Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis

Graphical Abstract

Highlights
- Systems-level approach reveals GATA4 roles in human cardiac development and function
- Heterozygous GATA4 missense mutation impairs cardiac gene program
- GATA4 G296S mutation disrupts TBX5 genome occupancy at cardiac super-enhancers
- PI3K signaling is a key “hub” in the GATA4 gene regulatory network

Authors
Yen-Sin Ang, Renee N. Rivas, Alexandre J.S. Ribeiro, ..., Beth L. Pruitt, Michael P. Snyder, Deepak Srivastava

Correspondence
dsrivastava@gladstone.ucsf.edu

In Brief
A human missense mutation that causes congenital heart defects disrupts the cooperation between transcription factors at cardiac super-enhancers and gives rise to aberrant gene expression.
Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis

Yen-Sin Ang,1,2,3 Renee N. Rivas,1,2,3 Alexandre J.S. Ribeiro,4 Rohith Srivas,5 Janell Rivera,1 Nicole R. Stone,1,2,3 Karishma Pratt,1 Tamer M.A. Mohamed,1,2,3 Ji-Dong Fu,1 C. Ian Spencer,1 Nathaniel D. Tippens,6 Molong Li,1 Anil Narasimha,5 Ethan Radzinsky,1 Anita J. Moon-Grady,2 Haiyuan Yu,6 Beth L. Pruitt,4 Michael P. Snyder,5 and Deepak Srivastava1,2,3,7,*

1Gladstone Institute of Cardiovascular Disease and Roddenberry Center for Stem Cell Biology and Medicine, San Francisco, CA 94158, USA
2Department of Pediatrics
3Department of Biochemistry and Biophysics
University of California, San Francisco, San Francisco, CA 94143, USA
4Department of Mechanical Engineering
5Department of Genetics and Center for Genomics and Personalized Medicine
Stanford University, Stanford, CA 94305, USA
6Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14850, USA
7Lead Contact
*Correspondence: dsrivastava@gladstone.ucsf.edu
http://dx.doi.org/10.1016/j.cell.2016.11.033

SUMMARY

Mutation of highly conserved residues in transcription factors may affect protein-protein or protein-DNA interactions, leading to gene network dysregulation and human disease. Human mutations in GATA4, a cardiogenic transcription factor, cause cardiac septal defects and cardiomyopathy. Here, iPS-derived cardiomyocytes from subjects with a heterozygous GATA4-G296S missense mutation showed impaired contractility, calcium handling, and metabolic activity. In human cardiomyocytes, GATA4 broadly co-occupied cardiac enhancers with TBX5, another transcription factor that causes septal defects when mutated. The GATA4-G296S mutation disrupted TBX5 recruitment, particularly to cardiac super-enhancers, concomitant with dysregulation of genes related to the phenotypic abnormalities, including cardiac septation. Conversely, the GATA4-G296S mutation led to failure of GATA4 and TBX5-mediated repression at non-cardiac genes and enhanced open chromatin states at endothelial/endocardial promoters. These results reveal how disease-causing missense mutations can disrupt transcriptional cooperativity, leading to aberrant chromatin states and cellular dysfunction, including those related to morphogenetic defects.

INTRODUCTION

Combinatorial interactions between transcription factors (TFs) result in tissue-specific gene expression that dictates cell identity and maintains homeostasis. TFs activate or repress gene transcription by recruiting other TFs, co-activators, or co-repressors. Super-enhancers (SEs), clusters of putative enhancers densely occupied by Mediator complex and TFs, are implicated as regulators of cell identity in development and disease (Heinz et al., 2015; Whyte et al., 2013). SEs differ from typical enhancers (TEs) in size, motif density, and transcriptional activation, rendering them more sensitive to changes in molarity of TF complexes. Dysregulation at SEs may contribute to human developmental disorders in embryogenesis and postnatal disease.

Developmental malformations occur in > 5% of human births. Congenital heart defects (CHD) are most common (~0.8% live births) and are often due to haploinsufficiency of developmentally regulated cardiac TFs (Srivastava, 2006). Heterozygous mutations in TFs GATA4 and TBX5 cause familial CHD with overlapping phenotypes, and we showed that they co-immunoprecipitate when overexpressed. They are mutated in sporadic CHD and are associated with cardiomyopathies (Rajagopal et al., 2007; Zhao et al., 2014; Porto et al., 2010). We reported a heterozygous disease-causing GATA4 glycine-to-serine missense mutation (G296S) that impaired in vitro interaction of GATA4 and TBX5 (Garg et al., 2003). Mice with compound heterozygous Gata4 and Tbx5 mutations develop atrioventricular septal defects (AVSD), providing genetic evidence for their interaction (Maitra et al., 2009).

Gata4—a TF with WGTAR-recognizing zinc fingers—is expressed in developing myocardial, endocardial, and endodermal cells (Heikinheimo et al., 1994). Gata4 deletion causes extraembryonic and foregut endodermal malformations (Kuo et al., 1997; Molkentin et al., 1997), and it is essential in regulating cardiomyocyte (CM) proliferation and septal development (Misra et al., 2012; Rojas et al., 2008). Deleting Gata4 in CMs causes cardiac decompensation and Gata4−/− mice have cardiac hypoplasia and reduced hypertrophic response to pressure overload (Bisping et al., 2006; Oka et al., 2006).
Thus, Gata4 is essential in a dose-sensitive fashion for heart development and homeostasis.

Although Gata4 and Tbx5 are critical for mouse cardiogenesis, the gene targets or signaling pathways they co-regulate in human CMs and how they regulate human septal formation are unclear (Stefanovic et al., 2014; Xie et al., 2012). Complete loss of Tbx5 or Nkx2.5, Gata4-interacting partners, showed that these TFs interdependently modulate each other's genomic occupancy in mouse cardiac differentiation (Luna-Zurita et al., 2016). Yet, it is unknown if this depends on protein-protein interactions and if dose-dependent perturbations in co-occupancy underlie heart disease.

