SDS in TE overnight at 65°C. 200 uL of uncrosslinked material was treated with RNase A for 2 hours, proteinase K for 2 hours, and extracted 2 times with phenol chloroform isoamyl alcohol. This was followed by ethanol precipitation with a glycogen coprecipitant, 80% ethanol wash and final resuspension in TE. Nucleic acid yield was determined via PicoGreen (Invitrogen). Adapter ligation and size selection (200-400 bp) were performed using a Beckman Coulter SPRI TE nucleic acid extractor. Fragments were PCR amplified for 13 cycles followed by sequencing on an Illumina HiSeq 2000 system.

3.6.11 ChIP-seq analysis pipeline

Single end 40 bp reads were aligned to the mouse genome (mm9) using Bowtie (Langmead et al., 2009). Sequences were extended +200 bp and allocated in 25-bp

bins. Input DNA was used as a background model. A Poissonian model was used to determine statistically enriched bins with a P-value threshold set at $1x10^{-12}$ as described previously (Marson et al., 2008). Genomic browser tracks were generated using the Integrated Genome Viewer (Robinson et al., 2011).

Chapter 4

Summary, Future Directions, and Perspective

Our studies have investigated the organization and regulation of chromatin in the context of a differentiating cardiovascular cell. We have utilized embryonic stem cell differentiation as a model system, which has several advantages. First, ES cells can be expanded easily and indefinitely, which allows for the isolation of large numbers of relevant cell types. This is important given that standard techniques used to identify genomic regions occupied by histone modifications or DNA-binding factors require several million cells. Secondly, highly efficient differentiation protocols exist which can limit cell heterogeneity. Our studies have demonstrated that directed differentiation can indeed lead to cultures highly enriched for cardiomyocytes (Chapter 2.3.1). Lastly, ES cells differentiate in a dish, and so multiple stages of differentiation can be isolated or manipulated with relative ease. This was utilized in our studies to generate genome-wide maps of chromatin modifications for multiple stages of cardiac differentiation (Chapter 2.3.2-2.3.3) or to mediate deletion of *Brg1* at multiple timepoints during differentiation (Chapter 3.3.2).

4.1 A bird's-eye view of chromatin and transcription in cardiac differentiation

Our studies describe a global and dynamic chromatin landscape during cardiac differentiation by use of multiple histone modification maps, which can collectively identify important chromatin states, such as active, repressed, or poised TSSs and enhancers. We demonstrate that utilizing these maps can provide important insights into how transcription is likely controlled at both the global and molecular level. For instance, we have identified relationships between gene expression patterns and chromatin patterns by clustering genes based on these criteria and looking for statistical enrichment (Figure 2.9b). This analysis provides numerous insights as to how gene expression is likely regulated at the level of chromatin. For example, cardiomyocyte-specific gene expression is associated with several clearly distinct chromatin patterns, and interestingly, these chromatin distinctions segregate genes of different functional classes. Thus, while many different types of genes must be activated upon cardiomyocyte differentiation, many modes of chromatin regulation exist to achieve this activation, and genes within a given class tend to rely on similar modes. The central factors that coordinate these modes may be important linchpins for congenital or adult-onset heart disease mediated by dysregulation of specific gene classes.

We also observe a novel chromatin pattern that occurs at cardiac muscle genes. These genes show enrichment for H3K4me1 prior to gene activation and enrichment of H3K4me3 (Chapter 2.3.5). It is tempting to speculate that H3K4me1 may be functioning analogously to that described for enhancers. However, many outstanding questions

remain. For instance, is H3K4me1 functionally important in the activation of these genes? While H3K4me1 has been proposed to poise enhancers for subsequent activation, little functional evidence exists that this is the function of H3K4me1 at enhancers. Similarly for TSSs, is early enrichment of H3K4me1 an essential, regulated step in the activation of these genes or is it simply the result of low levels of activity of a histone methyltransferase which will eventually deposit H3K4me3? In either case, it is important to note that H3K4me1 enrichment likely reflects changes in the accessibility of chromatin which will plausibly facilitate transcriptional activation. However, the functional importance of H3K4me1 remains to be determined. Experiments to identify the histone methyltransferase involved and to disrupt its function will be critical to understand this phenomenon.

Through chromatin signature, we identified large numbers of putative enhancers, which we utilized to build potential transcription factor regulatory networks based on predicted transcription factor binding events (Chapter 2.3.6, 2.3.8). From this analysis, we identified *Gata4* and *Meis1* as likely to cooperate in regulating many common genes and demonstrate these factors synergistically activate enhancers in a reporter assay. However, additional experiments are needed to fully appreciate this interaction. Future directions should aim to determine whether *Gata4* and *Meis1* (as well as other factors in these two TF families) physically interact, and if so, whether this interaction affects DNA-binding and transcriptional activation. Furthermore, *Gata4* and *Meis1* are bona fide or potential causes of CHDs, respectively (Garg et al., 2003, Stankunas et al., 2008, Pfeufer et al., 2010). It will be informative to determine if these two factors genetically interact and if compound heterozygotes for each gene have cardiovascular defects.

It should be emphasized that the work described here represents only a building

block to the goal of a comprehensive picture of the chromatin landscape of cardiac differentiation. Importantly, our studies identify the genome-wide occupancy of only a handful of potential histone modifications. Unstudied modifications may represent additional modes of chromatin regulation not mentioned here. In addition, neighboring histone modifications within the same histone tail, nucleosome, or genomic region may provide context to the modifications described in our work, fine-tuning or altering their function and the chromatin states they represent. In addition, identifying the occupancy of important regulatory proteins such as chromatin regulatory complexes, histone variants, and cardiac transcription factors should complement this work. Indeed, many essential transcription factors that function at distinct steps of cardiac differentiation are already identified, but we still do not know what genes are regulated by these factors and how these TFs fraternize genome-wide or within enhancers. Genome-wide maps of multiple TFs will improve our ability to construct transcriptional regulatory networks and will further accelerate the building of a correct regulatory hierarchy in cardiac differentiation. Moreover, transcription factor collectives, which have been described in *Drosophila* cardiogenesis (Junion et al., 2012), may reveal insights into the multiple modes of chromatin regulation that we have observed. In conclusion, our study represents an important, but incomplete, step in understanding chromatin during cardiogenesis.

4.2 Temporal analysis of *Brg1* function reveals novel aspects of its role in cardiogenesis

We used an ES cell differentiation model along with temporally-inducible *Brg1* loss of

function to study stage-specific roles for *Brg1* in cardiac differentiation, and by doing so have uncovered a surprising function for *Brg1* in the regulation of gene repression and H3K27me3 levels (Chapter 3.3.3, 3.3.5). This approach was important for a number of reasons. First, *Brg1* is essential for very early embryonic development and so studies investigating its cardiac function *in vivo* have depended on tissue-specific Cre lines. However, these Cre lines mediate deletion after the initiation of cardiac differentiation. This makes inducible deletion of *Brg1* during directed differentiation of ES cells, where *Brg1* can be deleted at any point, a nice complementary model to these studies. Secondly, *Brg1* is required for ES cell maintenance and pluripotency. Thus, an inducible system is essential to allow for long term culture of these cells. Finally, the directed differentiation approach used in the work described here is variable and optimal conditions for efficient differentiation vary between cell lines. Using a tamoxifen-inducible deletion system limits variability due to these confounding effects, essentially allowing for both control and treatment groups to be matched for each differentiation. This is important because otherwise variability in differentiation efficacy might mask subtle defects in cardiac differentiation or make analysis of a phenotype challenging. Given these facts, we feel such an inducible system should be used whenever possible for loss of function studies in directed differentiation.

