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

Joseph A. Wamstad,1,10 Jeffrey M. Alexander,2,3,10 Rebecca M. Truty,2 Avanti Shrikumar,1 Fugen Li,1 Kirsten E. Eilertson,2 Huiming Ding,1 John N. Wylie,2 Alexander R. Pico,2 John A. Capra,2 Genevieve Erwin,2,4 Steven J. Kattman,5 Gordon M. Keller,5 Deepak Srivastava,2,3,6,7 Stuart S. Levine,1 Katherine S. Pollard,2,8 Alisha K. Holloway,2 Laurie A. Boyer,1,* and Benoit G. Bruneau2,3,7,9,*

1Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
2Gladstone Institute of Cardiovascular Disease, San Francisco, CA 94158, USA
3Program in Biomedical Sciences
4Integrative Program in Quantitative Biology
University of California, San Francisco, CA 94158, USA
5McEwen Center for Regenerative Medicine, University Health Network, Toronto M5G 1L7, Canada
6Department of Biochemistry and Biophysics
7Department of Pediatrics
8Department of Epidemiology & Biostatistics and Institute for Human Genetics
9Cardiovascular Research Institute
University of California, San Francisco, CA 94158, USA
10These authors contributed equally to this work
*Correspondence: lboyer@mit.edu (L.A.B.), bbruneau@gladstone.ucsf.edu (B.G.B.)

http://dx.doi.org/10.1016/j.cell.2012.07.035

SUMMARY

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 preactivation 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.

INTRODUCTION

Developmental decisions during lineage commitment are precisely coordinated at the genome level as broad gene expression programs are jointly activated or repressed (Davidson, 2010).

Heart development requires the concurrent differentiation of cardiovascular cell types including endothelial cells, smooth muscle cells, and cardiomyocytes that must be organized into a complex organ. This process involves specification of pluripotent cells to mesodermal and cardiac precursors prior to terminal differentiation (Evans et al., 2010; Murry and Keller, 2008; Srivastava, 2006). Thus, heart development depends on precise 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 correlated during lineage commitment in the cardiac lineage.

Chromatin regulation is fundamental in specifying different cell types during embryonic development and in generating cellular responses to the environment. Studies in mammalian cells have shown that histone modifications are correlated with active, repressed, and poised expression states and define 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 also predict noncoding 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 poorly understood.

Emerging evidence indicates that faulty epigenetic regulation contributes to congenital heart disease (Chang and Bruneau, 2012). Mutations in the histone methyltransferase MLL2 in
humans cause congenital heart defects in Kabuki syndrome (Ng et al., 2010). 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). The H3K27 methyltransferase Ezh2 regulates gene expression programs that are important for heart development and homeostasis (Delgado-Olguin et al., 2012; He et al., 2012). In addition, epigenetic changes at cardiac-specific genes are observed during direct reprogramming of cardiac fibroblasts into cardiomyocytes (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 transcriptional 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 histone modifications at promoters to define chromatin states that accompany gene expression changes during cardiac differentiation. By using a dynamic model of lineage determination, we discovered previously unknown chromatin state transitions, including a preactivation pattern associated with a set of genes with cardiac functions. We also used chromatin marks to discover thousands of stage-specific enhancers that may define new transcriptional networks deployed during cardiac differentiation. Our data illustrate the strengths of a differentiation time course in identifying gene regulatory networks and reveal a chromatin-level determination of cell fates in the earliest stages of differentiation that may be key to heart development.