We used patient-derived induced pluripotent stem (iPS) cells to dissect GATA4 regulatory mechanisms in human cardiac development and function. We found that the heterozygous GATA4 G296S mutation impaired expression of the cardiac gene program and sonic hedgehog (SHH) signaling while upregulating genes of alternative fates, particularly the endothelial lineage and those related to cardiac septation. GATA4-dependent recruitment of TBX5 was disrupted at SE elements associated with genes for heart development and muscle contraction, and chromatin closure failed at loci involved in endothelial differentiation. This work reveals how a single missense mutation in a key cardiac TF leads to disease by dose-dependently regulating occupancy of TF complexes to enhancers and reveals potential nodes for therapeutic intervention.

RESULTS

Generation of Patient-Specific iPS Cells and Functional CMs

We reported a heterozygous c.886G>A mutation in human GATA4 linked to 100% penetrant atrial or ventricular septal defects (ASD; VSD), AVSD, or pulmonary valve stenosis (PS) (Figures 1A and S1A) (Garg et al., 2003). Mutant-GATA4 translated into a G296S missense substitution flanking the second zinc-finger domain, involved in DNA-binding and protein-protein interactions (Figure 1A, bottom). Our previous study found abnormalities in cardiac morphogenesis, but we now found GATA4 G296S patients with delayed-onset cardiomyopathy. This was characterized by decreased left ventricular systolic function and an unusual echocardiographic appearance of the right ventricle with deep trabeculations and thickening of papillary muscles in the left ventricle (Figure 1B and Movies S1 and S2). Deep trabeculation is typical of non-compaction, thought to reflect failure of ventricular CMs to mature.

We reprogrammed dermal fibroblasts from four subjects with the GATA4 G296S mutation and four family members without it into patient-specific iPS cells using non-integrating episomal vectors to generate purified CMs from the iPS cell lines (Figure S2A) (Lian et al., 2012; Tohyama et al., 2013). RNA-seq at various times showed stage-specific gene signatures for mesoderm, cardiac progenitor cells (CPCs), and CMs with expected gene ontologies (GO) (Figure S2B–S2D). iPS-CMs spontaneously contracted, expressed sarcomeric markers, and had membrane electrophysiology and gene expression similar to human CMs; 30% were binucleated (Figure S2E–S2H). Calcium flux showed proper drug responses. Electron microscopy indicated abundant mitochondria with defined Z-lines and sarcomeres (Figures 1E and 1F).

Impaired Contractility, Calcium Handling, and Metabolic Activity in Mutant CMs

We generated > 90% pure cTnT^+ day 32 (D32)-CMs from WT, G296S, and CRISPR-corrected isogenic iPS cells (Figure 2A), although mutant lines showed slight delays in onset of spontaneous contraction (Figures S2I and S2J). We built a micropatterning platform to measure contraction of single iPS-derived CMs (Figure 2B) (Ribeiro et al., 2015). Only 50% of patterned G296S CMs responded accurately to electrical pacing at 1Hz, compared to 70% of WT CMs. While WT cells did not respond to pacing at frequencies over 1Hz, 20% of G296S CMs beat at a faster rate (Figure 2C). G296S CMs had reduced contractile force generation per cell movement with decreased contraction time (Figures 2C and S2K), consistent with the cardiomyopathic phenotype in patients. Upon further differentiation at D70, G296S CMs were dysfunctional in response to electrical pacing and relaxation velocity, but force generation improved (Figure S2L–S2N).

In patch-clamp studies, G296S CMs had increased overshoot potential without altered maximum upstroke velocity or action potential duration (Figures 2D and S2O), suggesting a more depolarized membrane. Calcium transients in cell clusters had increased relative peak amplitude, suggesting defects in calcium ion handling (Figure 2E). When CMs were patterned onto 1-mm lines to induce uniaxial cell-cell communication, calcium flux in G296S CMs was higher (Figure 2F). A larger percentage of G296S CMs had disorganized sarcomeres (Figure 2G and S2P).

We hypothesized that the reduced contractile force came from defects in mitochondrial function or metabolic activity. Indeed, G296S CMs had decreased mitochondrial staining (Figure 2H), glycolytic capacity, and glycolytic reserve (Figure 2I). Although mitochondrial DNA (mtDNA) heteroplasmy is linked to neuropa-thogenicity, sequencing showed no increased de novo mutations of G296S mtDNA (Figure S2Q).

Attenuated Cardiac Gene Program in Mutant CPCs and CMs

We performed RNA-seq on isogenic iPS cells during differentiation into CPCs on day 7, contracting CMs before (D15-CMs) and after (D32-CMs) lactate purification (Figure 3A and Table S1). LASSO-regression algorithm predicted the iWT CM data represented the heart transcriptome (0.6–1). Transcriptomes of GATA4 G296S cells from each stage had lower cardiac scores (<0.6) (Figure 3B). In G296S cells, 2,228 genes were differentially expressed in at least one of the three stages with dynamic changes going from CPCs to mature CMs (Figure S3A and S3B). At all stages, 38 genes in Wnt-planar cell polarity pathway...
or vasculature-, endocardial-, heart-development, or cardiac progenitor differentiation were dysregulated (Figures S3C–S3D).

In G296S CPCs, Gene Set Enrichment Analyses (GSEA) showed decreased expression of genes typically present in cells receiving the SHH signal, including the PTCH1/PTCH2 receptors and GLI2/GLI3 transcriptional effectors (Figure 3C). In development, SHH secreted by pulmonary endoderm is received by neighboring atrial myocardium, resulting in growth of the atrial septum (Hoffmann et al., 2009); disrupting this in mice yields ASDs and AVSDs. Thus, downregulating genes required for SHH response is consistent with septal defects of GATA4-G296S patients. More broadly, downregulated genes were involved in heart development, cardiac chamber morphogenesis, myofibril assembly, heart contraction, and cardiac progenitor differentiation, suggesting incomplete activation of the myocardial gene program (Figures 3D and 3D). Uregulated genes (e.g., TAL1, ETS1, ROBO4, SOX17, TIE1, KDR, and KLF5) were involved in vasculature development, angiogenesis, extracellular matrix organization, integrin interactions, and calcineurin-NFAT transcription. Many are signaling or transcriptional regulators of the endocardial/endothelial program.