Our analysis of *Brg1* function has provided many interesting insights into how this factor might regulate cardiac differentiation at multiple stages. However, there are many outstanding problems to address. First and most importantly, we do not have direct evidence linking *Brg1* to any of the genes misregulated during mesoderm induction. This is crucial, as many of the genes significantly altered by *Brg1* loss are likely to be due to indirect effects. Chromatin immunoprecipitation experiments represent the best

way to provide evidence for direct regulation. These experiments are particularly important for supporting a direct role for *Brg1* in the regulation of Polycomb-mediated repression, as this is a role not broadly supported by the literature. Although utilized extensively in the past, antibodies against endogenous *Brg1* have not performed well in ChIP assays in our hands, and this has been seen by others as well (Dr. Gerald Crabtree, personal communication). This may be due to differences in antibody preparations or lots. However, ChIP for chromatin remodelers has been suggested to be inherently challenging due to the transient association of chromatin remodelers with the chromatin (Gelbart et al., 2005). Therefore, we feel occupancy data based on tagged forms of *Brg1* represent the best current, workable approach.

Given the overlap between our study and others who identify Polycomb target genes derepressed by loss of *Brg1* (Ho et al., 2011), we believe it is likely *Brg1* is playing some direct role. While it must be emphasized that evidence supporting this possibility depends on ChIP experiments, it is possible to speculate as to how *Brg1* might function in this context. *Brg1* has been demonstrated to directly associate with many different complexes with histone deacetylase (HDAC) activity and histone deacetylation is required for silencing of Polycomb target genes (Underhill et al., 2000, Sif et al., 2001, Pal et al., 2003, Bilodeau et al., 2006, Reynolds et al., 2011, Ren and Kerppola, 2011). Lysine 27 of histone H3 can be acetylated or methylated, and indeed, a dynamic equilibrium between these two marks has been proposed (Pasini et al., 2010). Thus, the potential for *Brg1* to modulate Polycomb repression through its association with HDACs is an intriguing hypothesis and remains to be further explored.

Perhaps most curious is the requirement for a chromatin remodeler in HDAC-containing complexes or for Polycomb function. Indeed, HDAC complexes associate

with other chromatin remodelers such as the Mi-2 proteins (Zhang et al., 1998). One possibility is that chromatin remodeling factors modulate DNA accessibility such that these repressive complexes can bind. However, why might these two activities (histone modification activity and chromatin remodeling activity) reside in a single complex? It has been proposed that chromatin regulators nearly unfailingly function as large protein complexes to interpret complicated signals embedded in the chromatin template (Wu et al., 2009). The existence of these complexes may also tether two activities that need to occur in a cooperative or coordinated fashion. While it is not clear how these activities might cooperate on the chromatin template, perhaps nucleosome sliding facilitates spreading of histone modifications to neighboring nucleosomes or remodelers equally space modified nucleosomes to facilitate their compaction into higher-order chromatin. Future studies should shed new light on these outstanding questions.

4.3 Concluding remarks

During the span of a Ph.D., scientific discovery seems to proceed with disorienting speed. For my Ph.D. dissertation, I have worked in collaboration with many to better understand the regulation of chromatin and transcription that occurs within the nucleus of a differentiating cardiac cell. We believe these studies have led to important developments and insights, which I hope will contribute to some small way to improve the quality of life of those suffering from congenital and adult-onset heart disease in the future. There are many questions this work leaves unanswered and new questions have emerged. These have been left to future studies to answer. It is my sincere hope that the work described here will serve as a useful resource or provide insights that may help

further unravel the mechanisms which regulate how cells differentiate and assemble into functional organs like the heart. There is much work to be done.

References

1. Adams CC and Workman JL (1995). Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Mol Cell Biol* 15: 1405-1421
2. Agalioti T, Lomvardas S, Parekh B, Yie J, et al. (2000). Ordered recruitment of chromatin modifying and general transcription factors to the IFN-beta promoter. *Cell* 103: 667-678
3. Agbottah E, Deng L, Dannenberg LO, Pumfery A, et al. (2006). Effect of SWI/SNF chromatin remodeling complex on HIV-1 Tat activated transcription. *Retrovirology* 3: 48
4. Anders S and Huber W (2010). Differential expression analysis for sequence count data. *Genome Biol* 11: R106
5. Ariumi Y, Serhan F, Turelli P, Telenti A, et al. (2006). The integrase interactor 1 (INI1) proteins facilitate Tat-mediated human immunodeficiency virus type 1 transcription. *Retrovirology* 3: 47
6. Armstrong JA, Papoulas O, Daubresse G, Sperling AS, et al. (2002). The Drosophila BRM complex facilitates global transcription by RNA polymerase II. *EMBO J* 21: 5245-5254
7. Asp P, Blum R, Vethantham V, Parisi F, et al. (2011). Genome-wide remodeling of the epigenetic landscape during myogenic differentiation. *Proc Natl Acad Sci U S A* 108: E149-158
8. Banach K, Halbach MD, Hu P, Hescheler J, et al. (2003). Development of electrical activity in cardiac myocyte aggregates derived from mouse embryonic stem cells. *Am J Physiol Heart Circ Physiol* 284: H2114-2123
9. Barron M, Gao M and Lough J (2000). Requirement for BMP and FGF signaling during cardiogenic induction in non-precardiac mesoderm is specific, transient, and cooperative. *Dev Dyn* 218: 383-393
10. Barski A, Cuddapah S, Cui K, Roh TY, et al. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* 129: 823-837
11. Basson CT, Bachinsky DR, Lin RC, Levi T, et al. (1997). Mutations in human

TBX5 cause limb and cardiac malformation in Holt-Oram syndrome. *Nat Genet* 15: 30-35

12. Beddington RS, Rashbass P and Wilson V (1992). Brachyury--a gene affecting mouse gastrulation and early organogenesis. *Dev Suppl* 157-165
13. Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, et al. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* 120: 169-181
14. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125: 315-326
15. Bertrand N, Roux M, Ryckebusch L, Niederreither K, et al. (2011). Hox genes define distinct progenitor sub-domains within the second heart field. *Dev Biol* 353: 266-274
16. Bi W, Drake CJ and Schwarz JJ (1999). The transcription factor MEF2C-null mouse exhibits complex vascular malformations and reduced cardiac expression of angiopoietin 1 and VEGF. *Dev Biol* 211: 255-267
17. Bilodeau S, Vallette-Kasic S, Gauthier Y, Figarella-Branger D, et al. (2006). Role of Brg1 and HDAC2 in GR trans-repression of the pituitary POMC gene and misexpression in Cushing disease. *Genes Dev* 20: 2871-2886
18. Blow MJ, McCulley DJ, Li Z, Zhang T, et al. (2010). ChIP-Seq identification of weakly conserved heart enhancers. *Nat Genet* 42: 806-810
19. Boheler KR, Czyz J, Tweedie D, Yang HT, et al. (2002). Differentiation of pluripotent embryonic stem cells into cardiomyocytes. *Circ Res* 91: 189-201
20. Bondue A, Lapouge G, Paulissen C, Semeraro C, et al. (2008). Mesp1 acts as a master regulator of multipotent cardiovascular progenitor specification. *Cell Stem Cell* 3: 69-84
21. Boyer LA, Plath K, Zeitlinger J, Brambrink T, et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441: 349-353
22. Bruneau BG, Nemer G, Schmitt JP, Charron F, et al. (2001). A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. *Cell* 106: 709-721
23. Bruneau BG (2008). The developmental genetics of congenital heart disease.

