Tbx20 regulates a genetic program essential to adult mouse cardiomyocyte function

Tao Shen, Ivy Aneas, Noboru Sakabe, Ralf J. Dirschinger, Gang Wang, Scott Smemo, John M. Westlund, Hongqiang Cheng, Nancy Dalton, Yusu Gu, Cornelis J. Boogerd, Chen-leng Cai, Kirk Peterson, Ju Chen, Marcelo A. Nobrega, Sylvia M. Evans

Research Article   Cardiology

Human mutations in or variants of *TBX20* are associated with congenital heart disease, cardiomyopathy, and arrhythmias. To investigate whether cardiac disease in patients with these conditions results from an embryonic or ongoing requirement for Tbx20 in myocardium, we ablated *Tbx20* specifically in adult cardiomyocytes in mice. This ablation resulted in the onset of severe cardiomyopathy accompanied by arrhythmias, with death ensuing within 1 to 2 weeks of *Tbx20* ablation. Accounting for this dramatic phenotype, we identified molecular signatures that posit Tbx20 as a central integrator of a genetic program that maintains cardiomyocyte function in the adult heart. Expression of a number of genes encoding critical transcription factors, ion channels, and cytoskeletal/myofibrillar proteins was downregulated consequent to loss of Tbx20. Genome-wide ChIP analysis of Tbx20-binding regions in the adult heart revealed that many of these genes were direct downstream targets of Tbx20 and uncovered a previously undescribed DNA-binding site for Tbx20. Bioinformatics and in vivo functional analyses revealed a cohort of transcription factors that, working with Tbx20, integrated multiple environmental signals to maintain ion channel gene expression in the adult heart. Our data provide insight into the mechanisms by which mutations in TBX20 cause adult heart disease in humans.

Find the latest version:

https://jci.me/59472/pdf

Research article

Tbx20 regulates a genetic program essential to adult mouse cardiomyocyte function

Tao Shen,[1,2] Ivy Aneas,[3] Noboru Sakabe,[3] Ralf J. Dirschinger,[1] Gang Wang,[1] Scott Smemo,[3] John M. Westlund,[3] Hongqiang Cheng,[4] Nancy Dalton,[4] Yusu Gu,[4] Cornelis J. Boogerd,[1] Chen-leng Cai,[1] Kirk Peterson,[4] Ju Chen,[4] Marcelo A. Nobrega,[3] and Sylvia M. Evans[1,4]

[1]Skaggs School of Pharmacy, UCSD, La Jolla, California, USA. [2]Key Laboratory of Geriatrics, Beijing Hospital and Beijing Institute of Geriatrics, Ministry of Health, Beijing, China. [3]Department of Human Genetics, University of Chicago, Chicago, Illinois, USA. [4]Department of Medicine, UCSD, La Jolla, California, USA.

Human mutations in or variants of *TBX20* are associated with congenital heart disease, cardiomyopathy, and arrhythmias. To investigate whether cardiac disease in patients with these conditions results from an embryonic or ongoing requirement for Tbx20 in myocardium, we ablated *Tbx20* specifically in adult cardiomyocytes in mice. This ablation resulted in the onset of severe cardiomyopathy accompanied by arrhythmias, with death ensuing within 1 to 2 weeks of *Tbx20* ablation. Accounting for this dramatic phenotype, we identified molecular signatures that posit Tbx20 as a central integrator of a genetic program that maintains cardiomyocyte function in the adult heart. Expression of a number of genes encoding critical transcription factors, ion channels, and cytoskeletal/myofibrillar proteins was downregulated consequent to loss of Tbx20. Genome-wide ChIP analysis of Tbx20-binding regions in the adult heart revealed that many of these genes were direct downstream targets of Tbx20 and uncovered a previously undescribed DNA-binding site for Tbx20. Bioinformatics and in vivo functional analyses revealed a cohort of transcription factors that, working with Tbx20, integrated multiple environmental signals to maintain ion channel gene expression in the adult heart. Our data provide insight into the mechanisms by which mutations in TBX20 cause adult heart disease in humans.

Introduction

Coordinated expression of transcription factors in adult heart regulates gene expression programs responsible for cardiomyocyte function. As many of these factors are also essential for cardiac development, it is difficult to assess whether adult phenotypes consequent to mutation of a given transcription factor reflect a developmental requirement or the acquisition of new functions in adult heart. The notion that transcription factors carry out a distinct set of biological roles in embryonic and adult myocytes raises the possibility that some forms of adult cardiac diseases may result from previously unappreciated adult-specific functions of transcription factors that possess well-established roles in cardiogenesis. Addressing this issue necessitates examination of transcription factor function specifically in adult cardiomyocytes.

Mutations in the gene encoding the T-box transcription factor TBX20 are associated with various forms of congenital heart disease in humans, including defects in septation, valvulogenesis, and chamber growth (1–3). Recently, population-based genome-wide association studies have identified a common single-nucleotide polymorphism associated with QRS duration in humans (4). These congenital cardiac defects are well aligned with previously demonstrated critical roles of Tbx20 in heart development, driving cardiac proliferation and a chamber-specific program of gene expression. Absence of Tbx20 leads to a severely hypoplastic heart and embryonic lethality at midgestation (5–8). Recently, *TBX20* mutations have also been found in patients with dilated cardiomyopathy (1). In contrast with previously described congenital morphological

phenotypes, it is not clear whether these functional abnormalities in adult heart reflect an earlier developmental requirement for TBX20 or an ongoing requirement in adult cardiomyocytes.

Previous studies of transcription factors associated with Tbx20 have demonstrated critical functions for these factors across wide developmental stages. Tbx20 interacts directly with Gata4, Gata5, Nkx2-5 (9), and Tbx5 (10) to regulate gene expression during embryogenesis. Mutations in genes encoding these factors in mice and humans consistently result in an array of structural and conduction cardiac defects of varying severity (11–23).
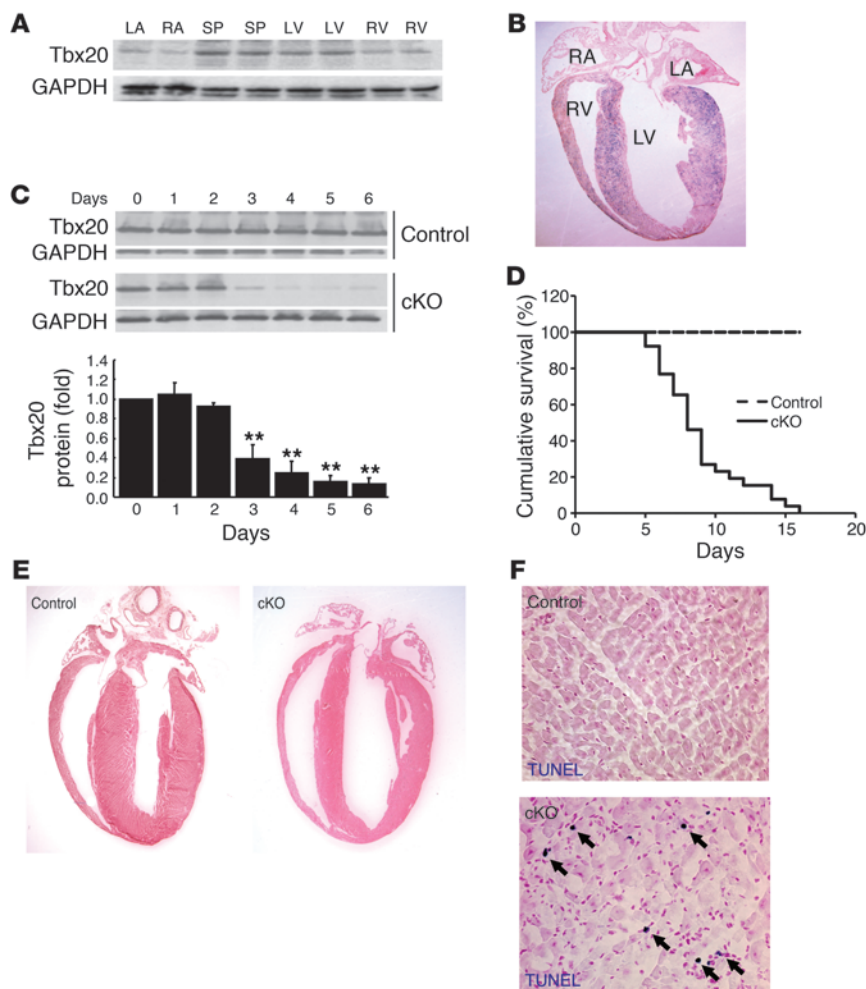
Tissue-specific ablation experiments in mice have been performed for some of these transcription factors, allowing insight into their function in developing or adult myocardium. Ablation of *Nkx2-5* in cardiomyocytes results in a hypoplastic atrioventricular (AV) node at birth, leading to development of AV block, but also leads to abnormal trabecular muscle growth, resembling some of the characteristic features of some affected human individuals (24). Cardiomyocyte-specific ablation of *Gata4* suggests a late developmental role of Gata4 for cardiomyocyte proliferation followed by an adult requirement for hypertrophy and cell survival in response to pressure overload (25, 26), and observations from heterozygous Gata4 mutant mice support this idea (27). Heterozygous hypomorphic or null Tbx5 mice exhibit electrophysiological abnormalities in addition to congenital cardiac malformations (28, 29). While these factors appear to be involved in congenital heart diseases, their additional adult disease phenotypes are quite distinct. As several of these studies demonstrate that these known cofactors of Tbx20 have important roles in adult cardiomyocytes, they raise the likelihood that Tbx20 itself may also be required in adult cardiomyocytes.

To date, no report has described phenotypes arising from cardiomyocyte-specific ablation of *Tbx20* in adult mice. Therefore, it is

Figure 1
Cardiomyocyte-specific ablation of *Tbx20* in adult mice results in cardiac chamber dilation and lethality. (**A**) Tbx20 protein expression in different fractions of the heart (representative result from 3 independent experiments). (**B**) α-*MHC–mER–Cre–mER* activity shown by X-gal in *R26R^lacZ/lacZ* hearts after 5 days of tamoxifen treatment. Original magnification, ×10. (**C**) Time course of *Tbx20* expression after initiation of tamoxifen treatment in α-*MHC–mER–Cre–mER/Tbx20^f/f* mice. Western blot (top) and densitometry analysis (bottom); representative results from 4 independent experiments. Data are expressed as mean ± SEM. **P < 0.01. (**D**) Survival of *Tbx20* conditional KO and control mice. Control, n = 25 (*Tbx20 flox*, n = 15; α-*MHC–mER–Cre–mER*, n = 10); *Tbx20* conditional KO (cKO), n = 26. (**E**) H&E staining of *Tbx20* conditional KO and control hearts 7 days after tamoxifen treatment. Original magnification, ×10. (**F**) TUNEL staining 5 days after initiation of tamoxifen treatment. Original magnification, ×200. Arrows indicate TUNEL-positive nuclei. LA, left atrium; RA, right atrium; SP, interventricular septum; LV/RV, LV/RV free wall.

not clear whether the association of mutations in *TBX20* with adult human cardiomyopathy results solely from a late-onset phenotype pursuant to congenital defects or reflects an ongoing functional requirement for Tbx20 in adult myocardium that is distinct from its developmental roles. Given that the primary role of Tbx20 during heart development is to promote proliferation and that adult cardiomyocytes are quiescent, there is a strong possibility that *Tbx20* acquires a new set of functions in the postproliferative heart.

To investigate functional roles of Tbx20 in adult myocardium, we ablated *Tbx20* specifically in adult cardiomyocytes. Remarkably, this resulted in the onset of global cardiac dysfunction that progressed rapidly to heart failure and lethality. Combining genome-wide determination of Tbx20-binding sites, transcriptome analyses, and in vivo enhancer assays, we demonstrate that Tbx20 directly regulates a number of genes critical to excitation-contraction functions in adult myocytes. Many of these direct downstream targets have been previously associated with cardiomyopathy in humans. We have also identified what we believe is a novel cohort of transcription factors that cooperate with Tbx20 to drive expression of a subset of Tbx20 targets. Taken together, our results identify Tbx20 as a key determinant of adult cardiac function and identify a genetic network regulating cardiac ion flux at a transcriptional level, establishing a molecular pathway connected to the development of heart failure.

