

Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart

Aibin He^{a,b,1}, Sek Won Kong^{a,b,c,1}, Qing Ma^{a,b}, and William T. Pu^{a,b,2}

^aDepartment of Cardiology and ^cChildren's Hospital Informatics Program, Children's Hospital Boston, Boston, MA 02115; and ^bHarvard Stem Cell Institute, Harvard University, Cambridge, MA 02138

Edited by Eric N. Olson, University of Texas Southwestern, Dallas, TX, and approved February 23, 2011 (received for review November 12, 2010)

Identification of genomic regions that control tissue-specific gene expression is currently problematic. ChIP and high-throughput sequencing (ChIP-seq) of enhancer-associated proteins such as p300 identifies some but not all enhancers active in a tissue. Here we show that co-occupancy of a chromatin region by multiple transcription factors (TFs) identifies a distinct set of enhancers. GATA-binding protein 4 (GATA4), NK2 transcription factor-related, locus 5 (NKX2-5), T-box 5 (TBX5), serum response factor (SRF), and myocyte-enhancer factor 2A (MEF2A), here referred to as "cardiac TFs," have been hypothesized to collaborate to direct cardiac gene expression. Using a modified ChIP-seq procedure, we defined chromatin occupancy by these TFs and p300 genome wide and provided unbiased support for this hypothesis. We used this principle to show that co-occupancy of a chromatin region by multiple TFs can be used to identify cardiac enhancers. Of 13 such regions tested in transient transgenic embryos, seven (54%) drove cardiac gene expression. Among these regions were three cardiac-specific enhancers of Gata4, Srf, and switch/sucrose nonfermentable-related, matrix-associated, actin-dependent regulator of chromatin, subfamily d, member 3 (Smardc3), an epigenetic regulator of cardiac gene expression. Multiple cardiac TFs and p300-bound regions were associated with cardiac-enriched genes and with functional annotations related to heart development. Importantly, the large majority (1,375/1,715) of loci bound by multiple cardiac TFs did not overlap loci bound by p300. Our data identify thousands of prospective cardiac regulatory sequences and indicate that multiple TF co-occupancy of a genomic region identifies developmentally relevant enhancers that are largely distinct from p300-associated enhancers.

gene regulation | motif enrichment analysis | in vivo biotinylation | TEA domain family member 1

A major challenge in deciphering the mammalian genome is the identification of regulatory elements that direct gene expression in the developing and adult organism. A subset of elements that activate transcription in a tissue (enhancers) can be identified by applying ChIP and high-throughput sequencing (ChIP-seq) to enhancer-associated features such as p300 (1). However, ChIP-seq of each enhancer-associated feature identifies only a fraction of the regulatory elements active in a tissue (2), and therefore delineation of the enhancers that are active in a tissue requires multiple complementary approaches.

Through sequence-specific DNA binding, transcription factors (TFs) play pivotal roles in translating the regulatory information encoded in genome sequences. In the heart, a number of transcription factors have been implicated in regulating gene expression (3, 4). Among these TFs are NK2 transcription factor related, locus 5 (NKX2-5), GATA-binding protein 4 (GATA4), T-box 5 (TBX5), and the MADS domain TF serum response factor (SRF) and myocyte-enhancer factor 2A (MEF2A). These factors, here defined as "cardiac TFs," are expressed in cardiomyocytes and are essential for initiating or maintaining cardiac gene expression (3, 5). These factors are expressed in multiple tissues, and specificity in promoting cardiac gene expression has been hypothesized to result from combinatorial interactions between these and other cardiac TFs (6). Although this principle has been established by

study of a handful of model genes (e.g., refs. 7–10), it has not been evaluated using unbiased, genome-wide approaches.

In this study, we used a modified ChIP-seq approach to define genome wide the binding sites of these cardiac TFs (1). We provide unbiased support for collaborative TF interactions in driving cardiac gene expression and use this principle to show that chromatin co-occupancy by multiple TFs identifies enhancers with cardiac activity in vivo. The majority of these multiple TF-binding loci (MTL) enhancers were distinct from p300-bound enhancers in location and functional properties.

Results

Genome-Wide TF Location Analysis. Currently available antisera for several cardiac TFs are not suitable for genome-wide ChIP-seq. To circumvent this limitation, we used a doxycycline-inducible dual adenovirus system to express biotinylated cardiac TFs (*SI Appendix, Fig. S1A*), permitting factor pulldown on streptavidin under stringent, uniform conditions and independent of antibodies (11). The use of this approach to tag TFs was validated previously for genome-wide ChIP analysis without reported effect on the factor's protein interactions or DNA binding properties (11–15). We expressed the biotinylated factors in the HL1 cardiomyocyte cell line, which expresses signature cardiac genes, forms organized sarcomeres, responds to β -adrenergic agonists, and exhibits spontaneous beating (16). Adenovirus and doxycycline doses were titrated so that GATA4^{flbio} and MEF2A^{flbio} were expressed at nearly endogenous levels (^{flbio} indicates FLAG and biotinylation epitope tags) (*SI Appendix, Fig. S1B*). Expression of SRF^{flbio}, TBX5^{flbio}, and NKX2-5^{flbio} under the same conditions resulted in moderate (two- to fivefold) overexpression (*SI Appendix, Fig. S1C*). The proteins were efficiently pulled down on streptavidin beads (*SI Appendix, Fig. S1D*), validating in vivo biotinylation of the tagged proteins.

We individually expressed each biotinylated TF in HL1 cells and performed biotin-mediated ChIP-seq (Table 1) (11). As controls, we sequenced mock samples that contained the biotinylating enzyme BirA but lacked epitope-tagged TF ("BirA" sample) as well as "input" chromatin. In addition, we used antibody-mediated ChIP to pull down chromatin associated with the transcriptional coactivator p300, a marker of active enhancers (1, 17). We used the peak-calling algorithm Sole-Search (18) to identify regions ("peaks") in which the experimental ChIP sample was enriched for tags compared with the input sample at a false discovery rate of 0.001. These parameters led to identification of thousands of peaks for each factor (Table 1 and *Dataset S1*). In contrast, the BirA control sample was enriched

Author contributions: A.H. and W.T.P. designed research; A.H. and Q.M. performed research; A.H. and S.W.K. contributed new reagents/analytic tools; A.H., S.W.K., and W.T.P. analyzed data; and A.H., S.W.K., and W.T.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Array and high-throughput sequencing data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE21529).