The percentage of G296S CPCs expressing high levels of GATA4, NKX2.5, and TBX5 were reduced, validating the gene expression decrease. Reciprocally, abundance of the endothelial-specific protein, KDR, was increased in GATA4/NKX2.5/ISL1-positive CPCs (Figures S3E and S3F). The percentage of
Figure 2. GATA4 G296S CMs Have Impaired Cardiac Function
(A) FACS analysis of cTnT+ CMs from representative WT and G296S differentiation after lactate purification.
(B) CMs micropatterned in arrays of single cells (top) and immunostained for α-Actinin or F-actin (bottom).
(C) Contractile measurements on micro-patterns. Percentage of single-CM responding to 1 Hz pacing in WT and G296S (left). Traction-force microscopy measurements of force production as a function of cell movement of CMs responding to 1Hz pacing (right). All measurements were done in triplicate with CMs generated independently from two patient lines. Data for patient 4 are shown.
(D) Action potential measurements of WT and G296S CMs. Overshoot potential (OSP) indicates highest membrane potential reached. Data shown are mean ± SEM from two WT and two G296S lines. *p < 0.05 (Mann-Whitney U test).
(E) Calcium flux measurement on microclusters. F/F₀ (Max), peak amplitude relative to baseline fluorescence between action potentials. Data shown are mean ± SEM from two WT and two G296S lines. *p < 0.05 (t test).
(F) Calcium flux measurement on patterned microtissues. CMs patterned on hydrogels of 10kPa-stiffness; 1-mm-long lines (left) and calcium flux measured as F/F₀ (center). Rates of rise and fall (right) between action potentials. Data are mean ± SEM. *p < 0.05 (t test).
(G) Percentage of CMs of individual sarcomeric classes observed by α-Actinin staining. Class IV represents the most disarrayed sarcomeric organizations. n > 150 CMs.
(H) Mitochondria staining intensity of single-CM micropatterns (top). Mitotracker red intensity relative to cell area was quantified (bottom). Data shown are mean ± SEM from 100 G296S lines. **p < 0.005 (t test).
(I) Seahorse measurements of glycolytic functions. Isogenic CM data shown are mean ± SEM. **p < 0.005, ***p < 0.0005 (t test).
See also Figure S2.
CD31-positive endothelial cells did not increase in unpurified D15 cultures (Figure S3G). Thus, the upregulated endothelial/endocardial program was likely due to failure in gene silencing, rather than more cells adopting an endothelial fate. When iPS cells were differentiated to promote the endothelial lineage (Theodoris et al., 2013), G296S cells were only marginally increased in propensity to commit into endothelial cells (Figure S3H).

In G296S D15-CMs, downregulated genes were critical in organ morphogenesis, heart development, and glycolysis (Figures 3E and 3F), consistent with the phenotypic abnormalities and impairment in glycolysis (Figure 2F). Like the CPC stage, upregulated genes at D15 participated in blood vessel development, cell-cell communication, and integrin and PI3K-Akt signaling pathways. Persistent expression of cardiac progenitor genes, such as ISL1, and upregulation of smooth muscle genes suggested alternative fate genes failed to be silenced as CMs matured. Increased expression of CAMK2D and CASQ2 was consistent with the increased calcium transients (Figures 2E and 2F).

In D32-CMs purified by lactate/glucose media, differentially expressed genes continued to show an attenuated cardiac gene program and persistent upregulation of the endothelial/endocardial gene program (Figures 3F and 3F'). GO terms for heart development, muscle contraction, cardiomyopathy, and cardiac septum development were enriched in downregulated genes. Upregulated genes were enriched for vasculature development, angiogenesis, and PI3K-Akt signaling. Genes in vascular and neuronal pathfinding were most upregulated in the neurogenesis category. Also, G296S CMs downregulated chamber myocardium genes and upregulated atrioventricular canal myocardium and smooth muscle-associated genes, suggesting a broader mis-specification in cell identity (Figure S3). Upregulating TBX2 was notable, given that it represses "working" myocardium genes and regulates atrioventricular canal development (Aanhaanen et al., 2011). Cellular respiration genes were reduced in G296S CMs (Figure S3J), consistent with decreased metabolic activity observed (Figures 2H and 2I). Quantitative PCR validation of the RNA-seq results showed a strong correlation for all three stages (Figure S3K).

**Open Chromatin Anomalies in GATA4 Mutant CPCs**

Chromatin accessibility is linked to TF occupancy and transcriptional output (Zaret and Carroll, 2011). To examine changes in open chromatin status (Figure 4A), we analyzed transposable-accessible chromatin by deep sequencing (ATAC-seq) in iWT and G296S CPCs (Table S2) (Buenrostro et al., 2013). In iWT CPCs, 14,532 ATAC-seq loci had 88% overlap with ENCODE DNase-hypersensitivity sites (DHSs) from human-CMs or ES-derived CPCs and at loci expected to be transposable-accessible (Figures 4B and 4C). Furthermore, > 75% had histone marks of activation (H3K4me3), but not repression (H3K27me3), and local-ized to introns (43%) of protein-coding genes (82%) (Figures 4C and 4D). In G296S CPCs, open-chromatin status was broadly reduced at cardiac genes (Figures 4B and 4E), consistent with their decreased expression (Figure 3D). Open chromatin status was increased at SOX17, a key regulator of hemogenic-endothelium (Clarke et al., 2013). This trend was also seen at 86 cardiac and 99 endothelial genes that were differentially expressed (Figure 4F).

Genomic loci with increased ATAC-seq signal were enriched for DNA motifs of core transcriptional regulators of endothelial cells (SOX17, KLF5, FOXO1, STAT6) and ETS-factors (GABPA, ELF5, ERG), suggesting that the endocardial/endothelial program was not effectively silenced in G296S CPCs (Figure 4G). These loci mapped to genes involved in AV valve morphogenesis, coronary vasculogenesis, and endocardial cushion development (Figure 4H), consistent with the AVSD diagnosis in the individual with the GATA4 mutation (Figure 1A). Hey1 and Hey2 are GATA4-interacting co-repressor proteins (Kathiriya et al., 2004) that were downregulated in G296S CMs and may contribute to failure of chromatin closure at endothelial/endocardial genes, while genes dependent on NFATc, an endothelial regulator, were upregulated (Figure 4I).