Nature 451: 943-948

24. Buckingham M, Meilhac S and Zaffran S (2005). Building the mammalian heart from two sources of myocardial cells. *Nat Rev Genet* 6: 826-835
25. Bultman S, Gebuhr T, Yee D, La Mantia C, et al. (2000). A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. *Mol Cell* 6: 1287-1295
26. Burns LG and Peterson CL (1997). The yeast SWI-SNF complex facilitates binding of a transcriptional activator to nucleosomal sites in vivo. *Mol Cell Biol* 17: 4811-4819
27. Cabili MN, Trapnell C, Goff L, Koziol M, et al. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* 25: 1915-1927
28. Cai CL, Liang X, Shi Y, Chu PH, et al. (2003). Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. *Dev Cell* 5: 877-889
29. Cairns BR (2007). Chromatin remodeling: insights and intrigue from single-molecule studies. *Nat Struct Mol Biol* 14: 989-996
30. Cao R, Wang L, Wang H, Xia L, et al. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* 298: 1039-1043
31. Chang CP, Neilson JR, Bayle JH, Gestwicki JE, et al. (2004). A field of myocardial-endocardial NFAT signaling underlies heart valve morphogenesis. *Cell* 118: 649-663
32. Chang CP and Bruneau BG (2012). Epigenetics and cardiovascular development. *Annu Rev Physiol* 74: 41-68
33. Chen L, Ma Y, Kim EY, Yu W, et al. (2012). Conditional ablation of Ezh2 in murine hearts reveals its essential roles in endocardial cushion formation, cardiomyocyte proliferation and survival. *PLoS One* 7: e31005
34. Chi T (2004). A BAF-centred view of the immune system. *Nat Rev Immunol* 4: 965-977
35. Chi TH, Wan M, Lee PP, Akashi K, et al. (2003). Sequential roles of Brg, the ATPase subunit of BAF chromatin remodeling complexes, in thymocyte development. *Immunity* 19: 169-182
36. Cirillo LA and Zaret KS (1999). An early developmental transcription factor

complex that is more stable on nucleosome core particles than on free DNA. *Mol Cell* 4: 961-969

37. Cirillo LA, Lin FR, Cuesta I, Friedman D, et al. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell* 9: 279-289
38. Cohen ED, Wang Z, Lepore JJ, Lu MM, et al. (2007). Wnt/beta-catenin signaling promotes expansion of Isl-1-positive cardiac progenitor cells through regulation of FGF signaling. *J Clin Invest* 117: 1794-1804
39. Creemers EE, Sutherland LB, McAnally J, Richardson JA, et al. (2006). Myocardin is a direct transcriptional target of Mef2, Tead and Foxo proteins during cardiovascular development. *Development* 133: 4245-4256
40. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* 107: 21931-21936
41. Crowley MA, Conlin LK, Zackai EH, Deardorff MA, et al. (2010). Further evidence for the possible role of MEIS2 in the development of cleft palate and cardiac septum. *Am J Med Genet A* 152A: 1326-1327
42. Cui K, Zang C, Roh TY, Schones DE, et al. (2009). Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell Stem Cell* 4: 80-93
43. Czermin B, Melfi R, McCabe D, Seitz V, et al. (2002). Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* 111: 185-196
44. Davidson EH (2010). Emerging properties of animal gene regulatory networks. *Nature* 468: 911-920
45. de Napoles M, Mermoud JE, Wakao R, Tang YA, et al. (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev Cell* 7: 663-676
46. Debril MB, Gelman L, Fayard E, Annicotte JS, et al. (2004). Transcription factors and nuclear receptors interact with the SWI/SNF complex through the BAF60c subunit. *J Biol Chem* 279: 16677-16686
47. Delgado-Olguin P, Huang Y, Li X, Christodoulou D, et al. (2012). Epigenetic repression of cardiac progenitor gene expression by Ezh2 is required for

postnatal cardiac homeostasis. *Nat Genet* 44: 343-347

48. Durocher D, Chen CY, Ardati A, Schwartz RJ, et al. (1996). The atrial natriuretic factor promoter is a downstream target for Nkx-2.5 in the myocardium. *Mol Cell Biol* 16: 4648-4655
49. Durocher D, Charron F, Warren R, Schwartz RJ, et al. (1997). The cardiac transcription factors Nkx2-5 and GATA-4 are mutual cofactors. *EMBO J* 16: 5687-5696
50. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473: 43-49
51. Evans SM, Yelon D, Conlon FL and Kirby ML (2010). Myocardial lineage development. *Circ Res* 107: 1428-1444
52. Faust C, Schumacher A, Holdener B and Magnuson T (1995). The eed mutation disrupts anterior mesoderm production in mice. *Development* 121: 273-285
53. Forcales SV, Albini S, Giordani L, Malecova B, et al. (2012). Signal-dependent incorporation of MyoD-BAF60c into Brg1-based SWI/SNF chromatin-remodelling complex. *EMBO J* 31: 301-316
54. Francis NJ, Kingston RE and Woodcock CL (2004). Chromatin compaction by a polycomb group protein complex. *Science* 306: 1574-1577
55. Fry CJ and Peterson CL (2001). Chromatin remodeling enzymes: who's on first? *Curr Biol* 11: R185-197
56. Gadue P, Huber TL, Nostro MC, Kattman S, et al. (2005). Germ layer induction from embryonic stem cells. *Exp Hematol* 33: 955-964
57. Gao X, Tate P, Hu P, Tjian R, et al. (2008). ES cell pluripotency and germ-layer formation require the SWI/SNF chromatin remodeling component BAF250a. *Proc Natl Acad Sci U S A* 105: 6656-6661
58. Garg V, Kathiriya IS, Barnes R, Schluterman MK, et al. (2003). GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature* 424: 443-447
59. Gebuhr TC, Kovalev GI, Bultman S, Godfrey V, et al. (2003). The role of Brg1, a catalytic subunit of mammalian chromatin-remodeling complexes, in T cell development. *J Exp Med* 198: 1937-1949
60. Gelbart ME, Bachman N, Delrow J, Boeke JD, et al. (2005). Genome-wide identification of Isw2 chromatin-remodeling targets by localization of a