RESULTS

Expression and Chromatin States in Cardiac Differentiation

We investigated how global patterns of gene expression and chromatin organization are coordinated in the cardiac lineage. We used directed differentiation of mouse embryonic stem cells (ESCs) to cardiomyocytes as a model system (Figure S1 available online). This approach reproduces normal cardiomyocyte differentiation (Kattman et al., 2011) and resulted in roughly 70% cardiac Troponin T (cTnT)-positive cardiomyocytes (Figure S1 and Movie S1). Differentiating cultures were highly enriched at earlier stages for the cardiac transcription factors Nkx2-5 andIsl1, indicating that these cells progress efficiently through normal cardiac differentiation (Figure S1). Based on marker gene analysis, we selected four stages of differentiation that represent key cell types in the transition from pluripotent cells to cardiomyocytes (Figure 1A): undifferentiated embryonic stem cells (ESC) expressing pluripotency genes (Pou5f1/Oct4 and Nanog), cells expressing mesodermal markers (Mesp1 and Brachyury) (MES), cells expressing cardiac transcription factors (Nkx2-5, Tbx5, andIsl1) but not yet beating (CP), and functional CM with cardiomyocyte-specific gene expression (Myh6 and Myh7).

We analyzed global expression patterns of polyadenylated transcripts and microRNAs (miRNAs) in the four cell types by RNA sequencing (RNA-Seq) and Nanostring, respectively. We identified over 13,500 genes expressed during the time course (reads per kilobase per million [RPKM] > 1). Genes were clustered by expression pattern using HOPACH (Pollard et al., 2005), yielding distinct clusters (Figure 1B, Figure S2, Table S1), including groups of genes specifically expressed at each stage (e.g., Clusters A, L, N, and S). The stage-specific clusters were enriched for expected Gene Ontology (GO) terms (Table S2). Using the Nanostring platform, we analyzed over 600 miRNAs and found that they also display dynamic and stage-specific expression (Figure 1C, Table S1). Our data confirmed expression of key ESC miRNAs, such as the miR290 cluster, as well as known cardiac miRNAs (including miR-1, miR-208, and miR-143) in the CM stage. These data also identify several other stage-specific miRNAs that represent potential new regulators of cardiac differentiation.

Long ncRNAs (lncRNAs) are noncoding polyadenylated transcripts with emerging roles in gene regulation (Pauli et al., 2011). LncRNAs are differentially expressed in mammalian cell types, suggesting roles in lineage commitment (Cabili et al., 2011). However, lncRNAs have not yet been implicated in heart development. Notably, we find that lncRNAs show striking stage-specific expression in our differentiation system (Figure 1D, Table S1). LncRNAs regulate gene expression in cis and trans and may also function as transcriptional enhancers (Ørom et al., 2010). We expected that if lncRNAs function in cis to regulate lineage commitment, then their neighboring genes should have functions related to this process. To test this idea, we determined GO enrichment for the two nearest genes relative to lncRNAs expressed >1 RPKM in at least one stage (Figures 1D and 1E). Consistent with our hypothesis, we found enrichment of genes involved in development, morphogenesis, and transcriptional processes (Figure 1F).

We find that lncRNAs identified in our data are significantly correlated in expression with their neighboring genes compared to randomly selected neighboring protein coding genes (hyper-geometric test \( p = 4 \times 10^{-32} \)). We tested the possibility that the observed correlations are attributable to coordinately regulated gene clusters, however, we find that lncRNA expression is more highly correlated with the nearest adjacent gene (\( p = 0.0275 \)) relative to our background model (Figure 1F, Extended Experimental Procedures). Our data suggest that some lncRNAs may regulate gene expression in cis during cardiomyocyte differentiation. Moreover, we find many correlated lncRNA-gene pairs are associated with known cardiac genes such as Gata6, Hand2 and Myocd (Figure 1G, Table S1). These data identify several potential noncoding regulators and will facilitate further study of lncRNAs in cardiogenesis.

Chromatin State Dynamics during Cardiac Differentiation

Chromatin structure is key to transcriptional regulation, yet its role during differentiation is largely unknown. To this end, we performed chromatin immunoprecipitation and massively parallel sequencing (ChIP-Seq) for several histone modifications at the time points examined for gene expression (ESC, MES, CP, and
Modifications included H3K27me3 and H3K4me3 (associated with inactive and active promoters respectively) and H3K4me1 and H3K27ac (associated with promoters and enhancers) (Barski et al., 2007; Cui et al., 2009; Ernst et al., 2011; 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 2 kb around the TSS of each gene (Figure 2A, Table S1, and Extended Experimental Procedures). Consistent with previous studies showing little variation of chromatin patterns at promoters across cell types (Heintzman et al., 2009), the largest cluster, Cluster 1, had high levels of H3K4me3, H3K27ac, and RNAP (all associated with active chromatin and transcription) across the time course. GO analysis showed that these genes are involved in fundamental cellular functions, such as metabolism and cell-cycle regulation (Table S3). Other clusters, however, revealed dynamic chromatin patterns, suggesting chromatin regulation is critical to cardiac gene expression and differentiation.