**Genome-wide Co-occupancy of GATA4 and TBX5 in Human CMs**

Open chromatin anomalies in mutant CPCs and GATA4's known function as a "pioneer factor" (Cirillo et al., 2002) led us to survey the genome-wide occupancy of GATA4 and TBX5 and histone marks of active-promoters (H3K4me3), repressed-promoters (H3K27me3), transcription elongation (H3K36me3), and active-enhancers (H3K27ac) (Tables S3 and S4). In WT CMs, chromatin immunoprecipitation with antibodies to the endogenous protein and deep sequencing (ChIP-seq) validated many direct targets of GATA4 or TBX5 identified in mouse studies (Figures 5A and S4A). These gene targets were co-bound by GATA4 and TBX5 (G4T5) and had high levels of H3K27ac, H3K4me3, and H3K36me3, but undetectable H3K27me3. GATA4 and TBX5 ChIP-seq signals positively correlated with gene expression (Figure S4B). GATA4, TBX5, and H3K27ac shared the strongest overlap in genome occupancy (Figure 5B), with nearly half of GATA4 sites co-bound by TBX5 (Figures S4C and S4D). The
Figure 4. Chromatin Accessibility Aberrations in G296S CPCs
(A) GSEA analyses of genesets for cardiac (top) and endothelial/endocardial (bottom) development. NES, normalized enrichment score. FDR, false discovery rate. Positive and negative NES indicate higher and lower expression in iWT, respectively.
(B) IGV browser tracks at chr14:23693015-24168059 show normalized ATAC-seq signal from WT (black) and G296S (red) matches normalized signal from ENCODE-DHS (blue) (gray regions).
(C) Heatmap of normalized read counts from ENCODE DHSs, H3K4me3, and H3K27me3 (D5CPC) around ATAC-seq loci identified in iWT CPCs. White and blue are low and high signal intensity, respectively.
(D) Pie-chart shows gene-body, upstream and downstream distribution (top), and coding and non-coding gene distribution (bottom) of 14532 iWT ATAC-seq loci.

(legend continued on next page)
2,428 sites co-bound by human G4T5 had higher ChIP-seq signals than sites bound by GATA4 or TBX5 alone (Figure 5C). Co-bound sites mapped to intronic (48%) and intergenic (35%) sites of genes for myofibril assembly, cardiac muscle development and contraction, CHD, and cardiomyopathy (Figures S4E and S4F). GATA4 and TBX5 motifs ranked at the top in motif analyses of G4T5 sites (Figure 5D). Motifs for TEAD4, MEF2C, NKX2.5, ISL1, SRF, and SMAD2/3 were enriched at these loci, indicating a TF code that maintains the cardiac gene program. Enrichment of motifs for endothelial regulators, FOXO1 and HOXB4, indicate a potential repressive role for G4T5 at these sites.

We systematically compared sites bound by GATA4, TBX5, or G4T5 in G296S and WT cells (Figure 5E; top). For GATA4 sites, 54% of sites were lost (L), 46% were unchanged (U), and 16% were ectopic sites gained (E) in mutants, suggesting dose sensitivity for DNA-binding at many sites and redistribution to others. For TBX5 sites, 26% were lost (L), 74% were unchanged (U), and 24% were ectopically gained (E). G296S had 34% fewer G4T5 co-bound sites than WT CMs (Figures S4D, S5A, and S5B), with 48% lost (L), 52% unchanged (U), and 21% gained (E). Next, we parsed the L, U, and E sites for the relative occupancy of GATA4, TBX5, and H3K27ac (Figure 5E; bottom). Consistent with the reduced DNA binding affinity of G296S GATA4, GATA4 occupancy was decreased particularly at G4L and G4T5L sites and correlated with increased TBX5 occupancy particularly at T5E and G4T5E. A broad increase in TBX5 occupancy suggested that loss of TBX5 occupancy at other sites was unlikely due to decreased TBX5 gene expression. The active enhancer mark H3K27ac was increased most at G4SE, T5E, and G4T5E sites. Nearly all of the changes were significant (Figure S5C), and GATA4, TBX5, and H3K27ac were not mis-localized at random genomic sites. From our RNA-seq data, genes mapping to G4T5 sites were largely downregulated in CPCs (Figure S5D), and TEAD4, MEF2C, Nkx2.5, ISL1, SRF, and SMAD2/3 were enriched at these loci, indicating a TF code that maintains the cardiac gene program. Enrichment of motifs for endothelial regulators, FOXO1 and HOXB4, indicate a potential repressive role for G4T5 at these sites.

To gain insights into a motif grammar that may explain why some loci were more sensitive to loss of G4T5 co-binding in the presence of the GATA4 G296S mutation, we compared the distance between GATA4 and TBX5 motifs within sites that lost G4T5 co-binding and sites unchanged in co-binding. The distance in GATA4 and TBX5 motifs was greater in G4T5 than G4T5 sites, regardless of strandedness (Figure 5G). G4T5 sites had fewer GATA4-TBX5 motif pairs than G4T5 sites, and motif pairs within G4T5 sites were preferentially located on the same strand compared to those in G4T5 sites (Figure 5H). Thus, protein-DNA interactions may compensate for disrupted protein-protein interactions. Also, motif analyses identified PRDM1, NR5A2, IRF1, PBX1, and HNF4A motifs in G4T5 sites, and TEAD4, EGR1, HIF1A, and MEIS1/3p-TBX5 motifs in G4T5 sites (Figure 5E). Cross-referencing the G4T5 sites to binding sites of > 200 ENCODE transcriptional regulators revealed closest proximity to p300-, CTCF-bound neuronal enhancers (Figure S5F). These results indicate that G4T5 cooperation is most robust when underlying cis-sequences are closely linked on the same DNA strand, ~75 bp apart, but may be most sensitive to perturbation at enhancers active in non-cardiac cells, perhaps due to weaker DNA interactions that require tethering of the TFs.