catalytically inactive mutant. *Genes Dev* 19: 942-954

61. Ghosh TK, Packham EA, Bonser AJ, Robinson TE, et al. (2001). Characterization of the TBX5 binding site and analysis of mutations that cause Holt-Oram syndrome. *Hum Mol Genet* 10: 1983-1994
62. Golob JL, Kumar RM, Guenther MG, Pabon LM, et al. (2011). Evidence that gene activation and silencing during stem cell differentiation requires a transcriptionally paused intermediate state. *PLoS One* 6: e22416
63. Griffin CT, Brennan J and Magnuson T (2008). The chromatin-remodeling enzyme BRG1 plays an essential role in primitive erythropoiesis and vascular development. *Development* 135: 493-500
64. Guan K, Furst DO and Wobus AM (1999). Modulation of sarcomere organization during embryonic stem cell-derived cardiomyocyte differentiation. *Eur J Cell Biol* 78: 813-823
65. Guenther MG, Levine SS, Boyer LA, Jaenisch R, et al. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* 130: 77-88
66. Han M and Grunstein M (1988). Nucleosome loss activates yeast downstream promoters in vivo. *Cell* 55: 1137-1145
67. Hang CT, Yang J, Han P, Cheng HL, et al. (2010). Chromatin regulation by Brg1 underlies heart muscle development and disease. *Nature* 466: 62-67
68. Hargreaves DC and Crabtree GR (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res* 21: 396-420
69. Hawkins RD, Hon GC, Yang C, Antosiewicz-Bourget JE, et al. (2011). Dynamic chromatin states in human ES cells reveal potential regulatory sequences and genes involved in pluripotency. *Cell Res* 21: 1393-1409
70. He A, Kong SW, Ma Q and Pu WT (2011). Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. *Proc Natl Acad Sci U S A* 108: 5632-5637
71. He A, Ma Q, Cao J, von Gise A, et al. (2012). Polycomb repressive complex 2 regulates normal development of the mouse heart. *Circ Res* 110: 406-415
72. He HH, Meyer CA, Shin H, Bailey ST, et al. (2010). Nucleosome dynamics define transcriptional enhancers. *Nat Genet* 42: 343-347
73. Heintzman ND, Stuart RK, Hon G, Fu Y, et al. (2007). Distinct and predictive

chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 39: 311-318

74. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459: 108-112

75. Heng JC, Feng B, Han J, Jiang J, et al. (2010). The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells. *Cell Stem Cell* 6: 167-174

76. Hiroi Y, Kudoh S, Monzen K, Ikeda Y, et al. (2001). Tbx5 associates with Nkx2-5 and synergistically promotes cardiomyocyte differentiation. *Nat Genet* 28: 276-280

77. Ho L, Jothi R, Ronan JL, Cui K, et al. (2009). 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* 106: 5187-5191

78. Ho L, Ronan JL, Wu J, Staahl BT, et al. (2009). 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

79. Ho L and Crabtree GR (2010). Chromatin remodelling during development. *Nature* 463: 474-484

80. Ho L, Miller EL, Ronan JL, Ho WQ, et al. (2011). esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. *Nat Cell Biol* 13: 903-913

81. Hoffman JI and Kaplan S (2002). The incidence of congenital heart disease. *J Am Coll Cardiol* 39: 1890-1900

82. Hsiao EC, Yoshinaga Y, Nguyen TD, Musone SL, et al. (2008). Marking embryonic stem cells with a 2A self-cleaving peptide: a NKX2-5 emerald GFP BAC reporter. *PLoS One* 3: e2532

83. Ieda M, Fu JD, Delgado-Olguin P, Vedantham V, et al. (2010). Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* 142: 375-386

84. Indra AK, Dupe V, Bornert JM, Messaddeq N, et al. (2005). Temporally controlled targeted somatic mutagenesis in embryonic surface ectoderm and fetal epidermal keratinocytes unveils two distinct developmental functions of BRG1 in

limb morphogenesis and skin barrier formation. *Development* 132: 4533-4544

85. Junion G, Spivakov M, Girardot C, Braun M, et al. (2012). A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell* 148: 473-486
86. Kagalwala MN, Glaus BJ, Dang W, Zofall M, et al. (2004). Topography of the ISW2-nucleosome complex: insights into nucleosome spacing and chromatin remodeling. *EMBO J* 23: 2092-2104
87. Kanamori M, Konno H, Osato N, Kawai J, et al. (2004). A genome-wide and nonredundant mouse transcription factor database. *Biochem Biophys Res Commun* 322: 787-793
88. Kassabov SR, Zhang B, Persinger J and Bartholomew B (2003). SWI/SNF unwraps, slides, and rewraps the nucleosome. *Mol Cell* 11: 391-403
89. Kattman SJ, Huber TL and Keller GM (2006). Multipotent flk-1+ cardiovascular progenitor cells give rise to the cardiomyocyte, endothelial, and vascular smooth muscle lineages. *Dev Cell* 11: 723-732
90. Kattman SJ, Witty AD, Gagliardi M, Dubois NC, et al. (2011). Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. *Cell Stem Cell* 8: 228-240
91. Keller G (2005). Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev* 19: 1129-1155
92. Kelly RG, Brown NA and Buckingham ME (2001). The arterial pole of the mouse heart forms from Fgf10-expressing cells in pharyngeal mesoderm. *Dev Cell* 1: 435-440
93. Kennison JA and Tamkun JW (1988). Dosage-dependent modifiers of polycomb and antennapedia mutations in Drosophila. *Proc Natl Acad Sci U S A* 85: 8136-8140
94. Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, et al. (2011). Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* 471: 480-485
95. Kim TK, Hemberg M, Gray JM, Costa AM, et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465: 182-187
96. Kim Y, Fedoriw AM and Magnuson T (2012). An essential role for a mammalian SWI/SNF chromatin-remodeling complex during male meiosis. *Development* 139: 1133-1140

97. King IN, Qian L, Liang J, Huang Y, et al. (2011). A genome-wide screen reveals a role for microRNA-1 in modulating cardiac cell polarity. *Dev Cell* 20: 497-510
98. Kornberg RD (1974). Chromatin structure: a repeating unit of histones and DNA. *Science* 184: 868-871
99. Kornberg RD and Lorch Y (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* 98: 285-294
100. Kouzarides T (2007). Chromatin modifications and their function. *Cell* 128: 693-705
101. Ku M, Koche RP, Rheinbay E, Mendenhall EM, et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* 4: e1000242
102. Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, et al. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev* 16: 2893-2905
103. Kuzmichev A, Zhang Y, Erdjument-Bromage H, Tempst P, et al. (2002). Role of the Sin3-histone deacetylase complex in growth regulation by the candidate tumor suppressor p33(ING1). *Mol Cell Biol* 22: 835-848
104. Laflamme MA, Chen KY, Naumova AV, Muskheli V, et al. (2007). Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. *Nat Biotechnol* 25: 1015-1024
105. Lamba DA, Hayes S, Karl MO and Reh T (2008). Baf60c is a component of the neural progenitor-specific BAF complex in developing retina. *Dev Dyn* 237: 3016-3023
106. Lange M, Kaynak B, Forster UB, Tonjes M, et al. (2008). Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex. *Genes Dev* 22: 2370-2384
107. Langmead B, Trapnell C, Pop M and Salzberg SL (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10: R25
108. Lee ER, Murdoch FE and Fritsch MK (2007). High histone acetylation and decreased polycomb repressive complex 2 member levels regulate gene specific transcriptional changes during early embryonic stem cell differentiation induced by retinoic acid. *Stem Cells* 25: 2191-2199