¹A.H. and S.W.K. contributed equally to this work.

²To whom correspondence should be addressed. E-mail: wpu@enders.tch.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1016959108/-DCSupplemental.

Table 1. Summary of ChIP-seq data

Sample	Reads ($\times 10^6$)	Peaks ($\times 10^3$)	Height	Length (bp)	BirA overlap	ChIP-qPCR (%)
GATA4 ^{flbio}	8.6	17.0	27	298	0	86
MEF2A ^{flbio}	17.1	1.3	25	232	0	80
NKX2-5 ^{flbio}	20.5	20.7	39	272	0	100
SRF ^{flbio}	9.9	24.1	26	276	0	75
TBX5 ^{flbio}	8.9	56.4	33	334	5	77
p300	12.8	1.5	32	259	0	100
BirA	4.4	0.1	16	207	NA	NA
Input	11.7	NA	NA	NA	NA	NA

BirA overlap, the number of peaks overlapping BirA peaks; ChIP-qPCR, the percentage of tested peaks with greater than twofold enrichment (*SI Appendix, Fig. S1E*); Height, number of reads mapping to the called peaks; Length, peak length; Peaks, the number of identified peaks; Reads, uniquely mapped reads.

for only 94 peaks (Table 1), and these peaks were subtracted from the final TF peak calls. The number of peaks identified by identical ChIP procedures varied considerably by TF, from 56,362 for TBX5 to 1,339 for MEF2A. This variation was not related to the number of uniquely mapped reads or to the degree of factor overexpression.

To validate the ChIP-seq peaks, we chose 15–20 peaks for each factor and measured their enrichment by chromatin pull down followed by quantitative PCR (ChIP-qPCR). Seventy-five percent to 100% of called peaks showed greater than twofold enrichment by ChIP-qPCR (Table 1, *Dataset S2*, and *SI Appendix, Fig. S1E*), indicating that the peak calls were reliable. We also sought to evaluate the extent to which TF binding detected in HL1 cells reflected TF binding in intact hearts. We were able to assess this correspondence for GATA4 by antibody-mediated ChIP from mouse heart. Nine of the 11 genomic regions occupied by GATA4 in HL1 cells also were occupied by GATA4 in adult mouse heart, suggesting that many peaks identified in HL1 cells also are present in adult hearts (*SI Appendix, Fig. S1F*).

To validate the ChIP-seq peaks further, we performed de novo motif discovery to test the expectation that the peaks should be enriched for the motif corresponding to the precipitated TF. Consensus and optimal in vitro DNA binding motifs were obtained from JASPAR, UniPROBE, and literature searches (19–22). Two independent de novo motif discovery approaches, Weeder (23) and MEME (24), were used to find motifs enriched in the top 500 ChIP peaks for each factor and yielded similar results. Motifs recovered from GATA4 and SRF peaks matched those reported previously (Fig. 1), indicating concordance between the in vivo binding site inferred from ChIP, the in vivo consensus sequence, and the optimal in vitro binding site. NKX2-5

and MEF2A peaks yielded motifs similar to previously reported motifs but differing at a single informative position. In the case of NKX2-5, we verified that this difference was not caused by the epitope tag (*SI Appendix, Fig. S2A*). TBX5 peaks were enriched for CG and GC dinucleotides not observed for the other four TFs (*SI Appendix, Fig. S2B and C*). After exclusion of these CG-/GC-rich peaks, we recovered motifs similar to the reported consensus but containing two additional flanking, highly informative bases (Fig. 1). These data further validate the ChIP-seq peaks and suggest that the in vivo binding motifs for NKX2-5, MEF2A, and TBX5 vary from the optimal in vitro motifs (*Discussion*).

Functional Role of TF Chromatin Occupancy. We characterized the location of cardiac TF and p300 peaks. These peaks had the highest density in the vicinity of the transcription start site (TSS), although the majority of peaks were intergenic (>10 kb from a RefSeq transcription unit) and intronic (*SI Appendix, Fig. S3A and B*), as is characteristic of enhancers. Intronic peaks tended to occur in the first intron (*SI Appendix, Fig. S3C*).

To begin to assess the functional significance of identified factor binding sites, we studied evolutionary conservation of DNA regions bound by cardiac TFs in the ChIP-seq data. Regions identified by ChIP-seq were significantly conserved (*SI Appendix, Fig. S3D*), and the degree of conservation was related to the peak rank, suggesting an overall relationship between strength of binding and function.

Next, we asked if our dataset contained enhancers that previously were reported to be regulated by cardiac TFs. Of 22 essential TF-enhancer interactions supported by transgenic analysis identified in a literature review, eight were recovered in our ChIP-seq experiments (*Dataset S3*), suggesting that many TF binding sites that we identified are functional in vivo. In addition, many reported enhancers bound additional TFs that were not noted previously. For instance, the NKX2-5-regulated promoter of *Ankrd1* (25) also was occupied by the other four cardiac TFs. The SRF-regulated enhancer of the cardiac microRNAs miR1-1/miR133a-2 (26) also was bound by TBX5. On the other hand, 14 previously reported cardiac TF-enhancer interactions were not recovered in our screen. It is likely that some TF-enhancer interactions that occur in fetal heart and in specific subregions of the myocardium are not represented in the HL1 cell line.