Results

Ablation of Tbx20 in adult mouse myocardium results in dilation of cardiac chambers and lethality within 15 days. To address whether Tbx20 may be playing a role in adult heart, we first determined the expression profile of *Tbx20* in hearts of adult mice. Tbx20 protein is present at the highest levels in the LV free wall and ventricular septum, at intermediate levels in the RV, and at the lowest levels in the left and right atria (Figure 1A).

To directly evaluate the role of *Tbx20* in adult cardiomyocytes, we generated adult cardiomyocyte-specific *Tbx20*-KO mice by injecting 6-week-old α-MHC–mER–Cre–mER/*Tbx20^f/f*/R26R–lacZ/lacZ mice with tamoxifen intraperitoneally daily for 5 consecutive days (referred to as *Tbx20* conditional KO mice in the remainder of the manuscript).

Following tamoxifen injections, X-gal staining to monitor α-*MHC–mER–Cre–mER* excision efficiency demonstrated Cre-mediated excision within a substantial number of ventricular myocytes, with little to no excision in atria (Figure 1B). *Tbx20* mRNA levels were determined by quantitative RT-PCR (qRT-PCR) (not shown), and protein levels were analyzed by Western blot analyses of total protein isolated from cardiac ventricles (Figure 1C). Relative to levels in control littermates, *Tbx20* mRNA and protein expression in mutant ventricles was substantially downregulated by day 3 following tamoxifen injec-
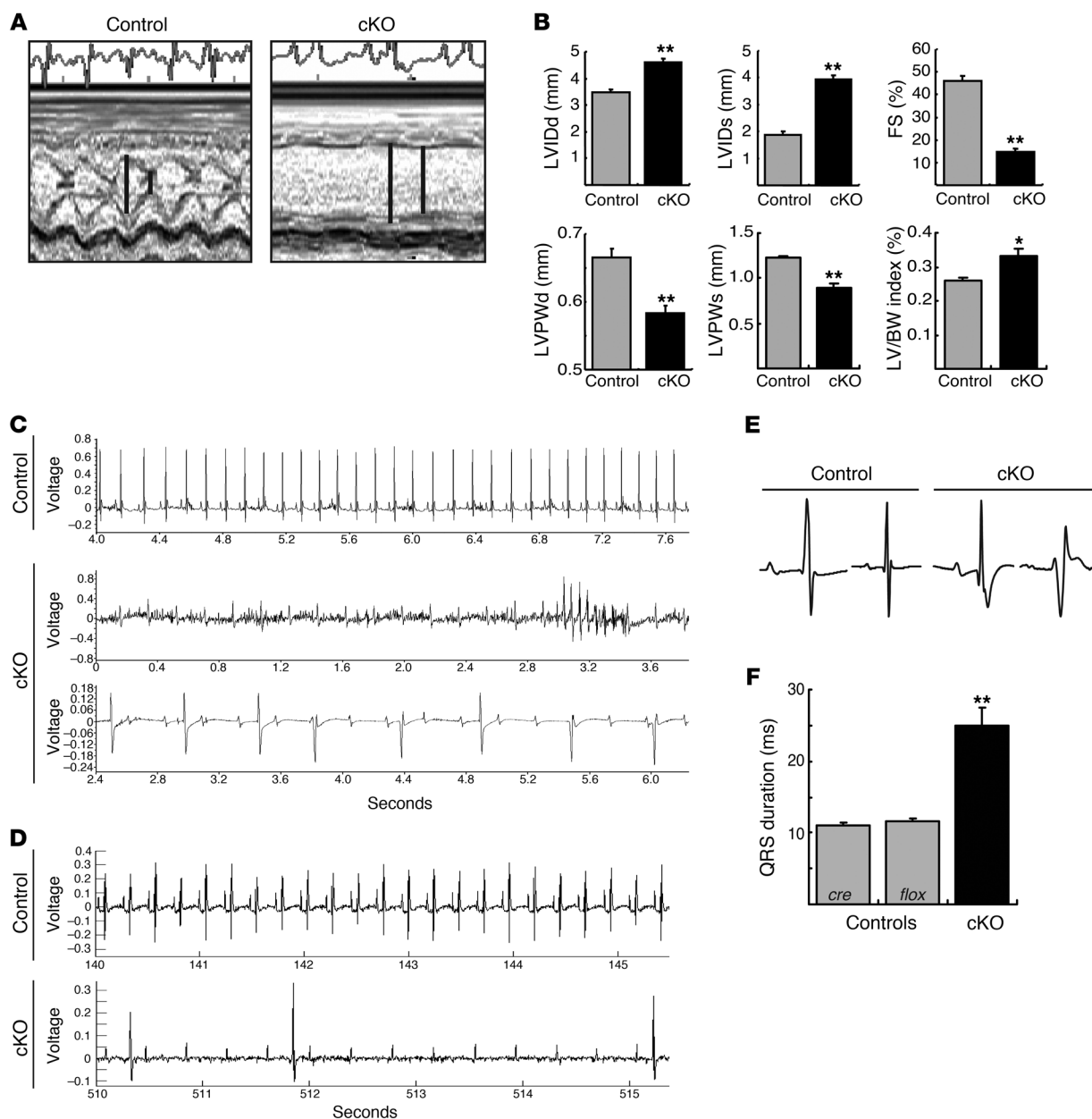
Figure 2
Tbx20 ablation in adult mouse myocardium results in cardiomyopathy and arrhythmia. (**A** and **B**) Echocardiographic M-mode tracings (**A**) and measurements (**B**) 6 days after tamoxifen treatment. *$P < 0.05$; **$P < 0.01$. Control (α-MHC–mER–Cre–mER), $n = 6$; *Tbx20* conditional KO, $n = 13$. (**C**) Telemetry tracings recorded 5–6 days after tamoxifen treatment showing examples of tachyarrhythmia (middle panel) and complete heart block (lower panel) in *Tbx20* conditional KO mice. (**D**) ECG recordings from explanted hearts, showing complete heart block in the mutant. (**E** and **F**) ECG morphology (**E**) and QRS duration (**F**) 5 days after tamoxifen treatment. Controls in both panels are α-MHC–mER–Cre–mER (left) (*cre*) and *Tbx20 flox* (right) (*flox*). **$P < 0.0001$. Control (α-MHC–mER–Cre–mER), $n = 18$; control (*Tbx20 flox*), $n = 11$; *Tbx20* conditional KO, $n = 14$. Data are expressed as mean ± SEM.

tion. Together, these results demonstrated the efficacy of the conditional KO system.

Mice with *Tbx20* ablated in cardiomyocytes exhibited a severe phenotype 3 to 4 days after initiation of tamoxifen treatment, with decreased motor activity and death occurring within 5–16 days (Figure 1D). Control littermates exhibited no adverse effects of tamoxifen treatment, with similar results in control mice carrying

the α-MHC–mER–Cre–mER transgene compared with mice without the transgene. Histological analysis with H&E staining of mutant hearts revealed enlarged right and left cardiac ventricles with thinner walls (Figure 1E). Ratios of heart weight to body weight and heart weight to tibia length were significantly increased in *Tbx20* conditional KOs (heart weight:body weight, 0.0075 vs. 0.0047, $P < 0.01$; heart weight:tibia length, 0.091 vs. 0.064, $P < 0.001$).
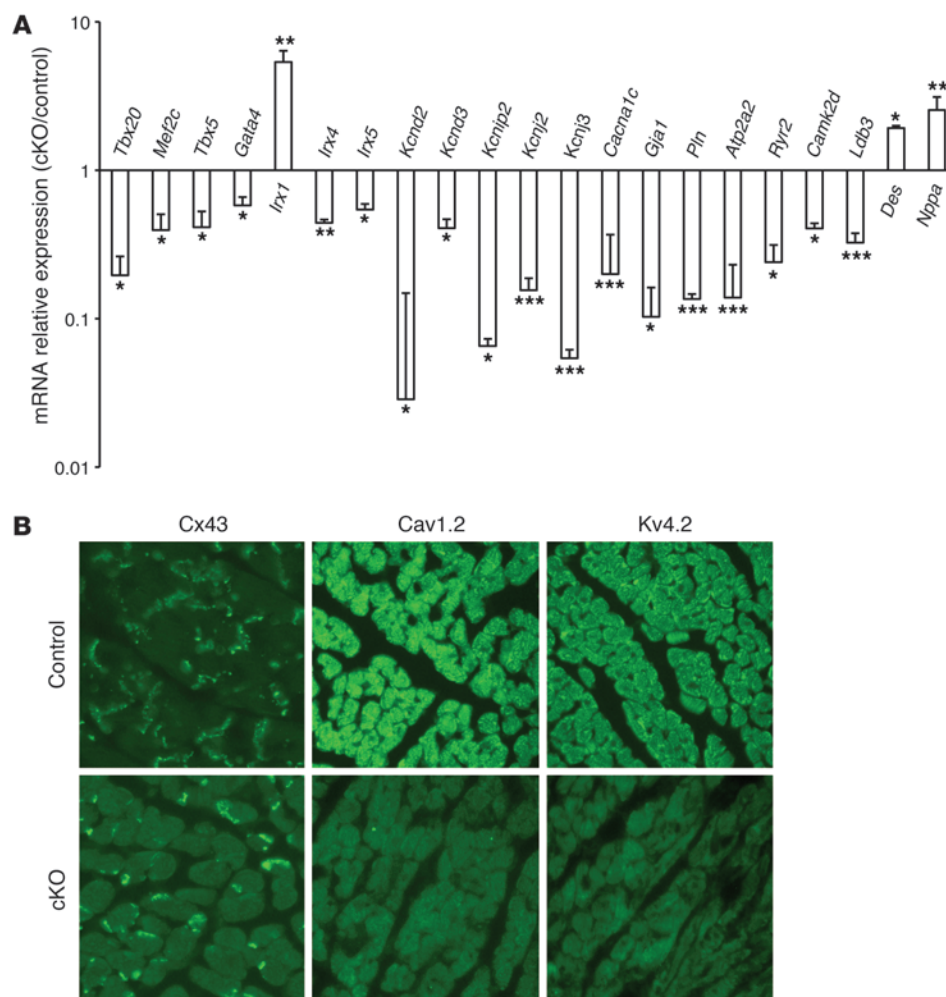
Figure 3
Loss of *Tbx20* in adult myocardium leads to changes in expression of important cardiac genes. (**A**) Genes with significantly altered expression in the hearts of *Tbx20* conditional KO mice relative to control mice 3 days after initiation of tamoxifen treatment. mRNA expression was analyzed by qRT-PCR and normalized to expression of *GAPDH*. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. Control, $n = 8$; *Tbx20* conditional KO, $n = 12$. Data are expressed as mean ± SEM. (**B**) Immunostaining for critical ion channel proteins 4 (Cx43) or 5 (Cav1.2 and Kv4.2) days after tamoxifen treatment. Original magnification, ×400 (Cx43); ×200 (Cav1.2 and Kv4.2). For each panel, 1 representative of at least 6 similar experiments is shown.

Increased cell death was observed in mutant hearts relative to controls by TUNEL staining 5 days after tamoxifen injection (Figure 1F). Early signs of fibrosis detected by trichrome staining were observed in mutant hearts 7 days after tamoxifen injection (Supplemental Figure 1A; supplemental material available online with this article; doi:10.1172/JCI59472DS1). Extensive fibrosis observed within a short time frame in our adult myocyte-specific KOs of Tbx20 was usually detected during acute ischemic episodes. The cause of the fibrosis is not yet clear, but will be a subject of future investigation.

Adult cardiomyocyte-specific Tbx20 conditional KO mice exhibit a severe cardiomyopathy with heart failure and arrhythmias. Our histological analysis suggested a dilated cardiomyopathy in mutant hearts. We therefore assessed heart function in vivo by echocardiography on mutant and control mice. *Tbx20* conditional KO mice showed enlarged LVs and reduced thickness of the LV walls, quantified by M-mode measurements 6 days after tamoxifen induction (Figure 2, A and B, and Supplemental Table 1). Changes in cardiac dimensions in mutant mice were accompanied by a dramatic decrease in LV systolic function, observed by a significant decrease in fractional shortening (%FS) and velocity of circumferential fiber shortening (VCF) compared with littermate controls. Reduction of cardiac function could be detected as early as 3 days after tamoxifen injection and was severe at day 4 after injection (Supplemental Data Files 1 and 2 [videos]).