Consistent with an impaired cardiac gene program (Figures 2, 3, and 4), G4T5 sites were involved in cardiac muscle contraction, cardiac septal defect, and cardiomyopathy (Figure 5I). To determine putative GATA4 and TBX5 targets, we examined all differentially expressed genes with a G4T5 site within 20 kb (Figure S5G). Genes with decreased GATA4 and TBX5 binding (G4DOWN_T5DOWN) were downregulated (Figure S5H), suggesting that differential gene expression is directly due to DNA binding aberrations by GATA4 and TBX5. GATA4 binding was decreased, and TBX5 binding concomitantly increased at 414 putative targets (Figure 5J). Importantly, GATA4 binding was reduced at 49% of sites near 82 upregulated endothelial genes (Figure 5K). Consistent with TBX5-motif enrichment in G296S CPCs (Figure 4G), TBX5 binding was increased at 64% of 207 TBX5 sites within these endothelial topologically associating domains, suggesting anomalous transcriptional activation by mis-localized TBX5 and perhaps other coactivators. This correlated with increased H3K4me3 and decreased H3K27me3 marks at endothelial TSS in G296S CMs (Figure S5J). Proximal promoters of upregulated endothelial genes were enriched in binding sites for GATA-, FOXO-, and ETS-family proteins (Figure S5J).

GATA4 binding sites within endothelial TADs and P3K genes mapped closely with binding sites of multiple co-repressors (Figure S5K) whose proteins were expressed at detectable levels (Figure S5L). Since the GATA4/HDAC complex mediates gene repression in mouse AV canal (Stefanovic et al., 2014), we performed HDAC2 ChIP-seq in D15-CMs. GATA4 and HDAC2 binding sites overlapped and sites that were TBX5-HDAC2 co-bound were enriched in genes for cardiovascular development, muscle cell differentiation, insulin, and integrin signaling (Figures S5M and S5N). At endothelial TADs in G296S CMs, ~30% of sites had less HDAC2 binding than WT-CMs, suggesting endothelial gene upregulation was partially attributed to decreased HDAC2-repression (Figure S5O).

GATA4 and TBX5 Co-regulate Human Cardiac SEs

Regions of high MED1 (Mediator Complex) occupancy across several kilobases mark SEs (Whyte et al., 2013), but MED1-classified SEs have not been described in human CMs. Here, we identified 213 SEs (top 4%) by MED1 ChIP-seq in WT CMs.
Figure 5. TF Mis-localizations in G296S CMs
(A) IGV browser tracks of indicated ChIP-seq signals at known GATA4 target loci (NPPA, NPPB) in WT CM. Gray boxes, significant peaks identified by MACS2. y axis shows reads/million/25 bp.

(legend continued on next page)
(Figure 6A and Table S4). These were proximal to cardiac-enriched genes with multiple constituents of G4T5 and robust H3K27ac enhancer marks (Figures 6B and S6A). MED1 ChIP-seq signal positively correlated to gene expression levels (Figure S6B). SEs had 11-fold more MED1 binding than TEs, were longer (3–80 kb; average, 10kb), and induced 4-fold more gene expression (Figures 6C and S6C). As expected, GATA4 and TBX5 binding was enriched in SE elements (Figures 6D and S6D), as were motifs for MEF2C, SRF, TEAD4, SMAD2/3, MEIS1, and NKX2.5, all critical for cardiac development (Figure 6E). SE elements were near genes involved in striated muscle development, cardiomyopathy, heart development, and cardiac muscle contraction (Figure 6F).

In contrast, MED1 ChIP-seq in G296S CMs identified 172 SE elements (Figure 6G). Comparison of SE elements showed loss of 34% (SE1^1), with 66% being unchanged (SE1^2) and 12% being ectopically gained (SE1^3) in mutant CMs (Figure 6H). TBX5 binding in the SE1^2 and SE1^3 elements were markedly reduced, despite comparable GATA4 DNA-binding (Figures 6I and S6E), most likely from disruption of the GATA4-TBX5 interaction and failure of GATA4 to recruit TBX5 to cardiac SEs. Key cardiac genes with lost SE elements included RBM20, SMYD1, and SRF (Figure 6J).

In line with a primed endothelial gene program in mutants, HES1 gained SE elements, as did several members of WNT signaling. RNA-seq showed altered expression of genes in SE elements in mutant CPCs. D15-CMs, and D32-CMs (Figure 6K). SE1 elements were enriched in MEF2A, TEAD4, and NFATC2 motifs, and SE6 elements were enriched in motifs of endothelial regulators, such as HIF1A and FOXP1, as well as MEIS1 and GATA4 motifs (Figure S6F). Downregulated genes from the RNA-seq data were disproportionally enriched for SE elements (Figure S6G).

To identify multivariate relationships between GATA4 and TBX5 binding with cardiac SE gene regulation, we used topological data analysis (TDA), which applies principal component analysis by singular value decomposition (Lum et al., 2013). Related genes are clustered into nodes, and clusters that share > 1 gene are connected via an edge. TDA accurately grouped SE genes into a distinct smaller network that was highly enriched for MED1, TBX5, GATA4, H3K27ac, H3K4me3, and H3K36me3, but not H3K27me3 (Figure S6H and Table S5). This predicted SE network was attenuated by GATA4 knockdown, which supports its biological importance (Figure 6L). Interestingly, TBX5 binding was better correlated in this SE network than GATA4 binding, suggesting that TBX5 is a better predictor of cardiac SE genes than GATA4.

SE elements mapped to several long-non-coding RNAs and TEs with undetermined cardiogenic functions. We hypothesized that they may be required to maintain CM function. Indeed, their depletion in CMs mostly induced abnormalities in contractility, calcium flux, and mitochondria mass (Figure S7A–S7C). Depleting MALAT1 and KLF9 induced a collapse of the cardiac transcriptional network (Figure S7D).