To define further the functional links between gene expression and chromatin occupancy by cardiac TFs and p300, we used shRNA to knock down GATA4 by more than 90% in HL1 cells (*SI Appendix, Fig. S4*). Microarray expression profiling revealed 251 genes that were differentially expressed downstream of GATA4 (nominal P value < 0.005 and absolute fold-change > 1.5; $n = 3$; *Dataset S4*). Of these genes, 149 (59%) were bound by GATA4. GATA4-bound genes were 3.6 times more likely to be differentially expressed than non-GATA4-bound genes (χ^2 test $P < 0.0001$). These 149 genes included genes important for

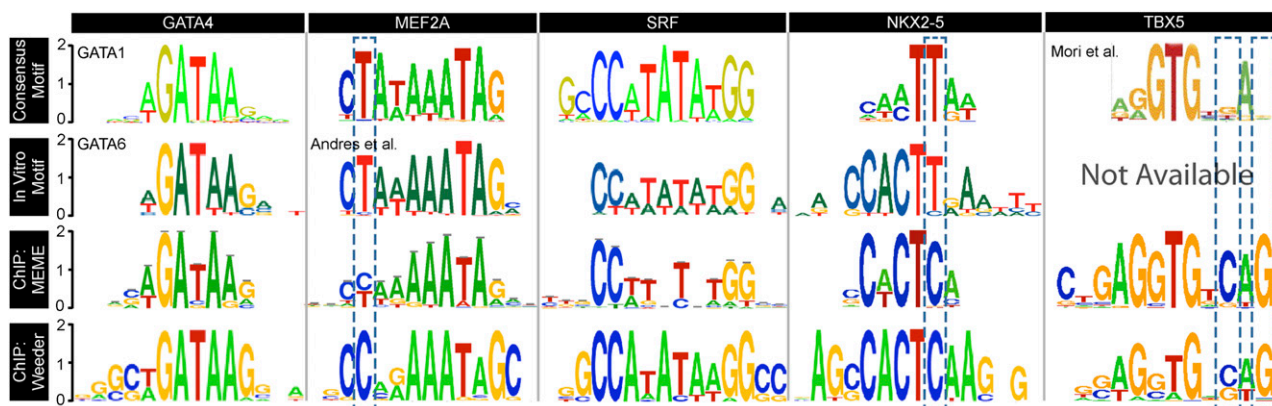


Fig. 1. In vivo TF binding motifs. De novo motif discovery of in vivo motifs by MEME and Weeder using the top 500 peaks of ChIP-seq data. All MEME E-values were less than 10^{-60} . For TBX5, high GC-content peaks were excluded. Motifs found by de novo discovery were compared with available consensus and optimal in vitro motifs from JASPAR, UniPROBE, or the indicated reference. Dashed boxes highlight differences between in vivo and in vitro motifs.

heart function, such as myosin-binding protein C, cardiac (*Mybpc3*), natriuretic peptide type A (*Nppa*) and type B (*Nppb*), phospholamban (*Pln*), and ryanodine receptor 2 (*Ryr2*) (Dataset S4). Indeed, Gene Ontology (GO) term analysis showed significant enrichment for terms related to heart development and function (Dataset S4). These data indicate that we have identified a number of GATA4-regulated enhancers important for expression of key cardiac genes.

Enriched Motifs in Cardiac TF-Bound Regions Identify TF Interactions. We hypothesized that the cardiac TFs and p300 interact with other TFs and that we could discover these interacting TFs by enrichment of their motifs in the ChIP-seq peak sequences. Therefore, we scanned the top 500 ChIP-seq regions for each factor with known TF binding motifs. We scored each motif for significantly increased binding frequency in ChIP-seq peaks compared with background (SI Appendix, Methods). This approach was validated by looking for enrichment of cardiac TF motifs among the top ChIP-seq peaks. Consistent with the results of the de novo motif discovery, ChIP-seq peaks for each TF were highly enriched for the cognate TF binding motif (SI Appendix, Fig. S5A and Dataset S5) and for motifs of factors previously reported to interact with the TF [e.g., GATA4-MEF2 and GATA4-NKX2-5 (7, 10, 27)].

When the analysis was extended to additional TF motifs, a number were strongly enriched (Fig. 2A; SI Appendix, Fig. S5B; Dataset S4). For instance, the TEA domain family member 1 (TEAD1) motif was highly enriched among sequences pulled down by p300, GATA4, NKX2-5, and MEF2A (Fig. 2A). TEAD1, also known as “transcriptional enhancer factor 1” (TEF-1), is a transcriptional regulator of muscle genes and is required for heart development (28). ChIP-qPCR of selected regions occupied by cardiac TFs and containing the TEAD1 motif confirmed in vivo TEAD1 occupancy in 11 of 15 cases (Fig. 2B), suggesting that a high percentage of TEAD1 motifs within cardiac TF ChIP peaks represent true binding sites. We next asked if TEAD1 interacts physically with p300 or core cardiac TF in HL1 cells. TEAD1 specifically coprecipitated with NKX2-5, SRF, GATA4, and p300 (SI Appendix, Fig. S5C and D), consistent with the greatest enrichment of TEAD1 motifs in the ChIP peaks of these four factors (Fig. 2B). Consistent with these data, TEAD1-SRF interaction was noted previously (29).

The motif of TF yin-yang 1 (YY1), recently implicated as a negative regulator of pathological cardiac hypertrophy (30), also was enriched in cardiac TF ChIP peaks. YY1 coprecipitated with NKX2-5 and SRF (SI Appendix, Fig. S5C and D). Thus, using a statistically based motif-scanning approach, we were able to extend the cardiac TF interaction network.

Binding of Multiple TF Marks Cardiac Enhancers. Cardiac TFs frequently occupied overlapping or nearly overlapping chromatin sites. TFs were defined as co-occupying a chromatin region when individual participant peaks were separated by 500 bp or less. Chromatin co-occupancy occurred with surprising frequency, with 21% of peaks co-occurring with the peak of at least one other cardiac TF (Fisher’s exact test; $P < 10^{-16}$) (Table 2). Four or five TFs co-occupied 1,715 regions, which were defined as MTL (Dataset S6). The co-occurrence of four or five factors at this frequency was very unlikely to occur by chance (permutation test; $P < 10^{-6}$). MTL-associated genes were highly enriched for genes with selective cardiac expression, as assessed by the cardiac enrichment score (defined as the ratio of expression in heart to the average expression in all other tissues using expression values from Affymetrix exon array profiling in GEO GSE15998; Fig. 3A). MTL-associated genes were more highly expressed in HL1 cells or heart than were randomly selected genes (one-sided Wilcoxon test; $P < 10^{-35}$), and the median expression of MTL-associated genes was higher in heart than in other tissues. All MTLs included GATA4, and MTL genes were 1.6-fold more likely to be differentially expressed in GATA4-depleted HL1 cells (χ^2 test; $P < 0.006$). These data provide unbiased support for the notion that collaborative interactions between cardiac TF direct cardiac gene expression.