ECG analyses showed a variety of abnormalities in *Tbx20* conditional KOs, including slower heart rate, changes in AV conduction, and altered ventricular depolarization and repolarization as well as tachyarrhythmias and bradyarrhythmias (Figure 2, C–F). To eliminate the effect of anesthesia, which is required for ECG analysis, we confirmed these findings by telemetric recordings in awake mice (Figure 2C) and on isolated hearts (Figure 2D) to further minimize effects caused by vagal innervation or circulating catecholamines. Mutant mice exhibited a widening of QRS complexes compared with controls (QRS duration 11.07 ± 0.38 ms vs. 24.92 ± 2.41 ms; $P < 0.0001$) and an elevation or depression of the ST segment (Figure 2, E and F). Mutant mice further showed sinus bradycardia and progressive AV block, resulting in complete heart block with a typically slow and in some cases polymorphic ventricular escape rhythm as a preterminal rhythm (Figure 2, C and D). Additionally, telemetry tracing revealed self-limiting polymorphic wide-QRS tachycardias in some of the mice (Figure 2C), suggesting a severe derangement of normal electrophysiology in mutant hearts.

Together, these results reveal an ongoing requirement for Tbx20 in adult heart function and show that loss of Tbx20 results in heart failure. Supporting this notion, we observed a reduction in *Tbx20* mRNA and protein levels in other murine models of heart failure (Supplemental Figure 2).
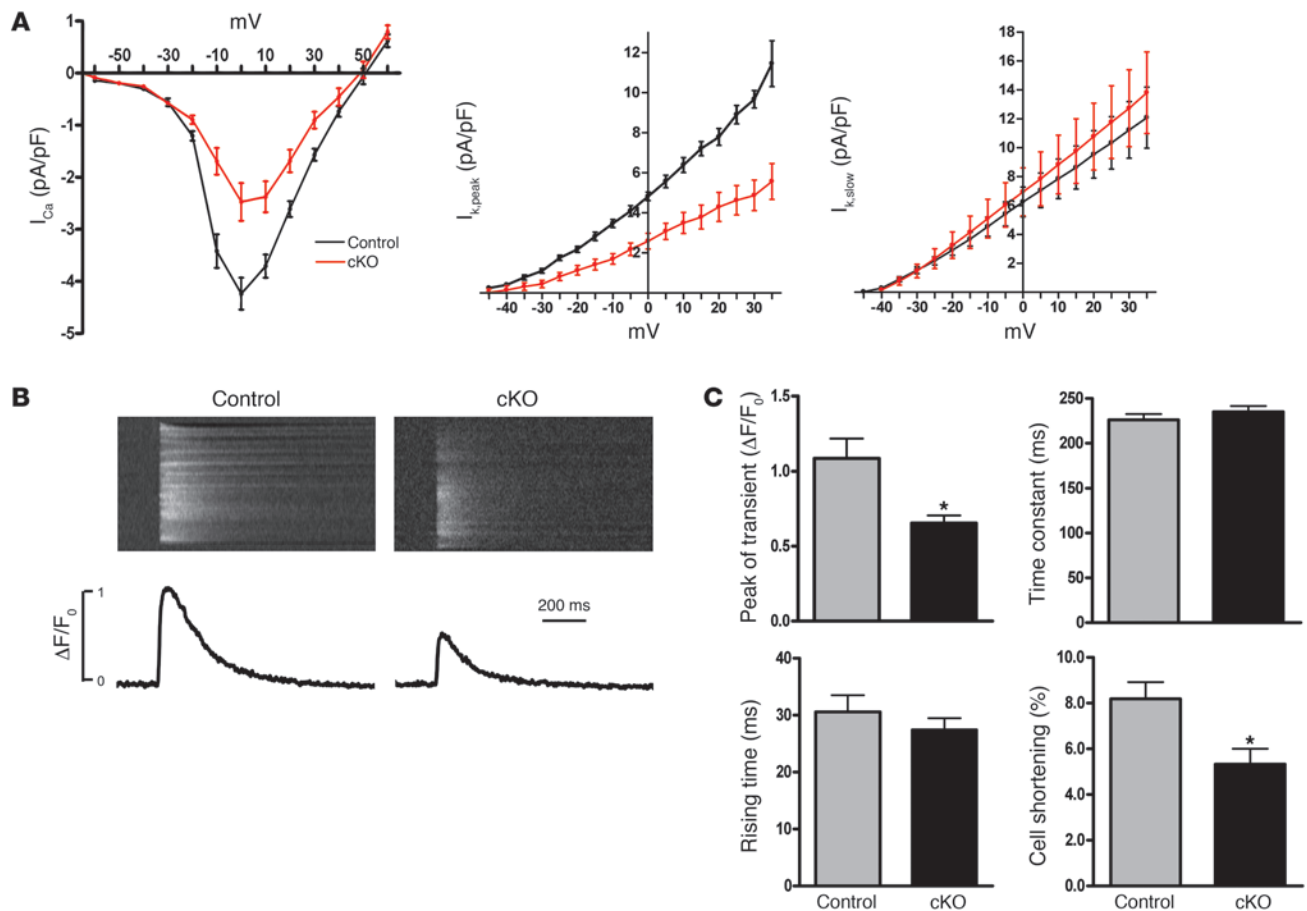
Figure 4
Single-cell electrophysiology analysis reveals altered ion flux regulation in adult *Tbx20*-null cardiomyocytes. (**A**) L-type calcium current (left), peak inactivating outward potassium current (middle), and slow outward potassium current (right) as measured by whole-cell patch clamp analysis. Control, *n* = 7; *Tbx20* conditional KO, *n* = 8. (**B** and **C**) Tracings (**B**) and measurements (**C**) from confocal microscopy calcium imaging show altered calcium cycling in *Tbx20⁻/⁻* cardiomyocytes. *$P < 0.05$. Control, *n* = 7; *Tbx20* conditional KO, *n* = 8. Data are expressed as mean ± SEM.

Loss of Tbx20 in adult cardiomyocytes leads to changes in expression of genes required for transcriptional regulation, ion transport, and contractility and disrupts critical gene expression gradients in myocardium. To evaluate molecular mechanisms underlying the fulminant disease in *Tbx20* conditional KOs, we analyzed expression of a number of candidate genes expressed in adult myocardium that might be involved in the development of the phenotype. Expression levels of candidates including transcription factors, ion channels, gap junction proteins, and proteins involved in calcium cycling as well as cytoskeletal and sarcomeric proteins were assayed by qRT-PCR (Figure 3).

At 3 days after tamoxifen induction, we found marked decreases in mRNA expression of cardiac transcription factors *Mef2c*, *Tbx5*, *Gata4*, *Irx4*, and *Irx5* in hearts of *Tbx20* conditional KO mice compared with controls, while the transcription factor Irx1 was substantially upregulated in mutant mice. Furthermore, we found downregulation in expression of numerous genes encoding proteins with critical functions in cardiomyocytes. These included genes encoding the gap junction protein Gja1 (connexin 43); potassium channel genes *Kcnd2*, *Kcnd3*, *Kcnj2*, *Kcnj3*, *Kcnip2*, and *Kcnh2*; calcium channel genes *Cacna1c*, *Cacna1g*, *Cacna2d1*, *Cacnb2*, and *Cachd1*; genes encoding calcium cycling or regulatory proteins *Serca2* (*Atp2a2*), phospholamban (*Pln*), ryanodine receptor 2 (*Ryr2*), and calmodulin-dependent pro-

tein kinase type II delta chain (*Camk2d*); and those encoding cytoskeletal proteins cypher (*Ldb3*) and desmin (*Des*). In contrast, genes for β-tubulin (*Tubb2b*, not shown) and atrial natriuretic factor (*Nppa*) were significantly upregulated (Figure 3A). Expression of other candidate genes, *Gata6*, *Tbx3*, *Irx2*, *Irx3*, *Nkx2-5*, or *Kcne1*, was not significantly altered in *Tbx20* conditional KO mice (not shown).

Several genes whose expression was downregulated in *Tbx20* conditional KO mice were further evaluated at the protein level by Western blot analysis of cardiac protein extracts obtained 5 days following tamoxifen injection (Supplemental Figure 1, B and C). Consistent with decreases in mRNA levels, most of these proteins were downregulated. Only 2 of the tested proteins, *Girk1* and *Serca2*, appeared unchanged despite a marked downregulation of the cognate *Kcnj3* and *Atp2a2* mRNAs. These differences may indicate a slower turnover of some proteins and/or compensatory alterations in posttranscriptional gene regulation.

Immunohistochemical analysis of the main ventricular gap junction protein α 1 (connexin 43) revealed a striking loss of cell-cell connections between neighboring cardiomyocytes compared with that in control hearts (Figure 3B). Staining with antibodies for the L-type calcium channel *Cav1.2* and the potassium voltage-gated channel Kv4.2 likewise demonstrated reduced expression in mutant hearts.
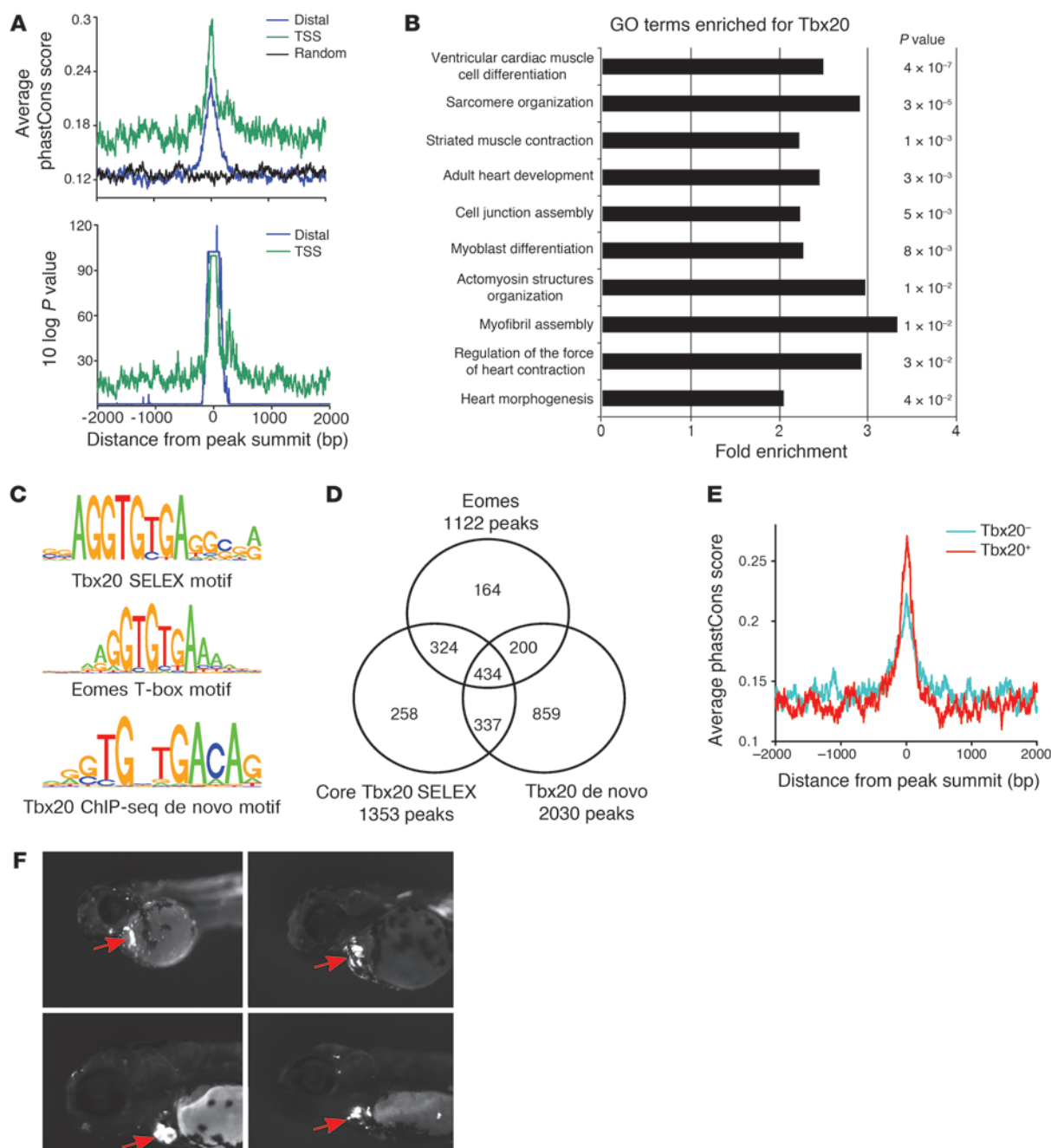
Figure 5
Validation of *Tbx20* ChIP peaks as long distance enhancers. (**A**) Distribution of *Tbx20* around the TSS. *Tbx20*-binding regions are more conserved than the rest of the genome. AUC of *Tbx20* peaks from −200 to +200 bp is 20.4% of the total possible area versus 12.75% for 1,000 sets of 4,012 random genomic regions with the same sizes of the *Tbx20* peaks (1.6-fold excess). (**B**) Selected GO terms of *Tbx20* peaks (for complete data set, see Supplemental Data File 3). (**C**) *Tbx20* de novo motif identified ab initio in *Tbx20* ChIP peaks, SELEX *Tbx20* motif, Eomes primary motif. (**D**) Distribution and overlap of different *Tbx20* motifs among the identified peaks. (**E**) *Tbx20*-binding regions without a *Tbx20* de novo motif are similarly conserved, being 8% less conserved at the ±200-bp region. (**F**) Validation of *Tbx20* ChIP peaks as long distance enhancers. Lateral view of a representative transgenic zebra fish embryo at 48 hours post-fertilization (hpf). Upper panel: *Kcnd2*, *Cacna1c*; lower panel: *Irx3/Irx5*, *Kcne*. Coordinates of the sequences tested near each gene are shown in Supplemental Table 4. Red arrow points to GFP expression in the heart.