109. Lee HS, Park JH, Kim SJ, Kwon SJ, et al. (2010). A cooperative activation loop among SWI/SNF, gamma-H2AX and H3 acetylation for DNA double-strand break repair. *EMBO J* 29: 1434-1445

110. Lee TI, Jenner RG, Boyer LA, Guenther MG, et al. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 125: 301-313

111. Lee TI, Johnstone SE and Young RA (2006). Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat Protoc* 1: 729-748

112. Lessard J, Wu JI, Ranish JA, Wan M, et al. (2007). An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron* 55: 201-215

113. Li QY, Newbury-Ecob RA, Terrett JA, Wilson DI, et al. (1997). Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. *Nat Genet* 15: 21-29

114. Lian X, Hsiao C, Wilson G, Zhu K, et al. (2012). Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *Proc Natl Acad Sci U S A*

115. Lickert H, Takeuchi JK, Von Both I, Walls JR, et al. (2004). Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. *Nature* 432: 107-112

116. Lien WH, Guo X, Polak L, Lawton LN, et al. (2011). Genome-wide maps of histone modifications unwind in vivo chromatin states of the hair follicle lineage. *Cell Stem Cell* 9: 219-232

117. Lindsay EA, Vitelli F, Su H, Morishima M, et al. (2001). Tbx1 haploinsufficieny in the DiGeorge syndrome region causes aortic arch defects in mice. *Nature* 410: 97-101

118. Loh YH, Wu Q, Chew JL, Vega VB, et al. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* 38: 431-440

119. Lomvardas S and Thanos D (2001). Nucleosome sliding via TBP DNA binding in vivo. *Cell* 106: 685-696

120. Lorch Y, LaPointe JW and Kornberg RD (1987). Nucleosomes inhibit the initiation

of transcription but allow chain elongation with the displacement of histones. *Cell* 49: 203-210

121. Lorch Y, Zhang M and Kornberg RD (1999). Histone octamer transfer by a chromatin-remodeling complex. *Cell* 96: 389-392

122. Luger K, Mader AW, Richmond RK, Sargent DF, et al. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* 389: 251-260

123. Lyons I, Parsons LM, Hartley L, Li R, et al. (1995). Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. *Genes Dev* 9: 1654-1666

124. Mahmoudi T, Parra M, Vries RG, Kauder SE, et al. (2006). The SWI/SNF chromatin-remodeling complex is a cofactor for Tat transactivation of the HIV promoter. *J Biol Chem* 281: 19960-19968

125. Maltsev VA, Rohwedel J, Hescheler J and Wobus AM (1993). Embryonic stem cells differentiate in vitro into cardiomyocytes representing sinusnodal, atrial and ventricular cell types. *Mech Dev* 44: 41-50

126. Marson A, Levine SS, Cole MF, Frampton GM, et al. (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* 134: 521-533

127. May D, Blow MJ, Kaplan T, McCulley DJ, et al. (2012). Large-scale discovery of enhancers from human heart tissue. *Nat Genet* 44: 89-93

128. McLean CY, Bristor D, Hiller M, Clarke SL, et al. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 28: 495-501

129. Mikkelsen TS, Ku M, Jaffe DB, Issac B, et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448: 553-560

130. Miller SA, Huang AC, Miazgowicz MM, Brassil MM, et al. (2008). Coordinated but physically separable interaction with H3K27-demethylase and H3K4-methyltransferase activities are required for T-box protein-mediated activation of developmental gene expression. *Genes Dev* 22: 2980-2993

131. Miller SA, Mohn SE and Weinmann AS (2010). Jmjd3 and UTX play a demethylase-independent role in chromatin remodeling to regulate T-box family member-dependent gene expression. *Mol Cell* 40: 594-605

132. Mjaatvedt CH, Nakaoka T, Moreno-Rodriguez R, Norris RA, et al. (2001). The outflow tract of the heart is recruited from a novel heart-forming field. *Dev Biol*

238: 97-109

133. Morrison AJ and Shen X (2009). Chromatin remodelling beyond transcription: the INO80 and SWR1 complexes. *Nat Rev Mol Cell Biol* 10: 373-384

134. Muller J, Hart CM, Francis NJ, Vargas ML, et al. (2002). Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell* 111: 197-208

135. Murry CE and Keller G (2008). Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell* 132: 661-680

136. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, et al. (2010). Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* 42: 790-793

137. Nimura K, Ura K, Shiratori H, Ikawa M, et al. (2009). A histone H3 lysine 36 trimethyltransferase links Nkx2-5 to Wolf-Hirschhorn syndrome. *Nature* 460: 287-291

138. Niwa H (2010). Mouse ES cell culture system as a model of development. *Dev Growth Differ* 52: 275-283

139. Nix DA, Courdy SJ and Boucher KM (2008). Empirical methods for controlling false positives and estimating confidence in ChIP-Seq peaks. *BMC Bioinformatics* 9: 523

140. O'Carroll D, Erhardt S, Pagani M, Barton SC, et al. (2001). The polycomb-group gene Ezh2 is required for early mouse development. *Mol Cell Biol* 21: 4330-4336

141. Ostlund Farrants AK, Blomquist P, Kwon H and Wrange O (1997). Glucocorticoid receptor-glucocorticoid response element binding stimulates nucleosome disruption by the SWI/SNF complex. *Mol Cell Biol* 17: 895-905

142. Pal S, Yun R, Datta A, Lacomis L, et al. (2003). mSin3A/histone deacetylase 2- and PRMT5-containing Brg1 complex is involved in transcriptional repression of the Myc target gene cad. *Mol Cell Biol* 23: 7475-7487

143. Pan G, Tian S, Nie J, Yang C, et al. (2007). Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* 1: 299-312

144. Pasini D, Bracken AP, Jensen MR, Lazzerini Denchi E, et al. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J* 23: 4061-4071

145. Pasini D, Bracken AP, Hansen JB, Capillo M, et al. (2007). The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol Cell Biol* 27: 3769-3779

146. Pasini D, Malatesta M, Jung HR, Walfridsson J, et al. (2010). Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. *Nucleic Acids Res* 38: 4958-4969

147. Pauli A, Rinn JL and Schier AF (2011). Non-coding RNAs as regulators of embryogenesis. *Nat Rev Genet* 12: 136-149

148. Pfeufer A, van Noord C, Marciante KD, Arking DE, et al. (2010). Genome-wide association study of PR interval. *Nat Genet* 42: 153-159

149. Pokholok DK, Harbison CT, Levine S, Cole M, et al. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* 122: 517-527

150. Pollard, KS and van der Laan, MJ (2005). Cluster Analysis of Genomic Data with Applications in R. *U.C. Berkeley Division of Biostatistics Working Paper Series.* Working Paper 167.http://biostats.bepress.com/ucbbiostat/paper167.