Physical interaction between GATA4 and MEF2 synergistically regulated expression of the model cardiac gene *Nppa* (10),

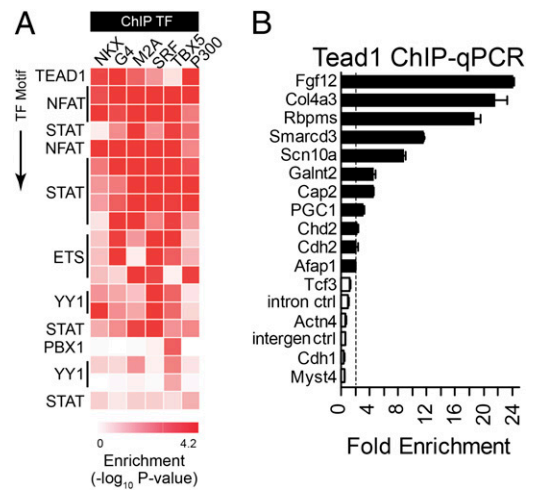


Fig. 2. Expansion of the cardiac TF interaction network by motif enrichment analysis. (A) Heat map showing statistical enrichment of selected JASPAR and TRANSFAC motifs among top 500 peaks bound by cardiac TF. A heat map of all analyzed motifs is shown in SI Appendix, Fig. S5. (B) TEAD1 ChIP-qPCR assay of cardiac TF peaks containing predicted TEAD1 motifs. Fold enrichment indicates TEAD1 compared with IgG1 ChIP and normalized to Actin- β (*Actb*) intronic control. Filled bars indicate greater than twofold enrichment (dotted line). *Actn4*, actinin α 4; *Afap1*, actin filament-associated protein 1; *Cap2*, adenylate cyclase-associated protein 2; *Cdh1*, cadherin 1 type 1; *Cdh2*, cadherin 2; *Chd2*, chromodomain helicase DNA binding protein 2; *Col4a3*, collagen type IV, α 3; *Fgf12*, fibroblast growth factor 12; *Galnt2*, UDP-N-acetyl- α -D-galactosamine:polypeptide N-acetylglucosaminyltransferase 2; *Myst4*, MYST histone acetyltransferase (monocytic leukemia) 4; *Rbpms*, RNA binding protein gene with multiple splicing; *Scn10a*, sodium channel, voltage-gated type X α subunit; *Tcf3*, transcription factor 3.

and GATA4-MEF2 co-occupied genes were 2.1-fold more likely to be down-regulated in GATA4-depleted HL1 cells than were other GATA4-bound genes (χ^2 test; $P = 0.0004$). To investigate the significance of this interaction further, we knocked down MEF2A and GATA4 in HL1 cells (SI Appendix, Fig. S4). Of 587

Table 2. Co-occurrence of core cardiac TF peaks

Transcription factors	Peaks	Genes	Examples
1 TF	120,278	14,132	<i>Hand1</i> , <i>Irx4</i> , <i>Tnni3k</i> ,
≥ 2 TF	18,343	9,824	<i>Tbx20</i> , <i>Mef2c</i> , <i>Hopx</i> , <i>Scn5a</i> , <i>Gja5</i>
≥ 3 TF	6,434	4,782	<i>Srf</i> , <i>Gata4</i> , <i>Gata6</i> , <i>Mef2d</i> , <i>Nppa</i> , <i>Nppb</i> , <i>Mybpc3</i>
≥ 4 TF	1,715	1,623	<i>Nkx2-5</i> , <i>Mef2a</i> , <i>Tbx5</i> , <i>Hand2</i> , <i>Pparg</i> , <i>Myocd</i>
5 TF	287	286	<i>Ankrd1</i> , <i>Atp2a2</i> , <i>Ryr2</i> , <i>Fhl2</i> , <i>Myh7</i>

Co-occurring peaks were defined as peaks separated by 500 bp or less. Peaks were mapped to the nearest gene within 100 kb. Some genes were associated with multiple peaks. *Ankrd1*, ankyrin repeat domain 1; *Atp2a2*, ATPase, Ca⁺⁺ transporting, cardiac muscle, slow twitch; *Fhl2*, four and a half LIM domains 2; *Gata6*, GATA-binding protein 6; *Gja5*, gap junction protein, α 5; *Hand1*, heart and neural crest derivatives expressed transcript 1; *Hand2*, heart and neural crest derivatives expressed transcript 2; *Hopx*, HOP homeobox; *Irx4*, Iroquois homeobox 4; *Mef2c*, myocyte enhancer factor 2C; *Mef2d*, myocyte enhancer factor 2D; *Myh7*, myosin, heavy chain 7, cardiac muscle, β ; *Myocd*, myocardin; *Pparg*, peroxisome proliferator-activated receptor γ ; *Ryr2*, cardiac muscle ryanodine receptor-calcium release; *Scn5a*, sodium channel voltage-gated, type V α subunit; *Tbx20*, T-box 20; *Tnni3k*, TNNI3 interacting kinase.