Tbx20⁻/⁻ cardiomyocytes exhibit altered ion currents, decreased calcium transient amplitudes, and reduced cell shortening during contraction. The combination of phenotypes and gene expression analysis in adult *Tbx20* myocardial mutant mice suggests that *Tbx20* conditional KO mice may be developing dilated cardiomyopathy in part due to altered ion flux regulation. To directly test this hypothesis, we performed electrophysiological analyses of isolated cardiomyocytes.
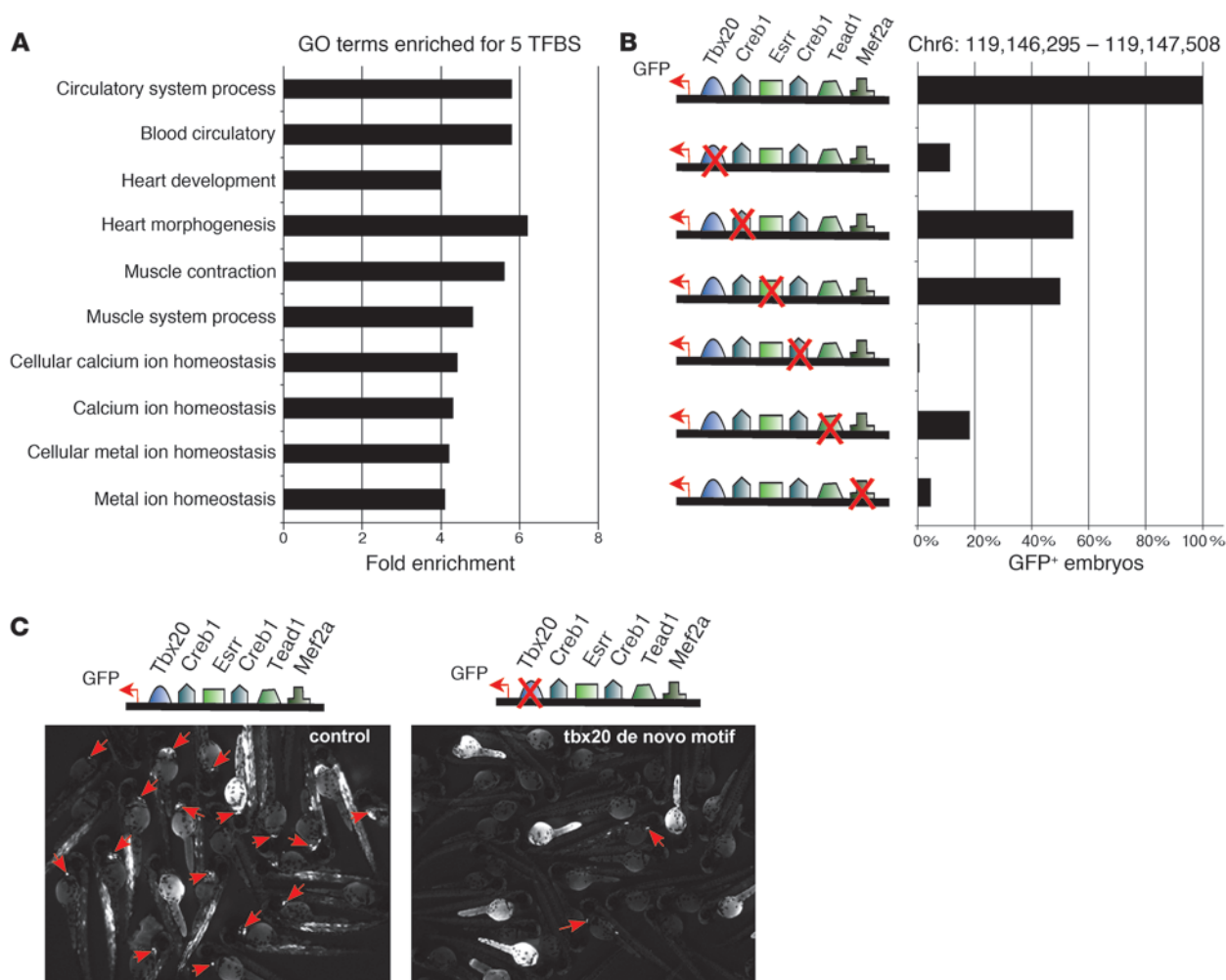
Figure 6
Cooccurrence of transcription factor–binding motifs for *Tbx20*, Mef2a, Tead1, Esrr, and Creb1 is important for gene expression regulation of ion transport genes. (**A**) Top10 enriched GO terms for peaks containing the combination of 5 TFBS. (**B**) Independent mutagenesis of motifs located in *Tbx20* peaks near *Cacna1c* gene. (**C**) Mutagenesis of *Tbx20* de novo motif shows abrogation of enhancer function (right) compared with control (left). Arrows point to Cacna1c enhancer driving expression of the reporter gene *GFP* in transgenic zebrafish heart.
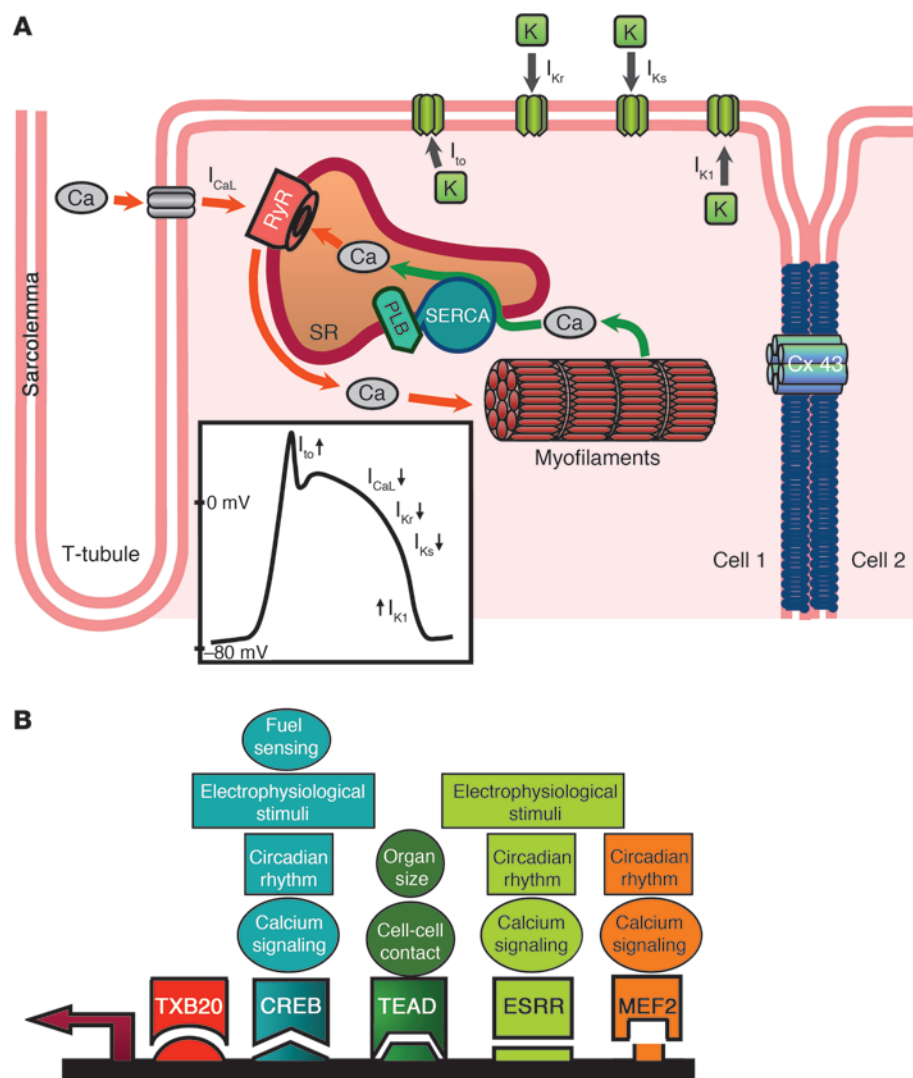
In accordance with the finding that voltage-gated, L-type calcium channels are significantly downregulated in mutant hearts, single-cell patch clamp analysis showed a significant reduction of L-type calcium current in *Tbx20⁻/⁻* cardiomyocytes compared with controls (Figure 4A). Additionally, also in line with the aforementioned experiments, the peak inactivating outward potassium current was significantly reduced (Figure 4A), while the slow component of the outward potassium current was not significantly altered (Figure 4A).

Expression analysis revealed downregulation of various other genes encoding proteins involved in ion flux and regulation of intracellular calcium levels. In keeping with these findings, calcium measurements by line scan confocal microscopy showed significantly reduced calcium transient peaks and significantly reduced cell shortening (Figure 4, B and C) in cardiomyocytes from *Tbx20* conditional KO mice.

ChIP-seq analysis in adult mouse hearts identifies direct downstream targets of Tbx20 with critical functions in cardiomyocytes. To improve our understanding of molecular mechanisms underlying the severe car-

diac phenotypes observed in *Tbx20* conditional KO mice, we sought to determine which cardiac genes with altered expression in mutant mice were directly regulated by Tbx20. Given the data presented previously, we hypothesized that Tbx20 is an upstream transcription regulator of many critical ion transport genes. To address this question, we generated a genome-wide map of Tbx20-binding regions in adult mouse heart and direct Tbx20 downstream gene targets, using ChIP followed by massive parallel sequencing (ChIP-seq).

As none of the commercially available Tbx20 antibodies that we tested were ChIP grade, we decided to use a tagging strategy. We identified a BAC encompassing *Tbx20* and its long-range cardiac transcriptional enhancers (M.A. Nobrega, personal communication). Utilizing BAC recombineering, we fused a GFP cDNA to the C terminus end of *Tbx20* and generated transgenic mice carrying this engineered BAC transgene. Because we used a BAC containing the *Tbx20* cardiac enhancers, the fused *Tbx20*-GFP protein recapitulated the endogenous cardiac expression pattern of *Tbx20*. This was confirmed both by RT-PCR and by immunohistochemistry

Figure 7
Tbx20 as an integrator of adult cardiomyocyte function. (**A**) A diagram detailing proteins involved in ion flux and action potential generation in cardiomyocytes that are encoded by genes that are direct downstream targets of *Tbx20*. These genes are downregulated in *Tbx20* conditional KO adult cardiomyocytes. (**B**) *Tbx20* is part of a cohort of transcription factors that integrate the current cellular state and availability of resources to modulate expression of genes that govern contraction in cardiomyocytes and therefore overall cardiac output.

(Supplemental Figure 3). We also determined, by qRT-PCR, that the fusion *Tbx20-GFP* mRNA is expressed at the same level as endogenous *Tbx20* in 6-week-old hearts, effectively doubling the level of *Tbx20* in the heart. Together, these results illustrate that the *Tbx20-GFP* fusion cassette is expressed in embryonic and adult hearts in a pattern recapitulating the endogenous *Tbx20* expression, and at quasiphysiological levels. We then employed a well-established, ChIP-grade polyclonal anti-GFP antibody for ChIP.