151. Racki LR, Yang JG, Naber N, Partensky PD, et al. (2009). The chromatin remodeller ACF acts as a dimeric motor to space nucleosomes. *Nature* 462: 1016-1021

152. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, et al. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470: 279-283

153. Rahl PB, Lin CY, Seila AC, Flynn RA, et al. (2010). c-Myc regulates transcriptional pause release. *Cell* 141: 432-445

154. Randazzo FM, Khavari P, Crabtree G, Tamkun J, et al. (1994). brg1: a putative murine homologue of the Drosophila brahma gene, a homeotic gene regulator. *Dev Biol* 161: 229-242

155. Reifers F, Walsh EC, Leger S, Stainier DY, et al. (2000). Induction and differentiation of the zebrafish heart requires fibroblast growth factor 8 (fgf8/acerebellar). *Development* 127: 225-235

156. Reisman D, Glaros S and Thompson EA (2009). The SWI/SNF complex and cancer. *Oncogene* 28: 1653-1668

157. Ren X and Kerppola TK (2011). REST interacts with Cbx proteins and regulates polycomb repressive complex 1 occupancy at RE1 elements. *Mol Cell Biol* 31: 2100-2110

158. Reyes JC, Barra J, Muchardt C, Camus A, et al. (1998). Altered control of cellular proliferation in the absence of mammalian brahma (SNF2alpha). *EMBO J* 17: 6979-6991

159. Reynolds N, Salmon-Divon M, Dvinge H, Hynes-Allen A, et al. (2012). NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. *EMBO J* 31: 593-605

160. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, et al. (2011). Integrative genomics viewer. *Nat Biotechnol* 29: 24-26

161. Robinson PJ and Rhodes D (2006). Structure of the '30 nm' chromatin fibre: a key role for the linker histone. *Curr Opin Struct Biol* 16: 336-343

162. Ruthenburg AJ, Allis CD and Wysocka J (2007). Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol Cell* 25: 15-30

163. Saga Y, Kitajima S and Miyagawa-Tomita S (2000). Mesp1 expression is the earliest sign of cardiovascular development. *Trends Cardiovasc Med* 10: 345-352

164. Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, et al. (2002). Active genes are tri-methylated at K4 of histone H3. *Nature* 419: 407-411

165. Schmidt D, Wilson MD, Spyrou C, Brown GD, et al. (2009). ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods* 48: 240-248

166. Schott JJ, Benson DW, Basson CT, Pease W, et al. (1998). Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science* 281: 108-111

167. Schubeler D, MacAlpine DM, Scalzo D, Wirbelauer C, et al. (2004). The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev* 18: 1263-1271

168. Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, et al. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* 311: 844-847

169. Sif S, Saurin AJ, Imbalzano AN and Kingston RE (2001). Purification and

characterization of mSin3A-containing Brg1 and hBrm chromatin remodeling complexes. *Genes Dev* 15: 603-618

170. Sparmann A and van Lohuizen M (2006). Polycomb silencers control cell fate, development and cancer. *Nat Rev Cancer* 6: 846-856

171. Srivastava D, Thomas T, Lin Q, Kirby ML, et al. (1997). Regulation of cardiac mesodermal and neural crest development by the bHLH transcription factor, dHAND. *Nat Genet* 16: 154-160

172. Srivastava D (2006). Making or breaking the heart: from lineage determination to morphogenesis. *Cell* 126: 1037-1048

173. Stankunas K, Hang CT, Tsun ZY, Chen H, et al. (2008). Endocardial Brg1 represses ADAMTS1 to maintain the microenvironment for myocardial morphogenesis. *Dev Cell* 14: 298-311

174. Stankunas K, Shang C, Twu KY, Kao SC, et al. (2008). Pbx/Meis deficiencies demonstrate multigenetic origins of congenital heart disease. *Circ Res* 103: 702-709

175. Stock JK, Giadrossi S, Casanova M, Brookes E, et al. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol* 9: 1428-1435

176. Strahl BD and Allis CD (2000). The language of covalent histone modifications. *Nature* 403: 41-45

177. Sun FL, Cuaycong MH and Elgin SC (2001). Long-range nucleosome ordering is associated with gene silencing in Drosophila melanogaster pericentric heterochromatin. *Mol Cell Biol* 21: 2867-2879

178. Surface LE, Thornton SR and Boyer LA (2010). Polycomb group proteins set the stage for early lineage commitment. *Cell Stem Cell* 7: 288-298

179. Takeuchi JK and Bruneau BG (2009). Directed transdifferentiation of mouse mesoderm to heart tissue by defined factors. *Nature* 459: 708-711

180. Takeuchi JK, Lou X, Alexander JM, Sugizaki H, et al. (2011). Chromatin remodelling complex dosage modulates transcription factor function in heart development. *Nat Commun* 2: 187

181. Tamkun JW, Deuring R, Scott MP, Kissinger M, et al. (1992). brahma: a regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. *Cell* 68: 561-572

182. Tan M, Luo H, Lee S, Jin F, et al. (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* 146: 1016-1028

183. Thorogood P (1997). Embryos, genes and birth defects. John Wiley & Sons

184. Treand C, du Chene I, Bres V, Kiernan R, et al. (2006). Requirement for SWI/SNF chromatin-remodeling complex in Tat-mediated activation of the HIV-1 promoter. *EMBO J* 25: 1690-1699

185. Tremethick DJ (2007). Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell* 128: 651-654

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

187. Trotter KW and Archer TK (2007). Nuclear receptors and chromatin remodeling machinery. *Mol Cell Endocrinol* 265-266: 162-167

188. Trouche D, Le Chalony C, Muchardt C, Yaniv M, et al. (1997). RB and hbrm cooperate to repress the activation functions of E2F1. *Proc Natl Acad Sci U S A* 94: 11268-11273

189. Tse C, Sera T, Wolffe AP and Hansen JC (1998). Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III. *Mol Cell Biol* 18: 4629-4638

190. Underhill C, Qutob MS, Yee SP and Torchia J (2000). A novel nuclear receptor corepressor complex, N-CoR, contains components of the mammalian SWI/SNF complex and the corepressor KAP-1. *J Biol Chem* 275: 40463-40470

191. Verzi MP, Shin H, He HH, Sulahian R, et al. (2010). Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2. *Dev Cell* 19: 713-726

192. Vissers LE, van Ravenswaaij CM, Admiraal R, Hurst JA, et al. (2004). Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat Genet* 36: 955-957

193. Waldo KL, Kumiski DH, Wallis KT, Stadt HA, et al. (2001). Conotruncal myocardium arises from a secondary heart field. *Development* 128: 3179-3188

194. Wang H, Wang L, Erdjument-Bromage H, Vidal M, et al. (2004). Role of histone H2A ubiquitination in Polycomb silencing. *Nature* 431: 873-878

195. Wang W, Cote J, Xue Y, Zhou S, et al. (1996). Purification and biochemical

heterogeneity of the mammalian SWI-SNF complex. *EMBO J* 15: 5370-5382

196. Wang W, Xue Y, Zhou S, Kuo A, et al. (1996). Diversity and specialization of mammalian SWI/SNF complexes. *Genes Dev* 10: 2117-2130