Dynamic Chromatin States Correlate with Distinct Expression Patterns

We considered that genes with similar temporal expression patterns would share a 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 and expression cluster and determined statistical enrichment (Figure 2B). We found that the mesoderm-specific expression cluster L is primarily associated with chromatin cluster 9. However, the ESC-specific expression cluster A, comprising genes rapidly silenced upon differentiation, correlated with several chromatin patterns, including chromatin cluster 5, where active marks are lost without gain of additional marks tested, and cluster 11, where active marks are gradually lost while gaining the repressive H3K27me3 modification (Figure 2A). Regulators of stem cell state fell into more than one cocluster (A11 Pou5f1/Oct4, and A5 Nanog), suggesting expression of pluripotency regulators is controlled by multiple epigenetic mechanisms.

We next examined how chromatin and expression patterns 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, genes in cocluster A11 include active genes with highly correlated chromatin states at the ESC stage (Figure 3). Upon differentiation, these genes were downregulated at the MES stage. However, this initial change in expression did not correlate with changes in chromatin until later in differentiation at the CP stage. Conversely, genes in the mesoderm-specific expression Cluster L9 (Figure 3B), which are expressed at the MES stage and repressed at the CP stage, correspondingly have highly correlated active MES and silent CP chromatin states.

Our analysis also revealed that chromatin patterns could distinguish functionally distinct genes with a similar expression pattern (Figure 3C). For example, expression cluster S comprises genes expressed at the CM stage that are associated with diverse chromatin patterns, such as H3K4me3 and no H3K27me3 (S3), H3K4me3 and H3K27me3 (S8 and S10), and gain of low levels of H3K4me3 during differentiation without H3K27me3 enrichment (S20, S23, S24, S26, S27, and S28). However, 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). This indicates that despite similar expression patterns, groups of functionally related genes can be distinguished at the chromatin level.

A Novel Chromatin State Transition during CM Differentiation

Our promoter clustering revealed a group of genes that showed enrichment for H3K4me1 enrichment prior to enrichment of H3K4me3 and RNA Pol II and transcriptional activation. As H3K4me1 has largely been associated with open chromatin at distal enhancers prior to activation (Creyghton et al., 2010; Ernst et al., 2011; Heintzman et al., 2009; Rada-Iglesias et al., 2011; Zentner et al., 2011), we hypothesized that H3K4me1 may mark a similar promoter state. Although we found H3K4me3 and H3K4me1 were often enriched at the same TSSs, a fraction (15%–20%) of genes marked by H3K4me1 was not H3K4me3 enriched (Figure 4A) and was poorly expressed (Figure 4B). These included many contractile protein genes, such as Actc1, for which H3K4me1 was present at the MES stage prior to transcriptional activation at the CP and CM stages (Figure 4C). Notably, H3K27me3 was not enriched at these promoters, suggesting that this group of genes was not repressed by Polycomb during cardiomyocyte differentiation.

To gain broader insights, we classified genes based on the pattern of H3K4me1 and H3K4me3 at their TSS. We identified three gene groups that showed interesting patterns of these modifications (Figures 4D and 4E). Group 1 gained H3K4me1 prior to H3K4me3 enrichment and transcriptional activation and was enriched exclusively for cardiovascular genes including those encoding contractile proteins associated with terminal differentiation and cardiomyocyte function. H3K4me1 was often maintained at these TSSs upon H3K4me3 enrichment and gene activation. Group II gained H3K4me1 over time but failed to gain H3K4me3 or robust expression and included muscle lineage genes such as Ckm, Ckmt2, and Tcap, whose expression is associated with cardiomyocyte maturation. Group III genes transiently gained H3K4me1 at specific stages, but showed no H3K4me3 enrichment during differentiation. These genes were not expressed above background levels throughout differentiation and function in noncardiac lineages. Group II and III genes did not acquire H3K27me3 suggesting Polycomb-independent silencing. Although the functional role of the H3K4me1 preactivation pattern is unknown, chromatin remodelling at early stages of cardiac differentiation may be necessary for gene activation during terminal differentiation.