Our ChIP analysis resulted in 4,012 peaks of Tbx20 binding at 2% false discovery rate. Around 23% of these peaks were within 6 kb of the transcription start site (TSS) of RefSeq genes ($P = 10^{-40}$), indicating that although Tbx20 directly binds the promoter regions of several genes, the majority of its regulation relies on binding to longer range DNA elements. Tbx20-binding regions within ±6 kb of the TSS were clustered at the center of this region, close to the TSS, which is the typical binding position for PolII and other transcription factors, supporting the role of Tbx20 as a core transcriptional coregulator. We observed a higher degree of evolutionary conservation of Tbx20-binding regions compared with the rest of the genome, suggesting that the peaks are functional binding sites (Figure 5A).

We then assigned all Tbx20 ChIP–binding regions to neighboring genes to identify putative direct downstream gene targets. Peaks were assigned to a given gene when mapping within the gene or within ±6 kb of the TSS. Peaks in distant intergenic regions were assigned to both flanking genes. As a result, the 4,012 peaks were assigned to 3,799 genes. Directly supporting our hypothesis that KO of *Tbx20* causes the observed cardiac phenotype by regulating expression of ion transport genes, 26 of the 32 genes for which we tested mRNA expression levels had a nearby ChIP peak. Among the 23 genes that were downregulated, 19 had a nearby ChIP peak, indicating that these genes are direct downstream targets of Tbx20.

To interpret the large number of putative Tbx20 downstream targets in functional terms, we performed a computational analysis for enrichment of Gene Ontology (GO) terms (30). The top enriched GO terms were related to heart muscle contraction, cardiac morphogenesis, muscle cell proliferation, myofibril assembly, cell junction assembly, and adult heart development (Figure 5B and Supplemental Data File 3). Interestingly, we observed a significant overlap between our Tbx20-binding regions and those recently reported for various cardiac transcription factors in cardiac HL1 cells (31). Supporting the notion that Tbx20 physically interacts with Gata4,

Nkx2-5, and Tbx5, we observed binding events for Tbx20 and Gata4 to the same genomic coordinates 25-fold more frequently than would be expected by chance. Cobinding of Tbx20 and Nkx2-5 to the same coordinates occurred 14-fold more often than by chance, and that for Tbx20 and Tbx5 also occurred with 14-fold enrichment (Supplemental Table 2; also see Supplemental Figure 4 for analysis of genes for which we measured mRNA expression). Finally, there was an 80-fold enrichment in the cooccurrence of binding regions for Tbx20 and p300, a signature for transcriptional enhancers (32). The overrepresentation of these cooccurrences was strongly statistically significant (Supplemental Table 2). We also observed a significant overrepresentation of binding sites for Gata and Nkx in the Tbx20-binding regions (see Supplemental Figure 4 and Supplemental Table 2), indicating that the high cooccurrence among Tbx20, Gata4, Nkx2.5, and Tbx5 was detectable both at the DNA and the protein-DNA interaction levels. Collectively, these results strongly validate our Tbx20 ChIP-seq data.

Computational analysis reveals a Tbx20-binding motif capable of predicting binding sites not recognized by previously identified motifs. As ChIP peaks putatively correspond to chromatin regions bound by Tbx20, we searched these sequences for previously described Tbx20-binding motifs, that is, the motif identified by selective evolution of ligands by exponential enrichment (33), referred to here as the SELEX Tbx20 motif, and the Eomes protein T-box motif (34). To our surprise, we found only a low number of peaks containing these motifs when compared with random sequences generated with third-, fourth-, and fifth-order Markov models (Eomes: 28% of the peaks, SELEX Tbx20: 4% of the peaks; Supplemental Table 3).

We then employed a de novo motif-finding program that recovered a motif that resembles the known core T-box motif AGGT-GTGA (33) and matches JASPAR's T (MA0009.1, $P = 2 \times 10^{-4}$), TRANSFAC's Tbx5 (M01044, $P = 6 \times 10^{-5}$), and UniProbe's Eomes-primary ($P = 2 \times 10^{-3}$) motifs (Supplemental Figure 5) (34–36), but does not fully match any of these motifs (Figure 5C and Supplemental Figure 6). While clearly displaying most nucleotides of the core T-box motif, our de novo motif contains one position with no nucleotide preference, creating a gap at the center of the SELEX Tbx20 motif (the asterisk in the consensus AGGTG*TGACAG). This de novo motif is present in 51% of the 4,012 Tbx20-binding regions at greater than 80% similarity to the consensus. Altogether, SELEX Tbx20, Eomes, and our de novo Tbx20 motifs could be found in 69% of the peaks (Figure 5D). The Tbx20 de novo motif was the only one found by MEME, it resembles the SELEX Tbx20 and T-box motifs, and it is the most overrepresented according to Clover. Therefore, we used the Tbx20 de novo motif (*Tbx20*+) in our subsequent analyses.

Our analyses of the remaining 31% of Tbx20-binding regions that did not contain an identifiable Tbx20 motif (Tbx20⁻) suggest that they represent functional binding sites of Tbx20. First, these regions display a significant degree of evolutionary sequence conservation and were only slightly less conserved than the binding regions containing recognizable Tbx20 motifs (Tbx20⁺) (Figure 5E). Second, genes associated either with Tbx20⁺ or Tbx20⁻ regions show GO enrichment for heart-related terms (Supplemental Data File 3). Furthermore, motif analyses in Tbx20-binding regions revealed 29 other overrepresented transcription factor–binding site (TFBS) motifs, 18 of which were also overrepresented in Tbx20⁺ regions (3/29 expected by chance). Taken together, these data suggest that at least a fraction of the peaks without a Tbx20 de novo motif are functional. Whether these peaks reflect sites where Tbx20 is recruited to the locus through protein-protein interactions or represent direct DNA-binding events using motifs unrecognized by the various matrices we used is unclear.

Functional testing of identified Tbx20-binding sites. To functionally test both Tbx20⁺- and Tbx20⁻-binding regions for their cardiac enhancer properties, we tested 41 individual binding regions in a zebrafish in vivo reporter assay (37–39). Because data from our murine adult conditional KO model strongly suggested that a number of critical targets of Tbx20 were involved in ion transport, we prioritized sequences that were near cardiac ion transport genes. Of 41 Tbx20-binding regions tested, 23 (56%) consistently drove GFP expression in zebrafish embryonic hearts (Figure 5F and Supplemental Table 4). In these assays, Tbx20⁺- and Tbx20⁻-binding sites were equally likely to behave as cardiac enhancers, further supporting the notion that Tbx20⁻ regions represent functional sequences.

Transcription factor–binding motifs for Tbx20, Mef2a, Tead1, Esrr, and Creb1 are required for coregulation of ion transport genes in adult heart. In addition to *Tbx20* and T-box motifs, we searched the Tbx20-binding region set for enrichment of all TFBS motifs from the JASPAR and UniProbe databases (34, 35) to uncover possible Tbx20 cofactors (Supplemental Data File 4).

Four transcription factor–binding motifs appeared particularly enriched in binding regions associated with genes encoding for ion channel and transport proteins:

JASPAR's (35) Mef2a (MA0052.1), Creb1 (MA0018.2), and Tead1 (MA0090.1) and UniProbe's (34) Esrr-primary (40) were overrepresented in peaks associated with both Tbx20⁺ and Tbx20⁻ (Supplemental Data File 4). Genes in which Mef2A, Tead1, Creb1, and Esrr motifs cooccur with Tbx20-binding regions were enriched for functional GO terms connected to ion transport/contraction and heart development/morphogenesis-related terms (Figure 6A), with more than 45% of enriched GO terms being ion transport/contraction (see Supplemental Table 5 and Methods for a detailed description of the analysis). This enrichment for the cooccurrence of the 5 transcription factor–binding motifs (those for Tbx20, Mef2A, Creb2, Tead1, Essr) was statistically significant in that only 0.4% of 10,000 random groups of any 5 TFBSs occurring in similar frequencies in the test motifs yielded comparable results.

To test the functionality of TFBSs for Tbx20, Mef2A, Creb2, Tead1, and Essr in Tbx20 peaks containing binding sites for each of these factors, we selected 2 representative Tbx20 target enhancers, one adjacent to *Cacna1c* and the other adjacent to *Irx4*, containing all 5 motifs and mutagenized individual instances of each motif in these enhancers, evaluating the impact of each mutation in the enhancer activity in our zebrafish in vivo reporter. Mutation of each individual binding site led to a reduction in enhancer activity, reflected in the decrease of the number of zebrafish expressing GFP in the heart by 50%–100% (Figure 6, B and C, and Supplemental Table 6).

Taken together, these results suggest that Tbx20, in concert with a network that includes Mef2, Tead, Creb, and Esrr, regulates calcium homeostasis and ion flux in adult heart.

Discussion

Genes critical to cardiac structure and function are direct targets of Tbx20 in adult heart. Previous studies have demonstrated a requirement for Tbx20 in heart development (5–8), and human studies have demonstrated that *TBX20* is a cardiac disease gene (1–3). In this study, we have shown that conditional ablation of *Tbx20* in adult cardiomyocytes causes a cardiomyopathy of unusual severity, with death occurring within 2 weeks. Mutant mice exhibited dilated

hearts, with a rapid loss of systolic function accompanied by a conduction delay, leading to bradyarrhythmias and tachyarrhythmias. The observation of arrhythmias in *Tbx20*-KO mice is of particular interest given recent genome-wide association studies demonstrating association between common variants in the *Tbx20* locus and susceptibility to arrhythmias in humans (4).

The severity of the phenotype of *Tbx20* conditional KO in adult cardiomyocytes demonstrates a pivotal role for Tbx20 in the hierarchy of factors regulating cardiomyocyte function, maintenance, and survival. This is the first study, to our knowledge, to establish an ongoing role for Tbx20 in adult heart function and to define direct targets of Tbx20 in adult heart, giving insight into mechanisms by which mutations in *TBX20* may result in cardiomyopathy in humans.

The expression of a number of genes critical to cardiac function was substantially downregulated in *Tbx20* mutant cardiomyocytes, and many of these genes are directly regulated by Tbx20, as demonstrated by our ChIP and in vivo enhancer analyses. Among direct downstream targets of Tbx20 are several transcription factors, potassium channel components, proteins involved in calcium cycling, connexins, myofibrillar proteins, and cytoskeletal proteins that have previously been shown to play important roles in cardiac development and/or adult heart function (41–53).

At least some pathways regulated by Tbx20 in adult heart appear to be conserved between vertebrates and invertebrates. In *Drosophila*, the Tbx20 ortholog neuromancer is required in adult hearts to maintain myofibrillar architecture and physiological function. In neuromancer mutants, *Serca* and the Eag-like potassium channel Elk (homologous to Kcnh1) were downregulated (48).

While the severe cardiac phenotype in *Tbx20* conditional KO mice suggests that a number of biochemical and signaling pathways are altered in these animals, we established that at least part of these phenotypes arise from a direct role of Tbx20 in controlling broad aspects of intracellular ion flux and homeostasis. Together, these genes encode for proteins involved in excitation/contraction coupling and participate in virtually every major phase of the propagation of an action potential in cardiomyocytes (Figure 7A).

Several of Tbx20 direct downstream targets have been linked to human channelopathies (54, 55), dilated cardiomyopathy (reviewed in ref. 56), or congenital heart disease (reviewed in ref. 57). Interestingly, channelopathies are often associated with an increased susceptibility to arrhythmias, while hearts of affected individuals appear structurally normal (58). With the exception of some diseases, dilated cardiomyopathies and acquired causes of congestive heart failure typically present primarily as decreases in contractile function, with the incidence of arrhythmias increasing over time after systolic dysfunction has developed (59). In contrast, *Tbx20* conditional KO mice develop both aspects, with loss of contractility and susceptibility to arrhythmias within a very short amount of time, a situation that is typical of end-stage adult heart disease. Preliminary data from our labs show that downregulation of *Tbx20* may be a regular finding in heart failure with different etiology (Supplemental Figure 2). Further research is required to address whether dysregulation of *Tbx20* may be a common feature of end-stage heart failure and therefore a potential therapeutic target.

Taken together, the KO phenotypes of these genes with decreased expression in adult cardiac *Tbx20* mutants are consistent with the phenotypes observed in these animals and are likely to have contributed to them. Nevertheless, considering the number of Tbx20 downstream targets we have discovered in adult mouse heart, it seems probable that other genes and proteins not discussed here are also involved in the development of the phenotype.

A Tbx20-binding motif. Our analyses of genome-wide binding regions of Tbx20 in adult mouse hearts revealed that a paucity of regions contained previously described Tbx20-binding motifs (33, 34). Instead, we discovered a Tbx20-binding motif within Tbx20 ChIP regions that has some similarity but is distinct from known motifs. This motif was present in a majority, but not all, binding regions, suggesting that there are a number of sequence variations that may have distinct affinities for Tbx20.