197. Wang Z, Zang C, Rosenfeld JA, Schones DE, et al. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* 40: 897-903

198. Watt AJ, Battle MA, Li J and Duncan SA (2004). GATA4 is essential for formation of the proepicardium and regulates cardiogenesis. *Proc Natl Acad Sci U S A* 101: 12573-12578

199. Wilkinson DG, Bhatt S and Herrmann BG (1990). Expression pattern of the mouse T gene and its role in mesoderm formation. *Nature* 343: 657-659

200. Wilson NK, Foster SD, Wang X, Knezevic K, et al. (2010). Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell* 7: 532-544

201. Wilson NK, Timms RT, Kinston SJ, Cheng YH, et al. (2010). Gfi1 expression is controlled by five distinct regulatory regions spread over 100 kilobases, with Scl/Tal1, Gata2, PU.1, Erg, Meis1, and Runx1 acting as upstream regulators in early hematopoietic cells. *Mol Cell Biol* 30: 3853-3863

202. Wobus AM, Wallukat G and Hescheler J (1991). Pluripotent mouse embryonic stem cells are able to differentiate into cardiomyocytes expressing chronotropic responses to adrenergic and cholinergic agents and Ca2+ channel blockers. *Differentiation* 48: 173-182

203. Wu JI, Lessard J and Crabtree GR (2009). Understanding the words of chromatin regulation. *Cell* 136: 200-206

204. Yang JG, Madrid TS, Sevastopoulos E and Narlikar GJ (2006). The chromatin-remodeling enzyme ACF is an ATP-dependent DNA length sensor that regulates nucleosome spacing. *Nat Struct Mol Biol* 13: 1078-1083

205. Zeitlinger J, Stark A, Kellis M, Hong JW, et al. (2007). RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. *Nat Genet* 39: 1512-1516

206. Zeng W, Ball AR, Jr. and Yokomori K (2010). HP1: heterochromatin binding proteins working the genome. *Epigenetics* 5: 287-292

207. Zentner GE, Tesar PJ and Scacheri PC (2011). Epigenetic signatures distinguish

multiple classes of enhancers with distinct cellular functions. *Genome Res* 21: 1273-1283

208. Zhang SS, Kim KH, Rosen A, Smyth JW, et al. (2011). Iroquois homeobox gene 3 establishes fast conduction in the cardiac His-Purkinje network. *Proc Natl Acad Sci U S A* 108: 13576-13581

209. Zhang SX, Garcia-Gras E, Wycuff DR, Marriot SJ, et al. (2005). Identification of direct serum-response factor gene targets during Me2SO-induced P19 cardiac cell differentiation. *J Biol Chem* 280: 19115-19126

210. Zhang Y, LeRoy G, Seelig HP, Lane WS, et al. (1998). The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell* 95: 279-289

211. Zhang Y, Smith CL, Saha A, Grill SW, et al. (2006). DNA translocation and loop formation mechanism of chromatin remodeling by SWI/SNF and RSC. *Mol Cell* 24: 559-568

212. Zhao Q, Wang QE, Ray A, Wani G, et al. (2009). Modulation of nucleotide excision repair by mammalian SWI/SNF chromatin-remodeling complex. *J Biol Chem* 284: 30424-30432

213. Zhao R, Watt AJ, Battle MA, Li J, et al. (2008). Loss of both GATA4 and GATA6 blocks cardiac myocyte differentiation and results in acardia in mice. *Dev Biol* 317: 614-619

214. Zhou VW, Goren A and Bernstein BE (2011). Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet* 12: 7-18

Appendix: Supplementary Tables

Supplementary Table 1

Cluster	GO (Biological Process)	# genes in cluster	# genes measured	Z-score
A	- Stem cell maintenance	12	30	5.79
	- Prophase	4	5	5.44
	- Chromosome organization involved in meiosis	7	15	4.98
B	- Nucleic acid metabolic process	314	1062	17.55
	- Cellular macromolecule biosynthetic process	175	544	14.15
	- Ribonucleoprotein complex biogenesis	30	39	12.16
C	- Actin filament-based movement	3	15	6.64
	- Serine family amino acid metabolic process	3	22	5.31
	- Regulation of cell shape	4	42	4.92
D	- Regulation of actin cytoskeleton organization	3	76	4.93
	- Phospholipid metabolic process	3	121	3.66
	- Phospholipid metabolic process	3	137	3.35
E	- Cytoskeleton organization	6	316	3.33
	- Regulation of cell projection	3	102	3.31

	organization - Taxis	3	128	2.79
F	- Transmembrane receptor protein tyrosine kinase signaling pathway - Hemostasis - Glycolysis	8 4 3	145 59 36	5.51 6.78 8.33
G	- Positive regulation of lipase activity - Activation of caspase activity - Autophagy	3 4 3	19 34 36	6.75 6.59 4.63
H	- Translation - Translational elongation - Amino acid transport	15 3 3	276 17 52	6.49 6.2 3.01
I	- Regulation of cytokinesis - Centrosome duplication - Nucleotide-excision repair	5 3 6	9 6 28	10.14 7.4 6.4
J	- Nucleic acid metabolic process - Chromosome organization - Establishment of spindle orientation	95 42 3	1062 421 7	8.95 6.59 5.34
K	- Nucleic acid metabolic process - Cell cycle - Cell division	195 68 47	1062 420 250	17.25 8.66 8.36
L	- Regionalization - Embryonic morphogenesis - Neural tube development	49 54 12	161 223 26	11.04 9.49 7.45

M	- Dolichol metabolic process	4	4	6.34
	- Cilium assembly	10	22	5.96
	- Retina homeostasis	4	5	5.49
N	- Biological regulation	398	4387	6.45
	- Cranial nerve morphogenesis	4	5	6.35
	- Homophilic cell adhesion	16	56	6.27
O	- Regulation of pigment cell differentiation	4	5	6.12
	- Gastrulation with mouth forming second	8	21	5.29
	- Cholesterol metabolic process	13	52	4.76
P	- Transition metal ion transport	3	55	3.39
	- Endocytosis	5	138	3.19
	- Response to organic cyclic compound	5	150	2.96
Q	- JAK-STAT cascade involved in growth hormone signaling pathway	3	4	7.64
	- Response to cortisol stimulus	4	8	7.02
	- Response to dexamethasone stimulus	3	5	6.75
R	- Electron transport chain	18	95	7.69
	- Hydrogen transport	9	42	5.94
	- ATP metabolic process	12	76	5.43
S	- Generation of precursor metabolites and energy	77	174	7.92
	- Oxidation-reduction process	169	504	7.55
		41	77	7.21

- Muscle system process

Supplementary Table 1. Enriched gene ontology terms for expression clusters. Top three biological process GO terms for each gene expression cluster. GO terms were ranked based on z-score. Each GO term is reported with the number of genes within the cluster belonging to the specific GO term, the number of genes belonging to the GO term measured in the RNA-seq dataset, and the z-score.