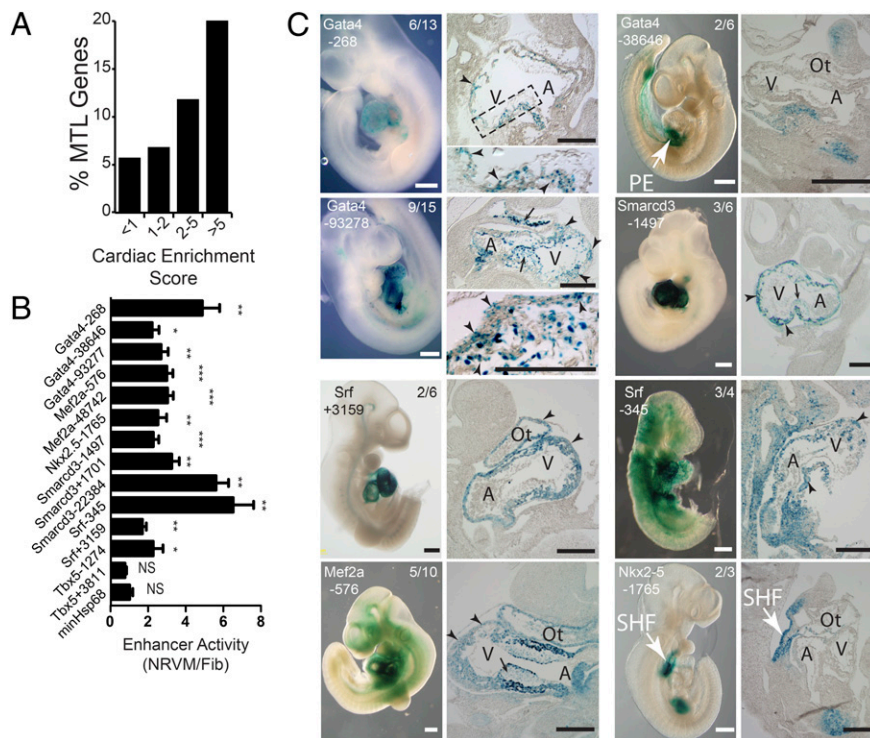


Fig. 3. Genomic regions co-occupied by multiple cardiac TFs direct cardiac gene expression. (A) Genes with a higher cardiac enrichment score (cardiac expression/average expression in other tissues) were associated more frequently with MTLs. (B) Enhancer activity in vitro. Enhancers cloned upstream of hsp68-lacZ were transfected into neonatal rat cardiomyocytes or cardiac fibroblasts along with pGL3-luc. Ratio of LacZ activity in neonatal rat ventricular myocytes (NRVM) to fibroblasts was plotted after normalization to luciferase activity. $n = 3$. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. NS, not significant. (C) Enhancer-hsp68-lacZ constructs were used to generate E10.5 transgenic embryos. Representative Xgal-stained embryos are shown. Numbers indicate embryos with cardiac expression over the total PCR⁺ embryos. Arrowheads indicate myocardial expression. Black arrows indicate activity in endocardium and endocardial cushions. PE, proepicardium; SHF, second heart field. (White scale bars: 400 μ m; black scale bars: 200 μ m.)

genes that were differentially expressed, 50 (8.5%) were associated with regulatory elements co-occupied by GATA4 and MEF2A (Dataset S4). GATA4, MEF2A, and GATA4-MEF2A chromatin co-occupancy increased the likelihood that genes were differentially regulated by GATA4 plus MEF2A knock down by 2.6, 3.3, and 4.6-fold, respectively (χ^2 test; $P < 0.0001$). These 50 genes were significantly enriched in GO terms related to heart development and function (Dataset S4). These data suggest that we have identified a number of enhancers directly and functionally regulated by GATA4 and MEF2A and that GATA4-MEF2A interaction is a significant contributor to transcriptional regulation by these factors.

Given that cardiac TFs collaboratively bind genomic regions and direct cardiac gene expression, we next asked if binding of multiple cardiac TFs to a genomic region predicts activity of the region as a cardiac enhancer. To answer this question, we tested genomic regions bound by multiple TF for enhancer activity in both in vitro and in vivo assays. To maximize the biological value of the tested genomic regions, we focused on enhancers flanking the cardiac TF themselves, plus switch/sucrose nonfermentable-related, matrix-associated, actin-dependent regulator of chromatin, subfamily d, member 3 (*Smarcd3*), a key epigenetic regulator of cardiac gene expression (31, 32). Although these genes are well known for their role in cardiac gene expression, each is expressed in multiple noncardiac tissues. We selected 13 enhancers flanking these genes that were bound by three or more cardiac TFs (SI Appendix, Fig. S6 A–D and Dataset S7) and positioned them upstream of a heat-shock protein 68- β -D-galactosidase (hsp68-lacZ) minimal promoter-reporter. In vitro, 12/13 enhancers showed greater activity in cardiomyocytes than in fibroblasts (Fig. 3B). In transient transgenic assays for enhancer activity in vivo, 7 of the 13 enhancers (54%) drove transgene expression in the embryonic day 10.5 (E10.5) heart or heart-forming regions (Fig. 3C). One of these, the *Nkx2-5* -1765 enhancer active in the second heart field had been reported previously (33). *Nkx2-5* -1765, *Gata4* -268, *Gata4* -93278, *Smarcd3* -1497, and *Srf* +3159 displayed predominantly cardiac activity, whereas *Srf* -345 and *Mef2a* -576 were active in multiple tissues including heart. Histological sections confirmed activity in cardiomyocytes (Fig. 3C, arrowheads). Some enhancers also were active in endocardium and endocardial cushion mesenchyme

(Fig. 3C, arrows). An eighth enhancer, *Gata4* -38646, was active in the septum transversum and proepicardium (Fig. 3C), in agreement with a previous report (34). The data reveal an extensive transcriptional network that governs cardiac gene expression (SI Appendix, Fig. S6E), in which each cardiac TF binds to its own regulatory regions as well as to the regulatory regions of the other cardiac TF. Moreover, these data demonstrate that binding of multiple cardiac TFs to a genomic region is predictive of regions with cardiac transcriptional activity.

We were interested in whether cardiac TFs bound at MTLs are each required for enhancer activity. ChIP-seq indicated that *Smarcd3* -1497 enhancer was bound by GATA4 as well as by TBX5, NKX2-5, and SRF (SI Appendix, Fig. S6D), and the GATA4 and TBX5 peaks contained consensus GATA4 and TBX5 binding sites (SI Appendix, Fig. S6D). ChIP-qPCR confirmed GATA4 occupancy of the *Smarcd3* enhancer in HL1 cells (Dataset S2), fetal heart, and adult heart (Fig. 4A). Single mutation of either site in the *Smarcd3* -1497 enhancer did not disrupt cardiac transcriptional activity (Fig. 4B). However, mutation of both sites in combination strongly decreased enhancer activity in the cardiac chambers and, to a lesser degree, in the outflow tract (Fig. 4B). Thus, at least for *Smarcd3* -1497, GATA4 and TBX5 binding is redundant for enhancer activity.