Enhancer Activity Correlates with Cardiac Specific Programs

Although regulation at promoters is important for gene regulation, distal enhancers are key regulators of 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). Using these modifications, we identified 81,497 putative distal enhancer regions during cardiac differentiation (Figures 5A and 5B, Figures 5Aa–5Ac, Table S4, and Extended Experimental Procedures). H3K4me1 marks most elements at each stage, whereas H3K27ac is enriched at a subset of these regions (Figure 5A). 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 our data to H3K4me1 and H3K27ac profiles from neural precursors, liver, and pro B cells (Creyghton et al., 2010) revealed that our enhancers are largely unique to the cardiac lineage (Figure S3D).

Our predicted set of enhancers significantly overlaps the smaller sets identified by p300/CRBP binding in fetal mouse heart, fetal or 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), which supports our enhancer predictions (Figures S3D and S3E). Consistent with evidence indicating that cardiac enhancers are conserved over a limited phylogenetic distance (Blow et al., 2010; May et al., 2012), we find our predicted enhancers have low overlap with highly conserved (>600) Phastcons elements, which increases...
significantly when considering only highly conserved elements in placental mammals (Figures S3F and S3G). Thus, our analysis identified many putative heart enhancers that likely function in heart development, including novel enhancers that may regulate the transition from pluripotency to a functionally differentiated state.

Histone modification patterns can distinguish active enhancers (which correlate with tissue-specific expression) from poised enhancers (which correlate with potential gene expression later in development) (Creighton et al., 2010; Ernst et al., 2011; Rada-Iglesias et al., 2011; Zentner et al., 2011). We classified our enhancers as active (H3K27ac+, H3K4me1+) or poised (H3K4me1+ only) at each stage of differentiation (Figure 5C). We find most enhancers at a given stage are poised, 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 5C), consistent with...
Figure 4. 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 preactivated gene, Actc1. ChIP-Seq (H3K4me1, H3K27ac, H3K4me3, y axis reads/million unique mapped reads), and RNA-Seq (FPKM) 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 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.
transcription initiation at these regulatory elements (Kim et al., 2010). Conversely, although H3K27me3 has been shown to demarcate poised enhancers elements (Rada-Iglesias et al., 2011; Zentner et al., 2011), this mark had a minimal overlap with our enhancer regions (Figure S4). Thus, while poised and active enhancers can be broadly defined by a limited set of histone modifications, there likely exist many other substates that comprise functionally distinct enhancer states.

To test whether active enhancers correlate with stage specific gene expression, we assigned active enhancers to their single nearest gene (Table S4) and found that genes associated with active enhancers are expressed at significantly higher levels than genes associated with unmarked or poised enhancers at each stage (Figure 5D). GO analysis of genes associated with active enhancers revealed enriched categories that progressively become cardiomyocyte-specific (Figure 5E, Figure S5A, Table S5). For example, enhancer-associated genes at the MES stage function in mesoderm and embryonic pattern specification, whereas at the CP stage they function in heart morphogenesis and cardiac tissue development. In CMs, we observe

**Figure 5. Identification of Enhancer Elements during Cardiac Differentiation**

(A) Total distal enhancers identified in ESC, MES, CP, and CM categorized by H3K27ac and H3K4me1 status at each stage. (B) Distribution of enhancers across the genome. (C) Density of ChIP-Seq reads ±4 kb relative to the midpoint of enriched regions for H3K4me1, H3K27ac, and RNA Polymerase (serine 5 phosphorylated form). (D) Boxplots of log2 transformed (FPKM) 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. Boxplots show interquartile ranges (IQR) with whiskers extending to the furthest data point that was no further than 1.5 times the IQR from the interquartile boundaries. (E) –log(Binomial FDR Q value) scores for GO Biological Process enriched in single nearest gene associated with active enhancers.
Figures A and B show chromatin state analysis for Myh7 and Nkx2-5 genes. Figure C illustrates the differentiation time course and transition state changes in ESC, CM, MES, and CP cells, with percentages indicating the proportion of cells in each state.
a transition to 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.