We hypothesize that Tbx20 functions may in part be regulated by the presence of such motifs, with variations in binding affinity. It is believed that T-box proteins likely bind to the same motifs and are often in competition for the same binding regions, as evidenced by the large overlap in ChIP-binding regions between Tbx20 in our study and Tbx5 (60). A canonical T-box site with high affinity to Tbx20 (33) may not allow for the rapid turnover of binding in response to small fluctuations in Tbx20 or other T-box family proteins in the nucleus. Instead, intermediate to low binding affinities may allow for fine-tuning of transcriptional regulation, particularly with different T-box proteins potentially competing or interacting at the same sites. The identification of *TBX20* mutations with both increased and decreased transcriptional activity in patients with heart disease supports this idea. Generalized low-affinity binding could provide one molecular explanation for the notorious dosage sensitivity of T-box proteins. Another possibility is that binding of Tbx20 to distinct motifs results in specific allosteric conformations of Tbx20, facilitating interaction with distinct cofactors. Further research addressing the DNA binding of Tbx20 and correlating the genome-wide distribution of the various motifs with impact on gene expression is required for an understanding of the details of its transcriptional regulation.

Tbx20 is part of a transcription factor cohort that integrates environmental information to regulate ion transport in adult heart. Our results suggest that Tbx20, Mef2A, Tead1, Esrr, and Creb1 are part of a transcription factor cohort that regulates important aspects of adult cardiac function, including ion flux in cardiomyocytes. A recent analysis of DNA-binding motifs within cardiomyocyte enhancers also identified an overrepresentation of the Tead1-binding site (60). Each of these transcription factors, working in concert with Tbx20, has been previously recognized as having an "integrator" function, regulating transcription in response to specific types of intracellular or external stimuli (40, 61–71). From this, the picture emerges of Tbx20 as a pivotal component of this integrator cohort (Figure 7B).

The finding that binding by Tbx20 is frequently associated with motifs for Mef2a, Tead1, Esrr, and Creb1, which are each required for enhancer activity in our experimentally tested constructs, suggests that important stimuli, such as intracellular calcium concentration, mechanical stress, organ size and cell-cell contacts, energy metabolism and cAMP level, and time of day are taken into account in the transcriptional regulation of ion transport proteins and other gene targets of this transcription factor cohort in adult cardiomyocytes (Figure 7B). These 5 factors may therefore function as parts of an integrator cohort that fine-tunes expression of continuously required proteins in response to the current cellular state, availability of resources, and the amount of contractile work required by the cardiomyocyte. This mechanism may ensure the continuous and economic maintenance of a cell type that is constantly subjected to mechanical stress and that converts chemical

energy into mechanical energy with mechanisms that critically rely on exact ion concentrations.

Little is known about factors regulating *Tbx20* expression in this context. Tfap2 transcription factors have been identified as repressors of *Tbx20* expression (72). A recent study showed that *Tbx20* is a direct transcriptional target of the Bmp/Smad pathway during development (73). Further research will be required to understand the biologic stimuli that drive *Tbx20* expression in adult heart.

In summary, the data presented here posit Tbx20 as a central and required transcriptional regulator in adult cardiomyocytes and suggest that it may act as part of a transcriptional cohort that integrates environmental, intracellular, and temporal information to regulate the expression of ion transport proteins in adult cardiomyocytes.

Methods

Generation of inducible cardiomyocyte-specific Tbx20-KO mice

The generation of *Tbx20* floxed (f/+ or f/f) mice and α-MHC–mER–Cre–mER mice has been reported previously (5, 74). Inducible cardiac-specific *Tbx20*-KO mice (α-MHC–mER–Cre–mER/*Tbx20^{f/f}*) were generated by breeding *Tbx20^{f/f}* mice with mice expressing Cre-recombinase fused to the mutated estrogen receptor mER-Cre-mER under the control of the α-MHC promoter (provided by J.D. Molkentin, University of Cincinnati, Cincinnati, Ohio, USA). Genotyping was performed by RT-PCR using primer sets described previously (5, 74). Six- to eight-week-old α-MHC–mER–Cre–mER/*Tbx20^{f/f}* male mice were treated with tamoxifen (Sigma-Aldrich) dissolved in sesame oil by intraperitoneal injection once a day for 5 consecutive days at a dosage of 30 mg/kg/d and used as inducible cardiac-specific *Tbx20*-KO mice. Control male mice carrying the genotypes α-*Tbx20^{f/f}* or MHC-mER-Cre-mER/*Tbx20^{+/+}* were treated with the same tamoxifen regimen. An additional control group was obtained by intraperitoneal injection of sesame oil in α-MHC–mER–Cre–mER/*Tbx20^{f/f}* mice. All experiments were performed in a mixed 129/black Swiss outbred background throughout the study. Cre excision efficiency in adult α-MHC–mER–Cre–mER/*Tbx20^{f/f}* mice was performed by X-gal staining after performing the experiment in mice carrying the Rosa26-LacZ reporter transgene (75) as well as by analysis of *Tbx20* mRNA and protein expression by Western blot and qRT-PCR.

Rat isoproterenol heart failure model

Male Wistar rats (10 to 12 weeks old) weighing 200–240 g were randomly divided into a control group and isoproterenol (ISO) treatment group. Animals received subcutaneous injections of ISO at a dose of 300 mg/kg/d ISO (Sigma-Aldrich) or saline for 2 consecutive days. The hearts were harvested at 4 weeks after the first injection (n = 4–8).

Mouse myocardial infarction model

Mice were anesthetized using 1%–1.5% isoflurane in oxygen and maintained at body temperature. The left-side thoracotomy was performed, and pericardium was partially stripped to expose the heart. The left anterior descending coronary artery was ligated with an 8-0 silk suture. Myocardial ischemia was confirmed by blanching of the anterior wall of the LV and ST segment elevation on the ECG. In sham-operated mice, the same surgical procedures were performed except for the ligation of the coronary artery. After the chest was closed with a 5-0 polyester suture, animals were kept warm under a heat lamp during the recovery period. The hearts were harvested 8 weeks after the operation. The LV and septum of sham-operated and myocardial infarction mice were dissected. For heart failure analyses, myocardium from noninfarcted regions was analyzed.

Surface ECG and radio telemetry recordings

Surface ECG recordings were performed on mice anesthetized with 1%–1.5% isoflurane or on isolated heart. Needle electrodes were used and placed in the conventional lead II position. A differential amplifier (DP-304; Warner Instruments Corp.) amplified the signals in the bandwidth of 0.1–1000 Hz, and signals were filtered using an adaptive 60-Hz filter (Humbug; Quest Scientific). Signals were digitized at 3000 Hz and analyzed using ECG Analysis Software (QRS Phenotyping).

To obtain ECG data from conscious mice over 24-hour periods, radio-frequency transmitters (TA 10ETA-F20; Data Science Inc.) were implanted into adult mice. Mice were kept in a facility at a 12-hour light/12-hour dark cycle (7 am to 7 pm). Data obtained from telemetry recordings were digitized at 1000 Hz and analyzed using Dataquest A.R.T. 4.0 (Data Science Inc.).

Single cell electrophysiology

Single ventricular myocytes were isolated from the hearts of 2-month-old *Tbx20*-KO and WT mice using an enzymatic method. The membrane currents were recorded under the whole-cell patch-clamp configuration with an Axon 200B amplifier. The pipette-filling solution for ICa recording contained the following: 120 mmol/l CsCl, 5 mmol/l MgATP, 10 mmol/l HEPES, 10 mmol/l TEA-Cl, 3 mmol/l NaCl, and 2 mmol/l EGTA (pH adjusted to 7.2 with CsOH). Bath solution for ICa recording contained the following: 137 mmol/l NaCl, 7 mmol/l CsCl, 1.2 mmol/l CaCl$_2$, 1 mmol/l MgCl$_2$, 1 mmol/l NaH2PO4, 10 mmol/l HEPES, 10 mmol/l D-glucose, and 0.02 mmol/l TTX (pH 7.4). In order to selectively examine L-type calcium channels in whole-cell current records, cells were voltage clamped at –45 mV to inactivate the T-type calcium current; then calcium currents were elicited by 500-ms depolarizations to different voltages (from –60 mV to 60 mV). I$_{Ca}$ was measured as the difference between peak inward current and the current at the end of the 500-ms pulse.

The pipette-filling solution for K$^+$ current recording contained the following: 120 mmol/l potassium aspartate, 20 mmol/l KCl, 5 mmol/l NaCl, 1 mmol/l MgCl$_2$, 5 mmol/l MgATP, 10 mmol/l HEPES, and 10 mmol/l EGTA (pH 7.2). Cell capacitance and series resistance were electronically compensated by 80%–90%. The bath solution for K$^+$ current recording contained the following: 140 mmol/l NaCl, 4 mmol/l KCl, 1 mmol/l MgCl$_2$, 1.2 mmol/l CaCl$_2$, 10 mmol/l HEPES, 10 mmol/l D-glucose, 0.02 mmol/l TTX, and 0.3 mmol/l CdCl$_2$ (pH 7.4). Cells were held at –80 mV, and K$^+$ currents were elicited by 4.5-second depolarizations to different test voltages (from –40 mV to 40 mV). The noninactivating component (I$_{ss}$) was measured as the sustained current remaining at the end of 4.5-second pulses. The I$_{k,slow}$ was measured at the point of 500 ms after the beginning of depolarizations. The I$_{k,peak}$ was measured as the difference between the maximal outward current and the I$_{k,slow}$.

Current amplitudes, measured in individual cells, were normalized to cell size (whole-cell membrane capacitance), and current densities (in pA/pF) were reported.

Confocal Ca$_2$$^+$ imaging

Myocytes loaded with Fluo-5 AM (10 μmol/l, 10 minutes) were measured using an Olympus FV-1000 confocal microscope with a ×40 oil immersion lens (NA 1.3). 488-nm laser excitation and 510- to 530-nm emission were used for detecting Fluo-5 fluorescent signal. All experiments were performed at room temperature (22–24°C). Off-line image processing used IDL software (version 6.4, Research Systems Co.).

Echocardiography

Mice were anesthetized using 1%–1.5% isoflurane in oxygen and maintained at body temperature. Echocardiography was performed by investigators blinded for the genotype of the animals using a Vevo 770, Vevo 2100 (VisualSonics), or HP Sonos 5500 (Philips).

Histological analysis

Hearts were flushed with PBS, relaxed in 50 mM KCl in PBS, and fixed in 4% paraformaldehyde (Sigma-Aldrich). Tissue was processed for paraffin-embedded sections or cryosections and subsequently analyzed by H&E staining according to the manufacturer's protocol (Sigma-Aldrich).

X-gal staining

Hearts presumably expressing β-gal were incubated in β-gal substrate (Roche Molecular) until sufficient staining was visible. Alternatively, hearts were fixed and sectioned as described above and sections were incubated in β-gal substrate until sufficient staining was visible.

Immunohistochemistry

8- to 10-μm sections were incubated with primary antibodies overnight at 4°C. The following primary antibodies were used: Connexin 43 (Abcam), Cav1.2a (Alomone), and Kv4.2 (Upstate). After washing with 0.25% Triton X-100 in PBS, sections were incubated with either fluorescently labeled (Molecular Probes; Invitrogen) or biotinylated secondary (Vector) antibodies for 2 hours. Sections were imaged by fluorescence microscopy.

Real-time RT-PCR

Quantitative real-time RT-PCR was performed using the iCycler IQ system (Bio-Rad) in combination with iQ SYBR Green Real-Time RT-PCR Supermix (Bio-Rad) according to the manufacturer's protocol. Briefly, total RNA was extracted using TriReagent (Sigma-Aldrich). 1–3 μg of RNA was then reverse transcribed to first-strand cDNA using random primers and Moloney murine leukemia virus reverse transcriptase (Promega) following the manufacturer's protocol. The primers used were as follows: 18S RNA, 5′-GGAAGGGCACCACCAGGAGT-3′ (forward) and 5′-TGCAGCCCCGGACATCTAAG-3′ (reverse). The RT-PCR profile was as follows: 95°C for 30 seconds and 40 cycles at 95°C for 15 seconds, 58°C for 30 seconds, and 72°C for 30 seconds. The amount of SYBR Green was measured at the end of each cycle. The cycle number at which the emission intensity of the sample rose above the baseline is referred to as the threshold cycle and was proportional to the target concentration. The data presented are the average of 4 independent experiments. For primers, see Supplemental Table 7.