Regulatory Hubs in a GATA4-TBX5 Network Centered on PI3K Signaling

We used a systems-biology approach to construct a GATA4-TBX5 gene regulatory network (GRN) by integrating down- and upregulated genes in G296S CMs (Figure 4), G4T5 bound genes in WT or G296S CMs (Figure 5) and genes with SE elements (Figure 6) with STRING datasets (Table S5). We predicted a “scale-free” network of 716 nodes connected by 2,353 edges with an average 6.6 neighbors and path length of 4.3 (Figure 7A). Nodes were connected by edges representing physical (protein-protein) or functional (genetic, co-expression, co-occurrence) interactions. At least five sub-networks connected through 20 regulatory “hubs” were identified. When we extracted the top 20 hubs as a sub-network connected by 70 edges, each had 27–53 neighbors—4- to 8-fold more than the average node in the GRN (Figure 7B). This sub-network had a significant interaction of p < 6.5e-11. Interestingly, the top four hubs were G4T5 co-bound genes linked to PI3K signaling: PIK3CA (α-catalytic subunit), PIK3R1 (regulatory subunit), and PTK2 and EGFR, the upstream signal transduction

(B) Metagenes plot of normalized ChIP-seq signals for indicated factors at 2,428 G4T5 co-bound sites (± 5 kb) identified in WT CM.

(C) Normalized GATA4 (left) or TBX5 (right) signal at sites that are G4T5 co-bound versus single TF bound. Boxplot and whiskers show mean, 25th, and 75th percentile, followed by 5th and 95th percentile. ***p < 0.0005 (Kolmogorov-Smirnov test).

(D) Known consensus motifs enriched in 2,428 G4T5 co-bound sites in WT CM.

(E) Venn diagram shows changes in GATA4, TBX5, or G4T5 bound sites between WT and G296S CMs. Number of sites lost in WT (L), gained in G296S (E), and unchanged (U) are shown (top row). Legend for metagenes of relative (G296S/WT) ChIP-seq occupancy at sites that are L (blue line), U (green), or E (red) (top row, far-right). 2nd to 4th rows show relative changes in GATA4, TBX5, and H3K27ac occupancy at these L, U, or E sites.

(F) FPKM values of genes mapped ≥ 20 kb of 1,186 G4T5 sites in iWT and G296S cells at three differentiated stages. Boxplot and whiskers show mean, 25th, and 75th percentile, followed by 5th and 95th percentile. **p < 0.005, *p < 0.0005 (Wilcoxon signed-rank test).

(G) Gap distances between GATA4 and TBX5 motifs within G4T5^I versusG4T5^II sites on the same (blue) or different (red) DNA strand. Boxplot and whiskers show mean, 25th, and 75th percentile, followed by 5th and 95th percentile. *p < 0.05 (Fisher’s exact test).

(H) Bar graph showing number of sites with ≥ 1 GATA4-TBX5 motif pairs (left) and number of motif pairs on same or different DNA strands (right) within G4T5^I versus G4T5^II sites. *p < 0.05, ****p < 0.00005 (Fisher’s exact test).

(I) GO analyses of 1186 G4T5^I sites. Significance shown as –Log10 Bonferroni p value after multiple hypothesis correction.

(J) Heatmap shows hierarchical clustering of 414 putative G4T5 target genes in D15-CMs/D32-CMs and changes to GATA4 and TBX5 binding. RNA-seq expression is row-scaled to show relative expression (left). ChIP-seq shows relative (Log2_FC) GATA4, TBX5 occupancy (right). One ChIP-seq peak with the largest fold difference was selected for each gene. Rows between GATA4 and TBX5 are approximately matched. Blue and red are low and high levels, respectively.

(K) Heatmap shows clustering of 82 endothelial genes and changes to GATA4 and TBX5 binding within endothelial TADs. RNA-seq (left) and ChIP-seq (right) show relative (Log2_FC) gene expressions and GATA4, TBX5 occupancy (right). Blue and red are down- and upregulation, respectively. Rows between RNA-seq and ChIP-seq results are approximately matched. See also Figures S4 and S5.
components. In PTK2, G4T5 co-occupancy was lost in GATA4 mutants (Figure 7B). ITGA2, ITGA9, and KDR were also hubs and involved in PI3K signaling. GO analysis showed enrichment for integrin, PI3K-Akt, Phosphatidylinositol and EGF signaling (Figure S7E). When CMs were further treated with a PI3K inhibitor (LY294002), iWT CMs had a decrease in force generation, but G296S CMs were insensitive (Figure 7C, left). While the PI3K inhibitor had some effect on beat rates, the IRS peptide increased beating rates in G296S CMs 3-fold greater than iWT CMs, suggesting a hyper-sensitivity to PI3K pathway activation (Figures 7C, 7D, and S7F). The evidence that mutant CMs exhibit dysregulated PI3K signaling provides a potential node for correcting the diseased GRN.

**DISCUSSION**

Here, we show that proper cardiac development and function require GATA4-TBX5 co-occupancy in MED1-bound, H3K27ac-marked SE elements to maintain an open chromatin state and activate cardiogenic gene transcription (Figure 7E). In GATA4 heterozygosity with a missense mutation that affects protein-protein interactions, a loss of TBX5 recruitment to SE elements is associated with failure to maintain open chromatin and diminished transcription of cardiac genes. The GATA4 G296S mutation allows mis-localization of TBX5 and perhaps other transcriptional activators, resulting in a failure to recruit HDAC2 and achieve a more closed chromatin signature at endothelial promoters. The result is aberrant activation of endothelial gene expression and alternative lineages. Furthermore, dysregulation of genes involved in the reception of SHH signals and cardiac septation provides a molecular basis for three-dimensional septal defects of GATA4 G296S patients, despite the two-dimensional model. These studies show how TF complexes cooperatively regulate genome-wide localization of trans-acting factors to control activation and repression of gene expression and how diseases occur when cooperativity is disrupted.

**GATA4 Maintains Homeostatic CM Function**

GATA4 is a well-known master regulator of early heart development, cardiac specification, and hypertrophy (Bisping et al., 2008). With MEF2C and TBX5, it reprograms fibroblasts to a CM-like fate, and the cooperativity shown here partially explains the induction of a cardiogenic program (Ieda et al., 2010; Qian et al., 2012). That mutant CMs have impaired contractility, calcium handling, sarcomeric organization, and metabolic activity is in line with GATA4 mutations associated with familial cardiomyopathy (Zhao et al., 2014), including those involving GATA4 G296S. Dysregulation of sarcomeric and metabolic genes explains many of the defects in human CMs. Our findings that GATA4 and putative co-repressors function in a negative feedback loop to limit PI3K signaling that becomes dysregulated in GATA4 mutants are consistent with reports of Gata4 mediating PI3K-dependent hypertrophic responses to physiological stress in mouse hearts (McMullen et al., 2004).