Enhancer Transitions during CM Differentiation

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 6A). The set of active and poised enhancers is largely unique at each stage, indicating that enhancers are highly cell type-specific even between closely related cell types. A subset of enhancers showed poised-to-active state transitions concomitant with activation of the proximal gene (e.g., Myh7 and Nkx2-5) (Figure 6B, Figure S6). However, most poised enhancers failed to acquire an active state during cardiomyocyte differentiation. These initially poised enhancers may be required to specify cell fates in other lineages during early development, suggesting that cells retain significant plasticity during lineage commitment.

The dynamic cell-type specificity of enhancer usage suggested that transitions between poised and active enhancer states occur rapidly between stages. Consistent with this, comparing enhancer states between each stage of cardiac differentiation showed that the fraction of active enhancers that transition through a poised state is largest during the MES to CP and CP to CM transition and lowest between unrelated cell types (Figure 6C and Figure S5). Although enhancer transitions between ESCs and the differentiated cell types remain below ~5%, they comprise 22% to 56% during the MES-CP transition, respectively (Figure S5). The cell type specificity and rapid state transitions of enhancers suggests dynamic enhancer usage is an important regulatory mechanism for coordinating tissue-specific gene expression.

Integrating Enhancers into Gene Networks

Although enhancers regulate global developmental gene expression patterns, integrating these genomic elements into the core transcriptional regulatory circuitry is challenging. Transcription factors (TFs) can act as master regulators of gene expression programs by binding to specific motifs within cis-regulatory elements. Given the stage-specific expression of TFs in our time course (Figure 7A, Figure S2), 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 and used these regions to search for motifs. We found over-represented motifs at each stage, including those for TFs that regulate the ESC state (OCT4_01, LRG1_Q5, and 500seq_marson) and cardiac development (GATA_Q6, MEF_Q6_01, MEIS1BHOXA9_02, and SRF) (Figure 7B). We compiled a list of TFs known to bind these highly conserved motifs and found strong correlations between TF expression and motif enrichment at each stage (Figure 7C). To address whether these predictions represented binding events, we analyzed ChIP-seq data 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 motifs at enhancers in ESCs (p = 2 × 10^-74 and 3 × 10^-26) (Figure S5D). GATA4 bound regions strongly correlate with enhancers that have a GATA_Q6 motif (p = 3 × 10^-74 and 2 × 10^-30) respectively at the CP and CM stage.

To construct gene regulatory networks connected to specific TF-motif pairs, we selected target genes positively correlated with TF expression. We find that many of these pairs were associated with genes that function in common pathways. For example, at the CP stage, MEIS, GATA, and NFATC enhancer-associated genes comprise networks implicated in cardiovascular development and function (Figure 7D, Figure S7). Although Meis1 and Meis2 have been implicated in heart development (Crowley et al., 2010; Pfeuffer et al., 2010; Stankunas et al., 2008), their targets in cardiac differentiation are unknown. Our data suggest MEIS1/2 regulate a subset of genes important in cardiac morphogenesis. Moreover, GATA factors potentially bind enhancers that may regulate important cardiac genes, including Nkx2-5, Mef2c, and Gata4 itself (Bruneau, 2008; Srivastava, 2006).

