Genetic variation in T-box binding element functionally affects SCN5A/SCN10A enhancer

Malou van den Boogaard,1 L.Y. Elaine Wong,1 Federico Tessadori,2 Martijn L. Bakker,1 Lisa K. Dreizehnter,1 Vincent Wakker,1 Connie R. Bezzina,3 Peter A.C. ’t Hoen,4 Jeroen Bakkers,2 Phil Barnett,1 and Vincent M. Christoffels1

1Department of Anatomy, Embryology, and Physiology, Heart Failure Research Center, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands. 2Hubrecht Institute-KNAW and University Medical Centre Utrecht, Utrecht, The Netherlands. 3Department of Experimental Cardiology, Heart Failure Research Center, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands.
4Center for Human and Clinical Genetics and Leiden Genome Technology Center, Leiden University Medical Center, Leiden, The Netherlands.

The contraction pattern of the heart relies on the activation and conduction of the electrical impulse. Perturbations of cardiac conduction have been associated with congenital and acquired arrhythmias as well as cardiac arrest. The pattern of conduction depends on the regulation of heterogeneous gene expression by key transcription factors and transcriptional enhancers. Here, we assessed the genome-wide occupation of conduction system–regulating transcription factors TBX3, NKX2-5, and GATA4 and of enhancer-associated coactivator p300 in the mouse heart, uncovering cardiac enhancers throughout the genome. Many of the enhancers colocalized with ion channel genes repressed by TBX3, including the clustered sodium channel genes Scn5a, essential for cardiac function, and Scn10a. We identified 2 enhancers in the Scn5a/Scn10a locus, which were regulated by TBX3 and its family member and activator, TBX5, and are functionally conserved in humans. We also provided evidence that a SNP in the SCN10A enhancer associated with alterations in cardiac conduction patterns in humans disrupts TBX3/TBX5 binding and reduces the cardiac activity of the enhancer in vivo. Thus, the identification of key regulatory elements for cardiac conduction helps to explain how genetic variants in noncoding regulatory DNA sequences influence the regulation of cardiac conduction and the predisposition for cardiac arrhythmias.

Introduction

Abnormal heart rhythms are common and cause significant morbidity and mortality. Arrhythmias may present around birth as part of a congenital heart defect, the most common etiology of birth abnormalities. However, the majority of arrhythmias develop with increasing age or as a comorbidity. These include sinus node dysfunction, atrioventricular (AV) block, and atrial fibrillation, the most common arrhythmia encountered in clinical practice causing significant morbidity and mortality (1). Sudden cardiac death from ventricular fibrillation, occurring mostly in the setting of coronary heart disease, is a leading cause of death (2). Although molecular, genetic, and physiological mechanisms have been described, most of the underlying causes are insufficiently understood (3).

Rhythmic and synchronized contraction of the atria and ventricles of the heart crucially relies on the proper generation and conduction of the cardiac electrical impulse. This process necessitates localized cardiomyocyte specialization in the different heart components (atria, AV node, ventricular conduction system, etc.) and demands that the cells possess specific electrophysiological characteristics. These intrinsic heterogeneities are rooted, at least in part, in the regional, compartment-specific differences in ion channel and gap junction gene expression, the defining factors of electrophysiology that derive action potential shape and intercellular propagation of the impulse (4–8).