MultiTF and p300 Binding Mark Distinct Subsets of Cardiac Enhancers.

p300 ChIP-seq identifies a subset of enhancers active in a tissue (1, 2, 17, 35). We mapped p300-bound cardiac enhancers in HL1 cells by antibody-mediated ChIP-seq and identified 1,504 p300 peaks (Table 1 and Dataset S1). ChIP-qPCR validated 16 of 16 p300 peaks (SI Appendix, Fig. S1E), indicating that a high fraction of the called peaks were enriched for p300 chromatin occupancy. p300 was recruited to genomic regions by association with cardiac TFs, because 89.7% of p300-bound regions were co-occupied by at least one of the cardiac TF (Fig. 5A). GATA4, the most frequent co-occupying TF, was found at 76% of the p300-bound regions, consistent with a physical interaction between p300 and GATA4 (36).

Like MTL-associated genes (Fig. 5A), p300-associated genes were highly enriched for genes with robust and selective cardiac expression, as quantified by cardiac enrichment score (Fig. 5B). Also like MTL genes, p300 genes were more highly expressed in

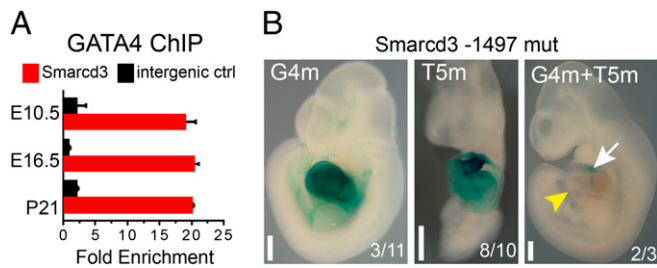


Fig. 4. GATA4 and TBX5 binding sites are required for *Smarcd3* -1497 activity. (A) Validation of GATA4 occupancy of *Smarcd3* -1497 by ChIP-qPCR from mouse heart at the indicated developmental stages. E, embryonic; P, postnatal. (B) Activity of *Smarcd3* -1497 enhancers containing mutation of GATA4 (G4m), TBX5 (T5m), or both motifs indicated in *SI Appendix, Fig. S6D*. Arrow indicates residual activity in outflow tract. Yellow arrowhead indicates loss of activity in cardiac chambers. Numbers indicate Xgal⁺ and PCR⁺ embryos. (Scale bars: 500 μ m.)

HL1 cells or heart than were randomly selected genes (one-sided Wilcoxon test; P value $< 10^{-11}$), and the median expression of MTL genes was higher in heart than other tissues. Thus, multiple TF binding and p300 binding are each means to identify enhancers of cardiac gene expression.

However, only a minority of enhancers were both MTL- and p300-bound (Fig. 5C). Each enhancer class stimulated gene expression to a similar degree (Fig. 5D). Compared with p300-bound enhancers, multiTF enhancers were more enriched between +5 kb and -5 kb of the TSS (Fig. 5E). GO term analysis indicated that both MTL and p300 genes were enriched for the biological process “heart development” (Fig. 5F). MTL genes also were enriched for terms related to transcription and chromatin, whereas p300 but not MTL genes were significantly enriched for “cell adhesion.” Interestingly, five of the seven enhancers with cardiac activity in vivo were not bound by p300 (Dataset S5), suggesting that these enhancers function independently of p300 recruitment. Collectively these data indicate that multiple cardiac TF binding and p300 binding identify distinct subsets of cardiac enhancers.

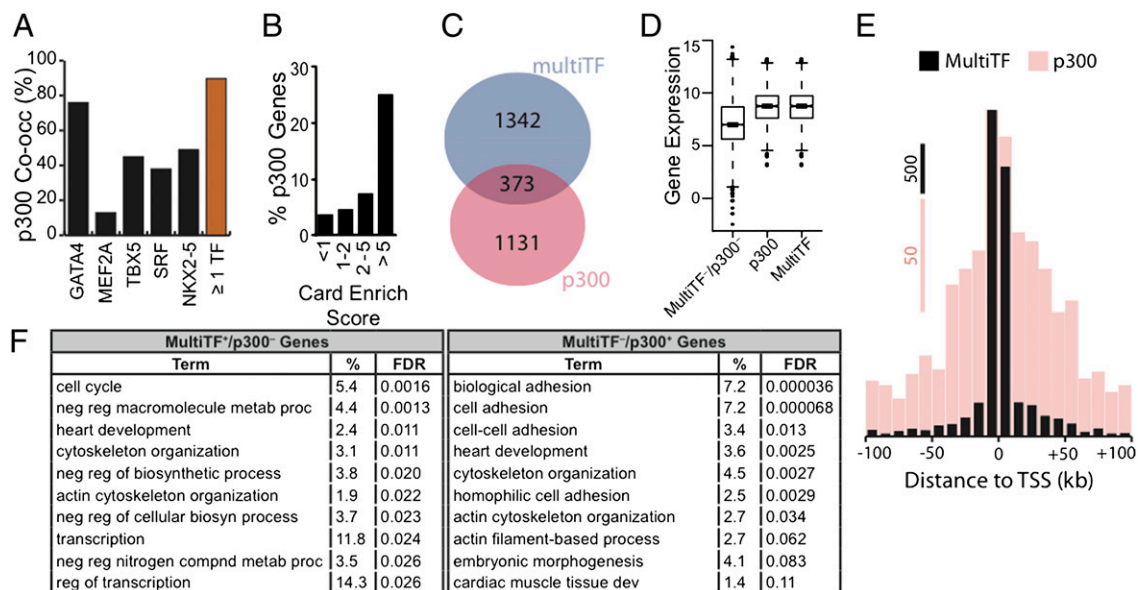


Fig. 5. MultiTF and p300 binding mark distinct sets of enhancers. (A) p300 frequently co-occupied genomic loci with cardiac TF, most notably GATA4. (B) Genes with higher cardiac enrichment scores were associated more frequently with p300. (C) The preponderance of multiTF and p300 enhancers did not overlap. (D) MultiTF and p300 genes were more highly expressed in HL1 cells than were genes that lacked these enhancers ($P < 10^{-16}$), but expression levels of multiTF and p300 genes were indistinguishable. Gene expression is indicated in log₂ scale. (E) MultiTF enhancers were located more proximal to the TSS than p300 enhancers. (F) Gene Ontology (GO) term analysis of MultiTF⁻/p300⁻ and MultiTF⁺/p300⁺ enhancers. Top 10 terms, fraction of positive genes within the set, and Benjamini–Hochberg false discovery rate (FDR) are shown for each class of enhancers.