We further tested our enhancer predictions by analyzing the effects of loss of function of particular TFs on gene expression. Genes in the predicted OCT4-regulated network were highly correlated with genes affected by Oct4 knockdown (p = 3 × 10^-44) (Loh et al., 2006). Despite considerable redundancy among GATA factors (Zhou et al., 2012), we found the GATA4 network also predicted many genes differentially regulated in Gata4 knockout HL-1 cells (He et al., 2011) (p = 9 × 10^-12 at CP stage, p = 9 × 10^-15 at CM stage). These data indicate that we have identified a set of TFs that may regulate specialized gene expression networks by binding to distinct sets of enhancers during cardiomyocyte differentiation.

Combinatorial interactions among transcription factors can increase the diversity of regulatory modules governed by a particular factor. We observed a significant overlap among target genes associated with enhancers containing MEIS1BHOXA9_02 and GATA_Q6 motifs at the CP stage (Figure 7E, Table S6). Upon further analysis, we identified groups of developmentally important genes regulated by only GATA, only MEIS, or MEIS and GATA together suggesting independent as well as coregulation.
by MEIS/HOX and GATA factors. Enhancers coenriched for both MEIS and GATA motifs are associated with genes important for cardiac development, such as Gata5, Irx4, Myocd, Zfpm2, Wnt2, and Smarcd3, and with genes that influence conduction system function (Hcn4), consistent with an association of MEIS1 with conduction parameters (Pfeufer et al., 2010). Notably, we find MEIS/GATA motifs are often enriched in the same enhancer (Figure 7F), suggesting a functional relationship between these factors.

We used luciferase reporter activation assays to test for coregulation by MEIS and GATA factors. We tested five enhancers with motifs for MEIS and GATA, including a Myocd enhancer active in the developing heart (Creemers et al., 2006). Cotransfection of the Myocd reporter with combinations of expression constructs for GATA4, MEIS1A, or MEIS2D, showed that this enhancer responded to GATA4 and was synergistically activated by the combination of GATA4 and MEIS1A (Figure 7G). This appeared to be specific to MEIS1A, because cotransfection of MEIS2D with GATA4 did not lead to synergistic activation. Furthermore, most (four of five) enhancers tested were synergistically activated by the combination of MEIS1A and GATA4 (Figure S7C). Thus GATA4 and MEIS1A can 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.

**DISCUSSION**

We have defined chromatin state transitions during cardiac differentiation that provide 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.

**Dynamic Epigenetic Transitions in Differentiation**

The rapid loss of expression of pluripotency-associated genes upon differentiation can be achieved by at least nine 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/Pou5f1). 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, whereas those involved in contractile function and sarcomere structure have a distinct pattern, although they share a similar expression profile. This implies that functionally related and coexpressed 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 importantly for adaptation to stress.

**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 preactivation pattern is consistent with the idea that molecular events in early lineage commitment mark genes for subsequent activation. A similar pattern has been observed in T cell differentiation, in which H3K4me2 precedes transcriptional activation for a select group of genes (Zhang et al., 2012). This preactivation pattern is likely important for genes that are not regulated by Polycomb complexes. The presence of H3K4me1 in anticipation of transcriptional activation is reminiscent of its presence at poised enhancers, for which only a minority show H3K27me3 enrichment. It is possible that H3K4me1 may mark regulatory elements such as proximal enhancers that may function to poise the TSS for activation. These data suggest that diverse mechanisms can poise specific classes of TSSs and enhancers for subsequent activation. We propose that early deposition of H3K4me1 at specific cardiac genes is a regulated step that facilitates later activation of these genes. It will be important to identify the chromatin regulator that is necessary to catalyze this modification as well as the factors required for its recruitment and whether the early deposition of this mark is required for proper activation of these genes.
Identification of Transcriptional Networks Based on Enhancer Predictions

Chromatin marks at genomic regions distal to the TSS provides a means to identify candidate enhancer elements (Creighton 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. For example, we identified networks for the GATA family of transcription factors that include GATA4, GATA5, and GATA6 in CP and CM stages. The GATA factors are known to orchestrate many developmental processes. In the developing heart, these transcription factors have broad roles in early differentiation, morphogenesis, and postnatal physiology (Zhao et al., 2008; Zhou et al., 2012). Mutations in GATA4 lead to congenital heart defects (Garg et al., 2003), and GATA6 is essential for heart development (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. Thus, we have identified a novel network potentially under control of Meis factors. Finally, we observed a striking overlap between GATA- and the MEIS1/BHOX9-binding sites at enhancers, and show that these enhancers can respond to GATA4 and Meis1 in reporter assays. Thus we have uncovered a previously unknown functional relationship between GATA and Meis TFs in heart development.