The development and function of the heart components is driven by a set of interacting DNA binding transcription factors and chromatin remodeling proteins that form evolutionarily conserved regulatory networks (9–11). These transcription factors interact with specific noncoding regulatory DNA sequences (i.e., enhancers) of many genes to activate or repress their activity. It is becoming increasingly clear that these same transcription factors play essential roles in regulating the expression of ion channel and gap junction genes, affecting cardiac electrophysiology. Hallmark examples include transcription factors essential for heart development (e.g., NKX2-5 and TBX5), which activate gap junction subunit–encoding genes (e.g., CX40, encoded by Gja5) and ion channel genes and which, when mutated, can bring about electrical abnormalities such as AV conduction block and atrial fibrillation in humans (12–17). TBX2 and TBX3 are specifically expressed in — and critical for the development of — the sinus node, AV canal (AVC), and inferred AV node and AV bundle (18-21). These factors may compete with TBX5 for the same DNA binding sites, but repress expression of gap junction proteins (e.g., CX40 and CX43) and ion channels (e.g., the sodium channel Scn5a) in pacemaker tissues of the sinus node and AVC (18, 22). The interplay between transcription factors expressed in broader patterns (i.e., TBX5 is expressed in atria and entire conduction system) and factors that display restricted patterns (i.e., TBX3 is expressed in sinus node and AV conduction system) provide essential contributions to the electrical pattern of the heart (refs. 11, 16, 18, 23–28, and Supplemental Figure 1; supplemental material available online with this article; doi:10.1172/JCI62613DS1).

A series of genome-wide association studies (GWAS) (29–33) revealed that common variants in the population link genomic...
regions that modulate cardiac impulse conduction and repolarization, disturbances that increase the risk of arrhythmia. Interestingly, these variants were linked to essential cardiac transcription factors (e.g., \textit{NKX2-5}, \textit{TBX5}, and \textit{TBX3}) and ion channels (e.g., \textit{SCN5A} and \textit{SCN10A}). Several of the variants were identified in noncoding sequences nearby these genes, which indicates that they represent potential enhancers controlling the localized and tissue-specific expression of these genes. These observations indicated that variants in regulatory elements of ion channel genes modulate development and homeostasis of the localized electrophysiological properties of the heart. However, the difficulties involved in the identification of functional enhancers in vivo has meant that insight into this process is lacking. Here we investigated the transcriptional network by assessing the genome-wide DNA interaction of the key cardiac transcription factors TBX3, TBX5, and \textit{NKX2-5} and \textit{GATA4} and of the ubiquitously expressed enhancer-associated factor p300 (34–37) in the mouse heart in vivo, and identified many candidate cardiac enhancers. We further analyzed how known human genetic variants in these cardiac enhancers may affect transcription factor binding, and thus their activity. For the first time to our knowledge, we provide functional explanations for the effects of variants in noncoding DNA on cardiac conduction and arrhythmias.

### Results

\textit{ChIP-seq analysis in adult mouse hearts and de novo definition of consensus transcription factor binding sites.} The genome-wide occupancy profile of TBX3 in the heart was assessed using ChIP–massive parallel sequencing (ChIP-seq) analysis (38). \textit{Tbx3} is only expressed in the conduction system of the heart, a very small fraction of the total adult heart that is too small for ChIP-seq analysis (22). Therefore, \textit{Tbx3} expression was induced in all myocardial cells using mice with a Cre-inducible \textit{Tbx3} expression cassette (CT3) (18) and a cardiomyocyte-specific tamoxifen-inducible Cre expression cas-