TUNEL

Nuclear fragmentation was detected by TUNEL staining with an apoptosis detection kit (R&D Systems) according to the manufacturer's protocol.

RNA in situ hybridization

Hearts were dissected and fixed overnight at 4°C in 4% paraformaldehyde in PBS, then put in 5%, 10%, and 20% sucrose each for 8 hours and embedded in 20% sucrose and 1:1 OCT. 15-micron sections were cut and mounted onto Superfrost Plus slides, dried for 15 minutes at room temperature, and stored at –80°C. Before use, tissue sections on slides were dried again at room temperature for 1 hour and then processed for in situ hybridization as described.

Western blot analysis

Western blot analyses were performed on heart lysates from *Tbx20* conditional KO mice and control mice using standard procedures. Antibodies used in this study were *Tbx20* (Santa Cruz Biotechnology Inc.), Connexin 43 (Abcam), CamKII delta (Abcam), Cav1.2a (Alomone), PLN (Abcam), Kir2.1 (UC Davis/NIH NeuroMab Facility), GIRK1 (Chemicon), Kv4.2 (Upstate), and Cypher (generated in our laboratory). GAPDH antibody (Sigma-Aldrich) was used for normalization. ImageJ software (NIH) was used to perform densitometric analyses (http://rsb.info.nih.gov/ij/).

Generation of Tbx20-GFP-BAC transgenic mice

We used a modified BAC to generate a transgenic line that stably expresses *Tbx20* tagged with eGFP fused at the C terminus. The eGFP/ampicillin cassette was amplified by using primers containing sequence-specific 50-bp overhang and inserted to BAC RP23-150G12 (Invitrogen) by RED/ET homologous recombination in *E. coli*. Modified BAC DNA was extracted using Nucleobond AX Kit (Macherey-Nagel) and used for pronuclear injections of CD1 embryos in accordance with standard protocols approved by the University of Chicago.

ChIP-seq

Six-week-old *Tbx20*-GFP mouse hearts were dissected and cross-linked in 1.5% formaldehyde for 10 minutes at room temperature. Samples were quenched by incubation with 2.5 M glycine for 5 minutes. Cells were washed and resuspended in Lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris, pH 8.1). The cross-linked chromatin was sonicated using a Misonix 4000 (amplitude 90 and 30 seconds on/30 seconds off) to an average size of 200 bp. Precleared chromatin extract was incubated overnight at 4°C with goat anti-GFP and immunoprecipitated with protein A–Sepharose beads. 10 ng of DNA was used to generate a standard Illumina sequencing library. Enrichment of selected target regions was confirmed by qRT-PCR. ChIP-seq data can be found at NCBI GEO (76) (GEO GSE29636).

Zebrafish in vivo reporter system

Each *Tbx20* ChIP peak sequence was RT-PCR amplified from the mouse genome and cloned in an eGFP reporter cassette with a minimal c-fos promoter upstream (77, 78). The reporter cassette contains 2 tol2 transposon sites that allow efficient integration of the test sequence (79). Each construct was injected in 100–200 1-cell–stage zebrafish embryos in the presence of tol2 transposase. Since patterns observed in mosaic G_0 fish are reproduced in the germline transmission to G_1 (77, 78), we evaluated enhancer properties in G_0 fish embryos. We required GFP expression in the hearts of at least 20% of the fish to call a sequence positive.

Mutagenesis of selected TFBS

Mutagenic primers were designed by using PrimerX (http://www.bioinformatics.org/primerx). TBX20 peaks were amplified from mouse genomic DNA by RT-PCR and cloned into an entry vector (pENTR/D-TOPO). Mutagenesis was performed with the GeneTailor Site-Directed Mutagenesis System (Invitrogen) according to the manufacturer's instructions. All constructs were verified by DNA sequencing. For mutagenesis primers, see Supplemental Table 8.

Bioinformatics analyses of ChIP data

ChIP peak calling. Illumina reads were aligned to the mouse genome (mm9) using Eland, and only reads uniquely mapped to the genome were used. Peaks were called against input chromatin using QuEST 2.3 (80) with ChIP-to-background cutoffs of 30 and 40 reads.

Gene database. RefSeq mm9 coordinates downloaded from UCSC (81) in February 2010 were clustered by gene symbol, resulting in 21,606 mRNA clusters/genes. Peaks were mapped to all available TSS.

De novo motif finding and motif similarity. All the 1.6-k and 4-k peaks had low complexity sequences and repeats (82) masked and given as input to MEME (83) using the ZOOPS model. Both sets yielded virtually the same result, but the analyses reported were performed with the motif identified in the 1.6-k set.

STAMP (84) (Smith-Waterman alignment, all other parameters set to default) was used to find the top 2 matches of the *Tbx20* de novo motif identified by MEME to known motifs in JASPAR, TRANSFAC, and Uni-Probe databases. Sequence logos were generated with the Bioconductor

package seqLogo by Oliver Bembom (University of California, Berkeley, Berkeley, California, USA).

Motif overrepresentation and mapping. Clover (85) was run with default parameters (P value cutoff = 0.05) on both peak sets against 4 control sequence sets: a mouse promoter set and chromosome 21 sequences provided by Clover and sequences generated from a third order and a fourth Markov model derived from each peak set.

MotifLocator from the Toucan package (86) with a 0.8 similarity cutoff and a third order Markov model background was employed to identify putative TFBS motifs from the JASPAR (35) and UniProbe (34) collections plus the SELEX *Tbx20* motif (33) and the de novo *Tbx20* motif identified by MEME. UniProbe matrices were end-trimmed if information content was less than 0.25 bits for all analyses.

Motif cooccurrence. Putative TFBS mapped with MotifLocator was counted in all peaks. As many TFBS have very similar motifs, resulting in artificial cooccurrence, TFBS occurrences overlapping by 50% of the length of the shortest motif were not considered. Corrected counts were then used to estimate the expected frequency of cooccurrence using Bayes' rule. Removed counts were later added or not, and the results were very similar. A χ^2 goodness-of-fit test was employed to calculate the P value for excess of peaks containing cooccurring TFBS. Multiple test correction with the Benjamini-Hochberg method implemented in the R package (87) was applied. When estimating the number of random sets of 4 motifs that would yield a similar number of significant combinations, we only employed JASPAR matrices that occurred in frequencies similar to those in the test set (24 matrices).

GO analysis. MGI associations of GO terms (30) to genes and the GO database were downloaded on August 1, 2010. GO terms were associated with RefSeq genes and both flanking intergenic regions, and using these associations, a "genome fraction" for each term was calculated (gene + intergenic regions length/total genome length). Peaks were assigned to a given GO term if they were located in the gene or intergenic regions annotated with the term. Significance of GO enrichment was assessed with the binomial test using the "genome fraction" as probability of success to correct for locus length bias (88). Concomitantly, a hypergeometric test was conducted on genes to which peaks were assigned. Only terms with $P < 0.05$ and fold-enrichment greater or lower than 2 in both sets were accepted as significant. The Benjamini-Hochberg method in the R package was used to correct for multiple tests. For the analysis of GO enrichment of genes associated with different motif combinations, only the hypergeometric test was used.

Conservation. The 30-way placental phastCons scores (89) were downloaded from the UCSC Genome Browser (81) and averaged per nucleotide for a given set of ChIP peaks. AUCs of phastCons scores were calculated by summing the average scores around ChIP peak summits. To estimate the P values in obtaining a similar AUC for a given peak set, 1,000 random sets of the same number of peaks being tested were generated and the AUC was calculated for each set. The number of sets with AUC greater than or equal to the test AUC was counted and divided by 1,000, resulting in the P value.

P value plots were calculated with t tests and corrected for multiple tests using the Benjamini-Hochberg method (R package).

Generation of random peak sets. Random sets containing the same number of *Tbx20* peaks of interest were generated by sampling size-matched sequences from the uniquely mappable regions of the genome. These regions were extracted from the mm9 Mappability track of the UCSC Genome Browser (crgMapabilityAlign36mer.bw).

Overlap of Tbx20 ChIP peaks with HL1 peaks. To assess enrichment of the fraction of *Tbx20* ChIP peaks overlapping HL1 peaks (31) above expectation for a random set of genomic coordinates, we generated 100 sets of 4,012 genomic intervals matching the size distribution of the *Tbx20* peaks, randomly sampled from the mappable regions of the genome. Overlap of these peak sets with HL1 peaks (at least 1 bp) was averaged and compared with the *Tbx20* ChIP set.

Statistics

Data are expressed as mean ± SEM. Statistical comparisons used 1-way ANOVA, followed by Bonferroni's procedure for multiple-group comparisons. $P < 0.05$ was considered statistically significant.

Animal use

All animal procedures were carried out in accordance with the guidelines set by the University of California San Diego and University of Chicago Institutional Animal Care Program and with IACUC approval.

Address correspondence to: Marcelo A. Nobrega, Department of Human Genetics, University of Chicago, 920 E. 58th Street CLSC 319, Chicago, Illinois 60637, USA. Phone: 773.834.7919; Fax: 773.834.8470; E-mail: nobrega@uchicago.edu. Or to: Sylvia M. Evans, Skaggs School of Pharmacy, UCSD, 9500 Gillman Road, La Jolla, California 92093, USA. Phone: 858.822.2452; Fax: 858.822.3027; E-mail: syevans@ucsd.edu.

Chen-leng Cai's present address is: Department of Developmental and Regenerative Biology, The Mount Sinai Medical Center, New York, New York, USA.

1. Kirk EP, et al. Mutations in cardiac T-box factor gene TBX20 are associated with diverse cardiac pathologies, including defects of septation and valvulogenesis and cardiomyopathy. *Am J Hum Genet.* 2007;81(2):280–291.
2. Liu C, Shen A, Li X, Jiao W, Zhang X, Li Z. T-box transcription factor TBX20 mutations in Chinese patients with congenital heart disease. *Eur J Med Genet.* 2008;51(6):580–587.
3. Posch MG, et al. A gain-of-function TBX20 mutation causes congenital atrial septal defects, patent foramen ovale and cardiac valve defects. *J Med Genet.* 2010;47(4):230–235.
4. Sotoodehnia N, et al. Common variants in 22 loci are associated with QRS duration and cardiac ventricular conduction. *Nat Genet.* 2010;42(12):1068–1076.
5. Cai CL, et al. T-box genes coordinate regional rates of proliferation and regional specification during cardiogenesis. *Development.* 2005;132(10):2475–2487.
6. Stennard FA, et al. Murine T-box transcription factor Tbx20 acts as a repressor during heart development, and is essential for adult heart integrity, function and adaptation. *Development.* 2005;132(10):2451–2462.
7. Takeuchi JK, et al. Tbx20 dose-dependently regulates transcription factor networks required for mouse heart and motoneuron development. *Development.* 2005;132(10):2463–2474.
8. Singh MK, et al. Tbx20 is essential for cardiac chamber differentiation and repression of Tbx2. *Development.* 2005;132(12):2697–2707.
9. Stennard FA, et al. Cardiac T-box factor Tbx20 directly interacts with Nkx2-5, GATA4, and GATA5 in regulation of gene expression in the developing heart. *Dev Biol.* 2003;262(2):206–224.
10. Brown DD, et al. Tbx5 and Tbx20 act synergistically to control vertebrate heart morphogenesis. *Development.* 2005;132(3):553–563.
11. Benson DW, et al. Mutations in the cardiac transcription factor NKX2.5 affect diverse cardiac developmental pathways. *J Clin Invest.* 1999; 104(11):1567–1573.
12. Dentice M, et al. Missense mutation in the tran-

scription factor NKX2-5: a novel molecular event in the pathogenesis of thyroid dysgenesis. *J Clin Endocrinol Metab*. 2006;91(4):1428–1433.

13. Garg V, et al. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature*. 2003;424(6947):443–447.

14. Goldmuntz E, Geiger E, Benson DW. NKX2.5 mutations in patients with tetralogy of fallot. *Circulation*. 2001;104(21):2565–2568.

15. Gutierrez-Roelens I, et al. A novel CSX/NKX2-5 mutation causes autosomal-dominant AV block: are atrial fibrillation and syncopes part of the phenotype? *Eur J Hum Genet*. 2006;14(12):1313–1316.

16. Hirayama-Yamada K, et al. Phenotypes with GATA4 or NKX2.5 mutations in familial atrial septal defect. *Am J Med Genet A*. 2005;135(1):47–52.