Discussion

Using ChIP-seq for five cardiac TFs and p300 in the HL1 cardiomyocyte cell line, we identified thousands of genomic regions that potentially contribute to cardiac gene expression. Our HL1 screen using *in vivo* biotinylated TF allowed us to use uniform, stringent pulldown conditions to perform ChIP and circumvented limitations of currently available antisera. The *in vivo* biotinylation approach has been validated in prior studies (11–15), and current data suggest that it does not alter TF activity. Limitations of our HL1 screen are that only a subset of the binding sites identified in this system may correspond to binding sites present *in vivo* at different stages of development and disease and that some *in vivo* binding sites may not be captured in this *in vitro* system. Nevertheless, our extensive *in vivo* validation studies indicate that the HL1 dataset captures functional cardiac TF chromatin occupancy in cardiomyocytes. Therefore, the dataset will be an invaluable resource for unraveling transcriptional regulatory mechanisms in cardiomyocytes.

Analysis of the DNA sequences bound by each cardiac TF revealed their *in vivo* binding motifs. There was considerable overlap between optimal *in vitro* and *in vivo* motifs. However, important differences suggest that the *in vivo* binding site is influenced by factors in addition to the intrinsic interaction between each TF and naked DNA. These other factors might include histones, epigenetic marks, and protein–protein interactions with other transcriptional regulatory molecules. Additionally, some ChIP peaks did not contain a recognizable TF binding motif, probably reflecting mechanisms of TF recruitment other than direct TF binding of DNA, such as “piggyback” recruitment, DNA looping, or interactions with modified histones or other TFs (37).

Analysis of sequences bound by cardiac TFs and p300 identified overrepresented TF motifs, suggesting interaction between the precipitated TF and the TF with enriched motifs. Direct testing of a subset of these predictions revealed interactions between TEAD1 and GATA4, between NKX2-5, SRF, and p300, and between YY1, NKX2-5 and SRF, thereby extending the cardiac TF interaction network. Sequential application of this approach will delineate more completely the extent of TF–TF interactions that drive tissue-specific gene expression.

Our genome-wide location analysis of five cardiac TFs revealed extensive binding of these TFs to their own and to each other’s

flanking regulatory regions, often with multiple-factor co-occupancy and in the absence of p300. Our analysis of these putative enhancers confirmed that a subset activated cardiac transcription *in vivo*, including the cardiac-specific enhancers of *Gata4*, *Smarcd3*, and *Srf*. These data suggest that the cardiac TFs participate in a transcriptional regulatory network involving interactions of cardiac TFs with their own regulatory regions, reminiscent of transcriptional networks that maintain pluripotency in embryonic stem cells (11).

Our analysis of the *Smarcd3* -1497 enhancer identified a prospective feed-forward regulatory circuit. This enhancer was directly bound and activated by GATA4. Bruneau and colleagues showed that SMARCD3 and GATA4 interact physically and together are sufficient to initiate the cardiac gene program in non-cardiac mesoderm (31, 32). Previously, ablation of both GATA4 and GATA6 was shown to down-regulate *Smarcd3* (38), although it was not clear if this effect reflected direct or indirect regulation or simply loss of cardiac mesoderm. Our data indicate that GATA4 and other cardiac TFs directly promote cardiac *Smarcd3* expression *in vitro* and *in vivo* and suggest a feed-forward circuit in which GATA4 stimulates expression of SMARCD3, which then cooperates with GATA4 to activate expression of cardiac genes.

The frequent co-occurrence of cardiac TFs provides unbiased support for collaborative TF interactions promoting cardiac gene expression. Moreover, the majority (7 of 13) of tested genomic regions with co-occurrence of multiple cardiac TFs exhibited cardiac transcriptional activity. Although we selected genomic regions flanking cardiac TF that were bound by multiple cardiac TFs for *in vivo* analysis, we do not think this selection substantially biased our analysis, because cardiac TFs are expressed in many noncardiac tissues, but multiTFs co-occupancy successfully identified regulatory regions that drove cardiac transcription *in vivo*. Indeed, the proportion of these MTL regions with cardiac activity may be higher than we found, because we

assayed activity at only one developmental stage. These observations indicate that multiTF co-occupancy is an effective means to recover tissue-restricted enhancers.

The majority of enhancers identified by binding of multiple TF were not co-occupied by p300. Indeed, screening for p300 enhancers would have missed the cardiac-specific *Gata4* -93278, *Smarcd3* -1498, and *Srf* +3160 enhancers that we recovered. The distribution of multiTF enhancers tended to be more proximal to the TSS than the p300 enhancers, and functional annotations of associated genes were distinct between these enhancer classes. These data suggest that multiple TF binding identifies enhancers different from those bound by p300 and is a complementary approach for enhancer identification.

Materials and Methods

Details of materials and methods are provided in *SI Appendix*. Procedures involving animals were performed under protocols approved by the Institutional Animal Care and Use Committee. Epitope-tagged TFs were expressed in HL1 cells (16) by adenoviral gene transfer. CHIP was performed as described (11), and precipitated DNA was sequenced on an Illumina GA2. Primers used for CHIP-qPCR are summarized in *Dataset S8*. Array and high-throughput sequencing data were deposited in the Gene Expression Omnibus as GSE21529. Transient transgenic analysis was performed by pronuclear injection (Cyagen Inc). Embryos were collected at E10.5 and were stained with 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (Xgal). Results are displayed as mean \pm SEM.

ACKNOWLEDGMENTS. W.T.P. was supported by National Heart, Lung, and Blood Institute Grants HL095712 and HL098166 and by charitable contributions from Edward Marram and Karen Carpenter. S.W.K and A.H. were supported by grants from the Hood Foundation and the American Heart Association, respectively.