---

**Figure 1**

In vitro validation of ChIP-seq datasets. (A) Overlap of all ChIP-seq binding regions with genes in refGene plus their promoters, defined here as the 1-kb upstream region of those genes. Values are percentage of all peaks. See Methods for a detailed description of peak definition. (B) Results obtained from the MEME motif discovery analysis. Binding motifs were in good agreement with those recently published for the HL-1 based ChIP-seq (35) and JASPAR motifs. (C) Number of overlapping binding regions between heart-derived ChIP-seq datasets. (D) ChIP-qPCR assay validating ChIP-seq peaks. Regions were randomly chosen from the group of ChIP-seq peaks that showed overlap between ChIP sequencing datasets for TBX3, GATA4 and \textit{NKX2-5}. Shown is enrichment relative to Hprt for \textit{NKX2-5}, GATA4, and TBX5 in WT adult mouse hearts and for TBX3 in TBX3-induced adult hearts. Ratios below represent the proportion of peaks that showed relative enrichment of more than 10. (E) Of the regions tested with ChIP-qPCR, 11 were tested for their response to \textit{NKX2-5}, GATA4, and TBX5 or TBX3 in Cos7 cells. All enhancers responded to stimulation with GATA4 and \textit{NKX2-5}. *P < 0.05.
sette (Myh6-mER-Cre-mER) (39). At 1 day after the last treatment with tamoxifen, animals appeared healthy, but showed strong downregulation of known TBX3 targets, including Gja1 (encoding CX43) and Gja5 (encoding CX40) (40). We reasoned that ectopically expressed TBX3 in the working myocardium would reveal binding sites normally occupied by TBX3 in the conduction system myocardium. However, TBX3 and TBX5 (members of the TBX2 subfamily expressed in adult heart) recognize highly similar binding sites, which indicates that this approach may also reveal TBX5 binding sites. Unfortunately, whereas ChIP of TBX5 was successful, we repeatedly failed to sequence enriched fragments because of technical problems during sequence library construction. Therefore, we performed ChIP–quantitative PCR (qPCR) and established that endogenous TBX5 in the adult heart interacted with the sites occupied by ectopically expressed TBX3 (see below). We also established the occupancy of endogenous NKK2-5 and GATA4, 2 cardiac transcription factors that cooperate with T-box factors (41). It has previously been shown that occupancy by the enhancer-associated protein p300 can predict the location of heart enhancers (34). To address this, we determined the occupancy profile of p300 (Supplemental Figure 2). In all cases we used ChIP-grade antibodies against the native factors.

We first evaluated the quality of our ChIP-seq experiments by determining the overlap in peaks between the TBX3 and control ChIP-seq datasets. These datasets possessed a total of 25,446 and 1,089 peaks, respectively, with an overlap of 131. The overlapping peaks were subtracted from the main TBX3 dataset as likely false positive peaks. Next, we determined the number of peaks for both the NKK2-5 and the GATA4 ChIP-seq datasets (Figure 1A), similarly removing peaks that overlapped with the control dataset. The resulting datasets were used to analyze the distribution of factor binding regions across the genome. For this analysis, a promoter region of 1 kb upstream of all genes was used to define the number of peaks associated with a promoter. The distribution of the remaining peaks was then assessed in terms of intragenic or intergenic. This analysis revealed that the majority of peaks in all 3 datasets were positioned within a gene sequence (exons and introns), with a significant enrichment of peaks associated with a promoter region ($P < 0.001$). Our p300 data in general were in good agreement with published data (34–36).

To investigate whether the ChIP peaks directly correlate to DNA binding elements for TBX3/TBX5, NKK2-5, and GATA4, we performed de novo motif discovery using MEME (42). Using 600 random peaks of each transcription factor, our MEME analysis yielded motifs (Figure 1B) in good agreement with those derived from ChIP-seq data of biotinylated TBX5, NKK2-5, and GATA4 expressed in HL-1 cells (ref. 35 and Supplemental Figure 3). We found near-perfect matches for the GATA4 motif in the JASPAR (43, 44) and TRANSFAC (45) databases. Interestingly, the NKK2-5 ChIP-seq motif showed a closer resemblance to the NKK3-2 consensus site found in the JASPAR database, compared with the NKK2-5 consensus site. The ChIP-seq motif for TBX3/TBX5 resembled the motif for T in the JASPAR database and for TBX5 in the TRANSFAC database.