17. Inga A, Reamon-Buettner SM, Borlak J, Resnick MA. Functional dissection of sequence-specific NKX2-5 DNA binding domain mutations associated with human heart septation defects using a yeast-based system. *Hum Mol Genet*. 2005;14(14):1965–1975.

18. Rauch R, et al. Comprehensive genotype-phenotype analysis in 230 patients with tetralogy of Fallot. *J Med Genet*. 2010;47(5):321–331.

19. Reamon-Buettner SM, Borlak J. Somatic NKX2-5 mutations as a novel mechanism of disease in complex congenital heart disease. *J Med Genet*. 2004; 41(9):684–690.

20. Schott JJ, et al. Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science*. 1998;281(5373):108–111.

21. Pfeufer A, et al. Genome-wide association study of PR interval. *Nat Genet*. 2010;42(2):153–159.

22. Li QY, et al. Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. *Nat Genet*. 1997;15(1):21–29.

23. Basson CT, et al. Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome. *Nat Genet*. 1997;15(1):30–35.

24. Pashmforoush M, et al. Nkx2-5 pathways and congenital heart disease; loss of ventricular myocyte lineage specification leads to progressive cardiomyopathy and complete heart block. *Cell*. 2004; 117(3):373–386.

25. Oka T, et al. Cardiac-specific deletion of Gata4 reveals its requirement for hypertrophy, compensation, and myocyte viability. *Circ Res*. 2006;98(7):837–845.

26. Zeisberg EM, et al. Morphogenesis of the right ventricle requires myocardial expression of Gata4. *J Clin Invest*. 2005;115(6):1522–1531.

27. Bisping E, et al. Gata4 is required for maintenance of postnatal cardiac function and protection from pressure overload-induced heart failure. *Proc Natl Acad Sci U S A*. 2006;103(39):14471–14476.

28. Bruneau BG, et al. A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. *Cell*. 2001;106(6):709–721.

29. Mori AD, et al. Tbx5-dependent rheostatic control of cardiac gene expression and morphogenesis. *Dev Biol*. 2006;297(2):566–586.

30. Ashburner M, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25(1):25–29.

31. He A, Kong SW, Ma Q, Pu WT. Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. *Proc Natl Acad Sci U S A*. 2011;108(14):5632–5637.

32. Visel A, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*. 2009; 457(7231):854–858.

33. Macindoe I, et al. Conformational stability and DNA binding specificity of the cardiac T-box transcription factor Tbx20. *J Mol Biol*. 2009;389(3):606–618.

34. Newburger DE, Bulyk ML. UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. *Nucleic Acids Res*. 2009; 37(Database issue):D77–D82.

35. Bryne JC, et al. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res*. 2008;36(Database issue):D102–D106.

36. Matys V, et al. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res*. 2003;31(1):374–378.

37. Gotea V, Visel A, Westlund JM, Nobrega MA, Pennacchio LA, Ovcharenko I. Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers. *Genome Res*. 2010;20(5):565–577.

38. Narlikar L, et al. Genome-wide discovery of human heart enhancers. *Genome Res*. 2010;20(3):381–392.

39. Taher L, et al. Genome-wide identification of conserved regulatory function in diverged sequences. *Genome Res*. 2011;21(7):1139–1149.

40. Dufour CR, et al. Genome-wide orchestration of cardiac functions by the orphan nuclear receptors ERRalpha and gamma. *Cell Metab*. 2007; 5(5):345–356.

41. Barry DM, Xu H, Schuessler RB, Nerbonne JM. Functional knockout of the transient outward current, long-QT syndrome, and cardiac remodeling in mice expressing a dominant-negative Kv4 alpha subunit. *Circ Res*. 1998;83(5):560–567.

42. Bettahi I, Marker CL, Roman MI, Wickman K. Contribution of the Kir3.1 subunit to the muscarinic-gated atrial potassium channel IKAch. *J Biol Chem*. 2002;277(50):48282–48288.

43. Guo W, et al. Targeted deletion of Kv4.2 eliminates I(to,f) and results in electrical and molecular remodeling, with no evidence of ventricular hypertrophy or myocardial dysfunction. *Circ Res*. 2005; 97(12):1342–1350.

44. Gutstein DE, et al. Conduction slowing and sudden arrhythmic death in mice with cardiac-restricted inactivation of connexin43. *Circ Res*. 2001; 88(3):333–339.

45. Gutstein DE, et al. Heterogeneous expression of Gap junction channels in the heart leads to conduction defects and ventricular dysfunction. *Circulation*. 2001;104(10):1194–1199.

46. Kuo HC, et al. A defect in the Kv channel-interacting protein 2 (KChIP2) gene leads to a complete loss of I(to) and confers susceptibility to ventricular tachycardia. *Cell*. 2001;107(6):801–813.

47. Periasamy M, et al. Impaired cardiac performance in heterozygous mice with a null mutation in the sarco(endo)plasmic reticulum Ca2+-ATPase isoform 2 (SERCA2) gene. *J Biol Chem*. 1999; 274(4):2556–2562.

48. Qian L, et al. Transcription factor neuromancer/TBX20 is required for cardiac function in Drosophila with implications for human heart disease. *Proc Natl Acad Sci U S A*. 2008;105(50):19833–19838.

49. Seisenberger C, et al. Functional embryonic cardiomyocytes after disruption of the L-type alpha1C (Cav1.2) calcium channel gene in the mouse. *J Biol Chem*. 2000;275(50):39193–39199.

50. Takeshima H, Komazaki S, Hirose K, Nishi M, Noda T, Iino M. Embryonic lethality and abnormal cardiac myocytes in mice lacking ryanodine receptor type 2. *EMBO J*. 1998;17(12):3309–3316.

51. Zaritsky JJ, Redell JB, Tempel BL, Schwarz TL. The consequences of disrupting cardiac inwardly rectifying K(+) current (I(K1)) as revealed by the targeted deletion of the murine Kir2.1 and Kir2.2 genes. *J Physiol*. 2001;533(pt 3):697–710.

52. Zheng M, et al. Cardiac-specific ablation of Cypher leads to a severe form of dilated cardiomyopathy with premature death. *Hum Mol Genet*. 2009; 18(4):701–713.

53. Zhou Q, et al. Ablation of Cypher, a PDZ-LIM domain Z-line protein, causes a severe form of congenital myopathy. *J Cell Biol*. 2001;155(4):605–612.

54. Priori SG, Napolitano C. Role of genetic analyses in cardiology: part I: mendelian diseases: cardiac chan-

nelopathies. *Circulation*. 2006;113(8):1130–1135.

55. Lehnart SE, et al. Inherited arrhythmias: a National Heart, Lung, and Blood Institute and Office of Rare Diseases workshop consensus report about the diagnosis, phenotyping, molecular mechanisms, and therapeutic approaches for primary cardiomyopathies of gene mutations affecting ion channel function. *Circulation*. 2007;116(20):2325–2345.

56. Watkins H, Ashrafian H, Redwood C. Inherited cardiomyopathies. *N Engl J Med*. 2011; 364(17):1643–1656.

57. Richards AA, Garg V. Genetics of congenital heart disease. *Curr Cardiol Rev*. 2010;6(2):91–97.

58. Roberts JD, Gollob MH. The genetic and clinical features of cardiac channelopathies. *Future Cardiol*. 2010;6(4):491–506.

59. Lakdawala NK, Givertz MM. Dilated cardiomyopathy with conduction disease and arrhythmia. *Circulation*. 2010;122(5):527–534.

60. He A, Kong SW, Ma Q, Pu WT. Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. *Proc Natl Acad Sci U S A*. 2011;108(14):5632–5637.

61. Chen Z, Friedrich GA, Soriano P. Transcriptional enhancer factor 1 disruption by a retroviral gene trap leads to heart defects and embryonic lethality in mice. *Genes Dev*. 1994;8(19):2293–2301.

62. Deblois G, Giguere V. Functional and physiological genomics of estrogen-related receptors (ERRs) in health and disease. *Biochim Biophys Acta*. 2011; 1812(8):1032–1040.

63. Gass P, et al. Deficits in memory tasks of mice with CREB mutations depend on gene dosage. *Learn Mem*. 1998;5(4–5):274–288.

64. Heallen T, et al. Hippo pathway inhibits Wnt signaling to restrain cardiomyocyte proliferation and heart size. *Science*. 2011;332(6028):458–461.

65. Hummler E, et al. Targeted mutation of the CREB gene: compensation within the CREB/ATF family of transcription factors. *Proc Natl Acad Sci U S A*. 1994;91(12):5647–5651.

66. Lin Q, Schwarz J, Bucana C, Olson EN. Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C. *Science*. 1997;276(5317):1404–1407.

67. Naya FJ, et al. Mitochondrial deficiency and cardiac sudden death in mice lacking the MEF2A transcription factor. *Nat Med*. 2002;8(11):1303–1309.

68. Potthoff MJ, Olson EN. MEF2: a central regulator of diverse developmental programs. *Development*. 2007;134(23):4131–4140.

69. Rudolph D, Tafuri A, Gass P, Hammerling GJ, Arnold B, Schutz G. Impaired fetal T cell development and perinatal lethality in mice lacking the cAMP response element binding protein. *Proc Natl Acad Sci U S A*. 1998;95(8):4481–4486.

70. Sawada A, Kiyonari H, Ukita K, Nishioka N, Imuta Y, Sasaki H. Redundant roles of Tead1 and Tead2 in notochord development and the regulation of cell proliferation and survival. *Mol Cell Biol*. 2008; 28(10):3177–3189.

71. Vong L, Bi W, O'Connor-Halligan KE, Li C, Cserjesi P, Schwarz JJ. MEF2C is required for the normal allocation of cells between the ventricular and sinoatrial precursors of the primary heart field. *Dev Dyn*. 2006;235(7):1809–1821.

72. Hammer S, et al. Characterization of TBX20 in human hearts and its regulation by TFAP2. *J Cell Biochem*. 2008;104(3):1022–1033.

73. Mandel EM, et al. The BMP pathway acts to directly regulate Tbx20 in the developing heart. *Development*. 2010;137(11):1919–1929.

74. Sohal DS, et al. Temporally regulated and tissue-specific gene manipulations in the adult and embryonic heart using a tamoxifen-inducible Cre protein. *Circ Res*. 2001;89(1):20–25.

75. Zambrowicz BP, Imamoto A, Fiering S, Herzenberg LA, Kerr WG, Soriano P. Disruption of overlapping

transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells. *Proc Natl Acad Sci U S A*. 1997;94(8):3789–3794.

76. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002;30(1):207–210.

77. Fisher S, Grice EA, Vinton RM, Bessling SL, McCallion AS. Conservation of RET regulatory function from human to zebrafish without sequence similarity. *Science*. 2006;312(5771):276–279.

78. Fisher S, et al. Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. *Nat Protoc*. 2006;1(3):1297–1305.

79. Kawakami K, Takeda H, Kawakami N, Kobayashi M, Matsuda N, Mishina M. A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. *Dev Cell*. 2004;7(1):133–144.

80. Valouev A, et al. Genome-wide analysis of transcription factor binding sites based on ChIP-seq data. *Nat Methods*. 2008;5(9):829–834.

81. Kent WJ, et al. The human genome browser at UCSC. *Genome Res*. 2002;12(6):996–1006.

82. Smit AFA, Hubley R, Green P. RepeatMasker Open-3.0 [computer program]. Seattle, Washington, USA: Institute for Systems Biology; 1996–2010.

83. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol*. 1994;2:28–36.

84. Mahony S, Benos PV. STAMP: a web tool for exploring DNA-binding motif similarities. *Nucleic Acids Res*. 2007;35(Web Server issue):W253–W258.

85. Frith MC, Fu Y, Yu L, Chen JF, Hansen U, Weng Z. Detection of functional DNA motifs via statistical over-representation. *Nucleic Acids Res*. 2004;32(4):1372–1381.

86. Aerts S, Thijs G, Coessens B, Staes M, Moreau Y, De Moor B. Toucan: deciphering the cis-regulatory logic of coregulated genes. *Nucleic Acids Res*. 2003;31(6):1753–1764.

87. The R Development Core Team. *R: A Language And Environment For Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2010.

88. Taher L, Ovcharenko I. Variable locus length in the human genome leads to ascertainment bias in functional inference for non-coding elements. *Bioinformatics*. 2009;25(5):578–584.

89. Siepel A, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*. 2005;15(8):1034–1050.