1. Visel A, et al. (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457:854–858.
2. Heintzman ND, et al. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 39:311–318.
3. Olson EN (2006) Gene regulatory networks in the evolution and development of the heart. *Science* 313:1922–1927.
4. Oka T, Xu J, Molkentin JD (2007) Re-employment of developmental transcription factors in adult heart disease. *Semin Cell Dev Biol* 18:117–131.
5. Naya FJ, et al. (2002) Mitochondrial deficiency and cardiac sudden death in mice lacking the MEF2A transcription factor. *Nat Med* 8:1303–1309.
6. Nemer G, Nemer M (2001) Regulation of heart development and function through combinatorial interactions of transcription factors. *Ann Med* 33:604–610.
7. Durocher D, Charron F, Warren R, Schwartz RJ, Nemer M (1997) The cardiac transcription factors Nkx2-5 and GATA-4 are mutual cofactors. *EMBO J* 16:5687–5696.
8. Sepulveda JL, et al. (1998) GATA-4 and Nkx-2.5 coactivate Nkx-2 DNA binding targets: Role for regulating early cardiac gene expression. *Mol Cell Biol* 18:3405–3415.
9. Garg V, et al. (2003) GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature* 424:443–447.
10. Morin S, Charron F, Robitaille L, Nemer M (2000) GATA-dependent recruitment of MEF2 proteins to target promoters. *EMBO J* 19:2046–2055.
11. Kim J, Chu J, Shen X, Wang J, Orkin SH (2008) An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* 132:1049–1061.
12. de Boer E, et al. (2003) Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc Natl Acad Sci USA* 100:7480–7485.
13. Rodriguez P, et al. (2006) Isolation of transcription factor complexes by *in vivo* biotinylation tagging and direct binding to streptavidin beads. *Methods Mol Biol* 338:305–323.
14. Yu C, et al. (2002) Targeted deletion of a high-affinity GATA-binding site in the GATA-1 promoter leads to selective loss of the eosinophil lineage *in vivo*. *J Exp Med* 195:1387–1395.
15. Lausen J, et al. (2010) Targets of the Tal1 transcription factor in erythrocytes: E2 ubiquitin conjugase regulation by Tal1. *J Biol Chem* 285:5338–5346.
16. Claycomb WC, et al. (1998) HL-1 cells: A cardiac muscle cell line that contracts and retains phenotypic characteristics of the adult cardiomyocyte. *Proc Natl Acad Sci USA* 95:2979–2984.
17. Blow MJ, et al. (2010) ChIP-Seq identification of weakly conserved heart enhancers. *Nat Genet* 42:806–810.
18. Blahnik KR, et al. (2010) Sole-Search: An integrated analysis program for peak detection and functional annotation using ChIP-seq data. *Nucleic Acids Res* 38:e13.
19. Portales-Casamar E, et al. (2010) JASPAR 2010: The greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res* 38(Database issue):D105–D110.
20. Newburger DE, Bulyk ML (2009) UniPROBE: An online database of protein binding microarray data on protein-DNA interactions. *Nucleic Acids Res* 37(Database issue):D77–D82.
21. Mori AD, et al. (2006) Tbx5-dependent rheostatic control of cardiac gene expression and morphogenesis. *Dev Biol* 297:566–586.
22. Andrés V, Cervera M, Mahdavi V (1995) Determination of the consensus binding site for MEF2 expressed in muscle and brain reveals tissue-specific sequence constraints. *J Biol Chem* 270:23246–23249.
23. Pavesi G, Mauri G, Pesole G (2001) An algorithm for finding signals of unknown length in DNA sequences. *Bioinformatics* 17 (Suppl 1):S207–S214.
24. Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: Discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res* 34(Web Server issue):W369–W373.
25. Kuo H, et al. (1999) Control of segmental expression of the cardiac-restricted ankyrin repeat protein gene by distinct regulatory pathways in murine cardiogenesis. *Development* 126:4223–4234.
26. Zhao Y, Samal E, Srivastava D (2005) Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature* 436:214–220.
27. Lee Y, et al. (1998) The cardiac tissue-restricted homeobox protein Csx/Nkx2.5 physically associates with the zinc finger protein GATA4 and cooperatively activates atrial natriuretic factor gene expression. *Mol Cell Biol* 18:3120–3129.
28. Yoshida T (2008) MCAT elements and the TEF-1 family of transcription factors in muscle development and disease. *Arterioscler Thromb Vasc Biol* 28:8–17.
29. Gupta M, et al. (2001) Physical interaction between the MADS box of serum response factor and the TEA/ATTS DNA-binding domain of transcription enhancer factor-1. *J Biol Chem* 276:10413–10422.
30. Sucharov CC, Dockstader K, McKinsey TA (2008) YY1 protects cardiac myocytes from pathologic hypertrophy by interacting with HDACS. *Mol Biol Cell* 19:4141–4153.
31. Lickert H, et al. (2004) Baf60c is essential for function of BAF chromatin remodeling complexes in heart development. *Nature* 432:107–112.
32. Takeuchi JK, Bruneau BG (2009) Directed transdifferentiation of mouse mesoderm to heart tissue by defined factors. *Nature* 459:708–711.
33. Searcy RD, Vincent EB, Liberatore CM, Yutzey KE (1998) A GATA-dependent nkx-2.5 regulatory element activates early cardiac gene expression in transgenic mice. *Development* 125:4461–4470.
34. Rojas A, et al. (2005) Gata4 expression in lateral mesoderm is downstream of BMP4 and is activated directly by Forkhead and GATA transcription factors through a distal enhancer element. *Development* 132:3405–3417.
35. Xi H, et al. (2007) Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS Genet* 3:e136.
36. Dai YS, Markham BE (2001) p300 Functions as a coactivator of transcription factor GATA-4. *J Biol Chem* 276:37178–37185.
37. Farnham PJ (2009) Insights from genomic profiling of transcription factors. *Nat Rev Genet* 10:605–616.
38. Zhao R, et al. (2008) Loss of both GATA4 and GATA6 blocks cardiac myocyte differentiation and results in acardia in mice. *Dev Biol* 317:614–619.