**GATA4 Promotes Cardiomyocyte and Represses Alternative Fate Gene Expression**

Our results show that GATA4 is critical for cardiac versus endothelial gene regulation in CPCs. GATA4’s function as a positive driver of cardiogenesis is unambiguous, but its potential as a repressor of endocardial/endothelial gene expression in CMs has been unknown. Scl/Tal1 promotes the hematopoietic gene program in hemogenic endothelium and prevents mis-specification into the cardiomyogenic fate by a combinatorial mechanism (Van Handel et al., 2012). Our data support this concept from the reciprocal angle where a disease-causing mutation of a TF that normally promotes cardiogenesis induces an ectopic endothelial gene program during CM differentiation. TAL1 was upregulated in G296S CPCs and may contribute to aberrant endothelial gene expression. G4T5 sites in CMs were enriched for motifs of key regulators of hemogenic endothelium, FOXO1 and HOXB4, and G4T5 occupancy normally was associated with gene repression at these sites. However, in GATA4 G296S mutants, loci of inappropriately open chromatin were enriched for motifs of endothelial regulators such as FOXO1 and numerous ETS components. In PTK2, G4T5 co-occupancy was lost in GATA4 mutants (Figure 7B). ITGA2, ITGA9, and KDR were also hubs and involved in PI3K signaling. GO analysis showed enrichment for integrin, PI3K-Akt, Phosphatidylinositol and EGF signaling (Figure S7E). When CMs were further treated with a PI3K inhibitor (LY294002), iWT CMs had a decrease in force generation, but G296S CMs were insensitive (Figure 7C, left). While the PI3K inhibitor had some effect on beat rates, the IRS peptide increased beating rates in G296S CMs 3-fold greater than iWT CMs, suggesting a hyper-sensitivity to PI3K pathway activation (Figures 7C, 7D, and S7F). The evidence that mutant CMs exhibit dysregulated PI3K signaling provides a potential node for correcting the diseased GRN.

**DISCUSSION**

Here, we show that proper cardiac development and function require GATA4-TBX5 co-occupancy in MED1-bound, H3K27ac-marked SE elements to maintain an open chromatin state and activate cardiogenic gene transcription (Figure 7E). In GATA4 heterozygosity with a missense mutation that affects protein-protein interactions, a loss of TBX5 recruitment to SE elements is associated with failure to maintain open chromatin and diminished transcription of cardiac genes. The GATA4 G296S mutation allows mis-localization of TBX5 and perhaps other transcriptional activators, resulting in a failure to recruit HDAC2 and achieve a more closed chromatin signature at endothelial promoters. The result is aberrant activation of endothelial gene expression and alternative lineages. Furthermore, dysregulation of genes involved in the reception of SHH signals and cardiac septation provides a molecular basis for three-dimensional septal defects of GATA4 G296S patients, despite the two-dimensional model. These studies show how TF complexes cooperatively regulate genome-wide localization of trans-acting factors to control activation and repression of gene expression and how diseases occur when cooperativity is disrupted.

**GATA4 Maintains Homeostatic CM Function**

GATA4 is a well-known master regulator of early heart development, cardiac specification, and hypertrophy (Bisping et al., 2008). With MEF2C and TBX5, it reprograms fibroblasts to a CM-like fate, and the cooperativity shown here partially explains the induction of a cardiogenic program (Ieda et al., 2010; Qian et al., 2012). That mutant CMs have impaired contractility, calcium handling, sarcomeric organization, and metabolic activity is in line with GATA4 mutations associated with familial cardiomyopathy (Zhao et al., 2014), including those involving GATA4 G296S. Dysregulation of sarcomeric and metabolic genes explains many of the defects in human CMs. Our findings that GATA4 and putative co-repressors function in a negative feedback loop to limit PI3K signaling that becomes dysregulated in GATA4 mutants are consistent with reports of Gata4 mediating PI3K-dependent hypertrophic responses to physiological stress in mouse hearts (McMullen et al., 2004).

**GATA4 Promotes Cardiomyocyte and Represses Alternative Fate Gene Expression**

Our results show that GATA4 is critical for cardiac versus endothelial gene regulation in CPCs. GATA4’s function as a positive driver of cardiogenesis is unambiguous, but its potential as a repressor of endocardial/endothelial gene expression in CMs has been unknown. Scl/Tal1 promotes the hematopoietic gene program in hemogenic endothelium and prevents mis-specification into the cardiomyogenic fate by a combinatorial mechanism (Van Handel et al., 2012). Our data support this concept from the reciprocal angle where a disease-causing mutation of a TF that normally promotes cardiogenesis induces an ectopic endothelial gene program during CM differentiation. TAL1 was upregulated in G296S CPCs and may contribute to aberrant endothelial gene expression. G4T5 sites in CMs were enriched for motifs of key regulators of hemogenic endothelium, FOXO1 and HOXB4, and G4T5 occupancy normally was associated with gene repression at these sites. However, in GATA4 G296S mutants, loci of inappropriately open chromatin were enriched for motifs of endothelial regulators such as FOXO1 and numerous ETS components. In PTK2, G4T5 co-occupancy was lost in GATA4 mutants (Figure 7B). ITGA2, ITGA9, and KDR were also hubs and involved in PI3K signaling. GO analysis showed enrichment for integrin, PI3K-Akt, Phosphatidylinositol and EGF signaling (Figure S7E). When CMs were further treated with a PI3K inhibitor (LY294002), iWT CMs had a decrease in force generation, but G296S CMs were insensitive (Figure 7C, left). While the PI3K inhibitor had some effect on beat rates, the IRS peptide increased beating rates in G296S CMs 3-fold greater than iWT CMs, suggesting a hyper-sensitivity to PI3K pathway activation (Figures 7C, 7D, and S7F). The evidence that mutant CMs exhibit dysregulated PI3K signaling provides a potential node for correcting the diseased GRN.
Figure 7. GATA4-TBX5 GRN Revealed Hubs Centered on PI3K Signaling

(A) GATA4-controlled GRN. Nodes are genes that are differentially expressed or G4T5 co-bound or have MED-1 SE elements. Edges are physical or functional interactions between nodes as extracted from STRING. Yellow, top 20 hubs with the most direct neighbors. Hubs are grouped into five sub-networks (pink circle).