Supplementary Table 2

Cluster	GO (Biological Process)	# genes in cluster	# genes measured	Z-score
1	- Nucleic acid metabolic process	833	1062	17.7
	- Cellular macromolecule metabolic process	1816	2695	17.55
	- Cell cycle	315	419	9.44
2	- Negative regulation of calcium ion transport into cytosol	3	3	11.29
	- Developmental process	107	2055	9.66
	- Positive regulation of locomotion	14	102	7.72
3	- Regulation of NF-kabbaB import into nucleus	7	12	4.50
	- Membrane lipid metabolic process	19	57	4.32
	- Lipopolysaccharide biosynthetic process	4	6	3.78
4	- Collagen catabolic process	3	8	5.93
	- Regulation of signaling	50	837	5.75
	- Axon extension	3	9	5.54
5	- Regulation of RNA metabolic process	19	1292	4.29
	- Regulation of cellular biosynthetic process	21	1507	4.26
	- Regulation of macromolecule biosynthetic process	20	1436	4.14

6	- L-serine biosynthetic process	3	4	7.68
	- Vitamin transport	4	11	5.84
	- Branching involved in mammary gland duct morphogenesis	4	12	5.54
7	- Regulation of embryonic development	5	29	6.48
	- Epithelial tube branching involved in lung morphogenesis	3	12	6.24
	- Antigen processing and presentation	5	40	5.29
8	- Response to pain	3	8	7.71
	- Alpha-beta T cell activation	3	14	5.62
	- Regulation of smooth muscle cell migration	3	17	5.00
9	- Anatomical structure morphogenesis	78	660	10.86
	- Regionalization	31	161	10.16
	- Multicellular organismal process	109	1274	9.14
10	- Positive regulation of survival gene product expression	3	5	7.51
	- Regulation of vasodilation	4	9	7.33
	- Cyclic nucleotide metabolic process	7	27	7.03
11	- Stem cell maintenance	7	30	8.08
	- Eating behavior	4	11	7.90
	- Positive regulation of steroid biosynthetic process	3	7	7.49

12	- Kidney epithelium development	5	7	9.03
	- Renal tubule development	3	3	8.41
	- Multicellular organismal process	106	1274	8.15
13	- Catabolic process	4	657	2.94
14	- Sulfur compound metabolic process	3	92	5.54
15	- Response to other organism	3	100	3.93
	- Proteolysis	6	462	2.90
	- Oxidation-reduction process	6	504	2.65
16	- Inflammatory response	3	80	5.19
	- Immune response	4	147	4.92
	- Coenzyme metabolic process	3	127	3.88
17	- Cellular protein metabolic process	5	1603	2.24
18	- Lipid catabolic process	4	90	5.01
	- Sensor perception	4	104	4.56
	- Inflammatory response	3	80	3.88
19	- Positive regulation of immune system process	5	157	6.38
	- Regulation of immune response	4	118	5.92
	- Lymphocyte activation	4	119	5.89
20	- Cardiac muscle tissue morphogenesis	6	29	19.14
	- Muscle contraction	9	65	19.05
	- Actin filament-based movement	3	15	13.30

21	- Hemostasis	3	36	10.62
	- Chemotaxis	3	128	5.28
	- Regulation of response to stimulus	3	236	3.58
22	- Translation	5	276	4.88
	- Positive regulation of multicellular organismal process	3	178	3.58
	- Proteolysis involved in cellular protein catabolic process	3	212	3.16
23	- Regulation of the force of heart contraction	3	17	13.1
	- Regulation of calcium ion transport	3	53	7.20
	- Cellular calcium ion homeostasis	4	98	6.92
24	- Negative regulation of growth	4	100	6.63
	- Actin cytoskeleton organization	4	174	4.72
	- Heart development	3	160	3.56
25	- Visual perception	3	31	10.26
	- G-protein coupled receptor protein signaling pathway	6	218	7.25
	- Ion homeostasis	3	208	3.36
26	- Inflammatory response	4	80	5.50
	- Regulation of angiogenesis	3	49	5.38
	- Immune response	4	147	3.67
27	- Response to heat	3	40	11.01
	- Heart development	4	160	7.02
	- Response to hypoxia	3	134	5.70

28	- Calcium ion homeostasis	3	102	5.42
	- Regulation of system process	4	194	5.05
	- System process	4	330	3.51
29	- Regulation of immune effector process	3	68	3.77
	- Secretion	5	169	3.64
	-Biological adhesion	8	361	3.62
30	- Positive regulation of peptidyl-tyrosine phosphorylation	3	32	5.34
31	- Regulation of transcription, DNA-dependent	9	1249	2.92
32	- ATP synthesis coupled electron transport	4	8	37.21
	- Response to inorganic substance	5	199	8.89
	- Response to oxidative stress	4	149	8.23
33	- Translation	9	276	7.45
	- Protein deubiquitination	3	43	6.71
	- Cell maturation	3	55	5.83
34	- Translation	12	276	9.38
	- Homophilic cell adhesion	4	56	7.21
	- G-protein coupled receptor protein signaling pathway	4	218	2.91

Supplementary Table 2. Enriched gene ontology terms for chromatin clusters.

Top three biological process GO terms for each chromatin cluster. GO terms were ranked based on z-score. Each GO term is reported with the number of genes within the

cluster belonging to the specific GO term, the number of genes belonging to the GO term found in the chromatin clustering analysis, and the z-score.

Supplementary Table 3

Antigen	Species (Host)	Clonality (Isotype)	Vendor
H3K4me1	Rabbit	Polyclonal (IgG)	Abcam (ab8895)
H3K4me3	Rabbit	Polyclonal (IgG)	Millipore (07-473)
H3K27me3*	Rabbit	Polyclonal (IgG)	Millipore (07-449)
H3K27me3**	Rabbit	Polyclonal (IgG)	Millipore (17-622)
H3K27ac	Rabbit	Polyclonal (IgG)	Abcam (ab4729)
RNAP II (S5Ph)	Rabbit	Polyclonal (IgG)	Abcam (ab5131)
CD140a (PDGFRα)	Rat	Monoclonal (IgG2a,k)	eBioscience (12-1401-81)
cTnT	Mouse	Monoclonal (IgG1)	Thermo Scientific (MS-295)
BRG1	Rabbit	Polyclonal (IgG)	Santa Cruz (H88)
Myosin heavy chain	Mouse	Monoclonal (IgG2b)	DSHB (MF20)
Actin	Mouse	Monoclonal (IgM)	Millipore (CP01)
TPM	Mouse	Monoclonal (IgG1)	DSHB (CH1)
GFP	Chicken	Polyclonal (IgY)	Abcam (ab13970)

Flk1	Rat	Monoclonal (IgG2a,k)	Kind gift from Gordon Keller
MLC2a	Mouse	Monoclonal (IgG2b)	Synaptic Systems (311-011)
MLC2v	Mouse	Monoclonal (IgG2a)	Synaptic Systems (310 111)
Isl1	Mouse	Monoclonal (IgG2b)	DSHB (39.4D5)
Tbx5	Rabbit	Polyclonal	Kind gift from Susan Mackem

* - used in Chapter 2 ChIP; ** - used in Chapter 3 ChIP; DSHB – Developmental Studies Hybridoma Bank

Supplementary Table 3. List of antibodies used in study. All antibodies used in the presented studies along with species and isotype information as well as where these antibodies were obtained.

Publishing Agreement

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

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

Author Signature _____ Date 7/29/12

(This page must be signed and dated by the author and include the correct pagination – as the last numbered page number of your document.)