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 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; Qian et al., 2012; Song et al., 2012; Takeuchi and Bruneau, 2009) and to elucidate how disruption of these diverse regulatory modules contributes to congenital heart disease.

EXPERIMENTAL PROCEDURES

Detailed experimental and analysis methods can be found in the Extended Experimental Procedures.

Cardiomyocyte Differentiation and Analysis

E14 Tg(Krox2-5-EmGFP) mouse ES cells (Hsiao et al., 2008) were cultured in feeder-free conditions using standard techniques. Directed differentiations and analyses were performed essentially as described in (Kattman et al., 2011).

RNA-Seq and Analysis Pipeline

Total RNA was isolated from 5 × 10⁶ cells using TRIzol Reagent. Sequencing libraries were prepared according to Illumina RNA-Seq library kit with minor modifications. Paired-end RNA-Seq 36 base pair reads were aligned to mm9 (Mus musculus assembly July 2007). DESeq (Anders and Huber, 2010) was used to normalize raw read counts and estimate differential gene expression. USeq 7.0 (Nix et al., 2008) was then 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. Expression was normalized to the interquartile range across the time course; interquartile numbers were used for clustering using a cosine angle distance metric and the Hapoch cluster package (http://www.bioconductor.org/packages/2.1/bioc/html/hopach.html).

ChIP-Seq and Analysis Pipeline

Genome-wide localization of histone modifications (H3K4me3, H3K4me1, H3K27me3, and H3K27ac) and the serine 5 phosphorylated form of RNA Pol II for each stage was determined via chromatin immunoprecipitation followed by high throughput sequencing. Chromatin immunoprecipitations were performed according to the Young lab protocol (Lee et al., 2006) with minor modifications. Details on analysis pipeline can be found in Extended Experimental Procedures.

Transient Transactivation Assays for Enhancer Validation

Candidate enhancers were cloned into a modified pGL3 luciferase reporter construct. Expression constructs for Gata4, Meis1a, and Meis2d (250 ng each) were cotransfected with each reporter construct (600 ng each) in 10T1/2 cells. Luciferase activity was assessed 40 hr later and normalized to Renilla luciferase activity. Conditions were analyzed for statistical differences using a one-way repeated-measures ANOVA followed by Tukey’s Multiple Comparison Test.

ACCESSION NUMBERS

All sequencing reported in this article have been deposited in the GNomEx database under accession numbers 44R and 7R2 and can be found at https://b2b.hci.utah.edu/gnomex/ by using the guest login.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, six tables, and one movie and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2012.07.035.

ACKNOWLEDGMENTS

We thank T. Sukonnik for cloning enhancer fragments; S. Thomas (Gladstone Bioinformatics Core) for help with Figure 3; J. Wythe and P. Devine for the modified pGL3 constructs; M. Cleary, E.N. Olson, and D. Wotton for constructs; and Bruneau, Boyer, and Pollard lab members for helpful discussions. This work was supported by the NHLBI Bench to Bassinet Program (U01HL08179, to B.G.B., L.A.B., K.S.P., and D.S.), NIH F32-HL104913 (J.A.W.), the Lawrence J. and Florence A. DeGeorge Charitable Trust/American Heart Association Established Investigator Award (B.G.B.), by William H. Younger, Jr. (B.G.B.) and Massachusetts Life Sciences Center (L.A.B.), L.A.B. is a Pew Scholar in the Biomedical Sciences. D.S. is a scientific founder of iPierian, and member of the Scientific Advisory Boards of iPierian and RegeneRx Biopharmaceuticals.

Received: February 11, 2012
Revised: May 21, 2012
Accepted: July 3, 2012
Published online: September 13, 2012
REFERENCES