(B) Sub-network plot of extracted top 20 hubs named by gene symbol. Number of edges from entire GRN shown beside each node. Blue or red are gene expressions down- or upregulated, respectively. Diamond, square, or circle represents genes that gained, lost, or were unchanged for G4T5 binding. Bolded border represents genes with SE elements.

(C) Relative change in force generation between iWT (black) and G296S (red) CMs after inhibition (circle) or activation (triangle) of PI3K signaling. Traction force microscopy (TFM) measurements of CMs responding accurately to 1 Hz pacing. Data are mean ± SEM, *p < 0.05, **p < 0.005, ***p < 0.0005 (Mann-Whitney U test).

(D) Beat rate measurements between iWT (black) and G296S (red) CMs after inhibition (circle) or activation (triangle) of PI3K signaling. TFM measurements of CMs responding accurately to 1 Hz pacing. Data are mean ± SEM, *p < 0.05,** p < 0.005, ***p < 0.0005 (Mann-Whitney U test).

(E) Proposed model. Top, cardiac gene loci in WT are open and permissive to G4T5 binding at MED1-bound SE elements, which activates transcription; G4T5 and HDAC2 repress aberrant endothelial gene transcription. Bottom, transcriptional and epigenetic consequences of GATA4 G296S. Cardiac gene loci have reduced open chromatin and TBX5 binding to SE elements which reduces transcription; aberrantly open chromatin is depleted of GATA4-HDAC2 but enriched for TBX5, along with motifs for ETS factors resulting in failure to silence endothelial gene transcription and other sites involved in septal development not depicted. See also Figure S7.
factors, suggesting loss of G4T5 repression. The reduction in HDAC2 recruitment and downregulation of the GATA4-interacting repressors, HEY1 and HEY2, provide a potential mechanism for de-repression of endothelial gene targets that may contribute to septal defects.

Even in a monolayer system, genome-wide analyses revealed gene expression and chromatin dysregulation of genes required for atrioventricular canal development, endocardial cushion formation, and septal morphogenesis in GATA4 G296S CMs. These observations suggest that iPS cells can be used as an in vitro model to understand cellular events leading to morphogenetic defects. Specifically, TBX5, a regulator of AV canal myocardium (Aanhaanen et al., 2011), was upregulated, and genes necessary for receiving the SHH signal in myocardium were downregulated. This was particularly interesting because exogenous SHH signals from pulmonary endoderm are received by the developing atrium, resulting in expansion of the posterior second heart field-derived dorsal-mesenchymal-protrusion that forms part of the atrial septum. Failure to respond to the SHH signal results in septal defects in mice (Hoffmann et al., 2009). Evidence for GATA4 regulating SHH signaling suggests a potential mechanism for septal defects observed in mice and humans haploinsufficient for GATA4.

**Combinatorial Regulation of Human Cardiac Enhancers**

Our results show a combinatorial TF binding code for activating the human cardiac gene program, similar to mouse CMs (Luna-Zurita et al., 2016), and reveal how disrupting this code by a missense mutation leads to epigenetic and transcriptional dysregulation and human disease. ATAC-seq analyses of open chromatin signature and genome-wide profiling of GATA4 and TBX5 binding sites provide a detailed catalog of TF-bound enhancers in humans and complement the sparse ENCODE data on cardiac cell types, which we leveraged in identifying a putative G4T5 corepressor. We found that GATA4 and TBX5 cooperation was robust when underlying cis-sequences were closely linked on the same DNA strand and in the same 5’–3’ orientation; in such situations, protein-DNA interactions may overcome a lack of protein-protein interaction between GATA4 and TBX5 in the mutant setting.

Until now, human cardiac SEs had not been identified by MED1 ChIP-seq. Our cataloging of SEs pinpoints transcriptional regulators and long noncoding RNAs that may be crucial in human cardiac development and function. TDA with machine learning distinguished genes with SE elements from other genes and placed TBX5 at a higher hierarchical level than GATA4 in mapping cardiac SEs. TE5 seem to have a different TF binding code than SEs. In GATA4 mutants, TBX5 binding was decreased at SEs but increased at many TEs. This difference suggests cardiac TFs operate via diverse rules at various enhancer sites, perhaps dictated by underlying cis-sequence and/or local chromatin configuration.

In conclusion, this study reveals a combinatorial TF code that ensures robust cardiac gene program, illustrates how human disease occurs when this code is altered by disrupting TF cooperativity, and highlights potential nodes for therapeutic intervention.
for assistance with manuscript preparation. Generous software support from Ayasdi was provided for topological data analysis. This work was supported by NIH P01 HL089707 (D.S.), U01 HL098179 (D.S.), R01 HL057181 (D.S.), U01 HL100406 (D.S.); R21 HL13099301 (B.L.P.), R01 GM097358 (H.Y.), R01 GM104424 (H.Y.), U1MI HG003933 (H.Y.), R01 HD082568 (H.Y.); SFARI Grant 367561 (H.Y.), R01 EB006745 (B.L.P.); by AHA 14POST18360018 (A.J.S.R.) and 13POST17390040 (Y.-S.A.); by NSF MICS-1136790 (B.L.P.); and by Damon Runyon Foundation DRG-2187-14 (R.S.). D.S. was supported by the Roddenberry Foundation, the L.K. Whittier Foundation, and the Younger Family Fund. This work was also supported by NIH/NCRR grant C06 RR018828 to the Gladstone Institutes. This work is dedicated to Ian Spencer. D.S. is a scientific co-founder of Tenaya Therapeutics.

Received: February 23, 2016
Revised: August 9, 2016
Accepted: November 17, 2016
Published: December 15, 2016

REFERENCES