TBX3, TBX5, NKK2-5, and GATA4 are known to interact with each other in the regulation of transcription (46, 47). Recent ChIP-seq analysis of biotinylated TBX5, NKK2-5, and GATA4 in HL-1 cells indicated that overlapping peak regions often represent bona fide regulatory elements (35). Therefore, we assessed the intersections of our in vivo TBX3/TBX5, NKK2-5, and GATA4 datasets (Figure 1C) and found that the intragenic portion of the 185 overlapping peak regions was significantly enriched for GO terms for heart and muscle system development (Supplemental Figure 4A). Comparison of terms associated with the molecular function of the NKK2-5–GATA4 overlap and the TBX3–NKK2-5–GATA4 intersect revealed that inclusion of TBX3 resulted mainly in an enrichment of terms associated with ion binding (Supplemental Table 4). Furthermore, we found an overlap of 63.2% of TBX3 peak regions with the TBX5 dataset derived from HL-1 cells (35), demonstrating good correlation between the binding peak data of these closely related T-box proteins, despite the difference between the 2 experimental systems. Conversely, the TBX20 peak dataset (48) showed only a 17.8% overlap with our TBX3 dataset, consistent with the observed difference in binding motif.

We selected 21 peak regions occupied by TBX3, NKK2-5, and GATA4 and validated binding in adult mouse hearts using ChIP-qPCR (Figure 1D). A mean enrichment of 70- to 80-fold was observed for the transcription factors at these genomic locations. Furthermore, of 45 TBX3-occupied sites tested, 43 were occupied by endogenous TBX3. In addition, 22 of 25 tested genomic sites were also occupied by endogenous TBX3 in the developing heart (Supplemental Figure 5). These data support the notion that TBX3 and TBX5 recognize the same sites in vivo and that ChIP of TBX3 ectopically expressed in the adult heart occupies sites bound by endogenous TBX3. Using a luciferase reporter analysis, we next

Table 1
Comparison of TBX3 peak–containing gene sets enriched or reduced in AVC with those induced or suppressed by TBX3

<table>
<thead>
<tr>
<th>TBX3 peak associated</th>
<th>TBX3-induced</th>
<th>TBX3-suppressed</th>
<th>Nonenriched controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enriched AVC</td>
<td>118 (20.4)</td>
<td>73 (13.9)</td>
<td>36 (6.2 ± 0.9)</td>
</tr>
<tr>
<td>Reduced AVC</td>
<td>117 (20.2)</td>
<td>182 (34.7)x</td>
<td>50.3 (9.6 ± 1.0)</td>
</tr>
</tbody>
</table>

$P < 0.01$ versus all other groups.

Table 2
GO term analysis associated with molecular function

<table>
<thead>
<tr>
<th>Genes</th>
<th>Count</th>
<th>%</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metal ion binding</td>
<td>67</td>
<td>26.9</td>
<td>$1.00 \times 10^{-3}$</td>
</tr>
<tr>
<td>Ion binding</td>
<td>68</td>
<td>27.3</td>
<td>$1.10 \times 10^{-3}$</td>
</tr>
<tr>
<td>Cation binding</td>
<td>67</td>
<td>26.9</td>
<td>$1.30 \times 10^{-3}$</td>
</tr>
<tr>
<td>Gated channel activity</td>
<td>11</td>
<td>4.4</td>
<td>$2.50 \times 10^{-3}$</td>
</tr>
<tr>
<td>Calmodulin binding</td>
<td>7</td>
<td>2.8</td>
<td>$2.80 \times 10^{-3}$</td>
</tr>
<tr>
<td>NAD(P)+-protein-arginine ADP-ribosyltransf. act.</td>
<td>3</td>
<td>1.2</td>
<td>$3.00 \times 10^{-3}$</td>
</tr>
<tr>
<td>Cation channel activity</td>
<td>10</td>
<td>4</td>
<td>$3.50 \times 10^{-3}$</td>
</tr>
<tr>
<td>Channel activity</td>
<td>12</td>
<td>4.8</td>
<td>$5.40 \times 10^{-3}$</td>
</tr>
<tr>
<td>Passive transmembrane transporter activity</td>
<td>12</td>
<td>4.8</td>
<td>$5.40 \times 10^{-3}$</td>
</tr>
<tr>
<td>Voltage-gated ion channel activity</td>
<td>8</td>
<td>3.2</td>
<td>$6.60 \times 10^{-3}$</td>
</tr>
<tr>
<td>Voltage-gated channel activity</td>
<td>8</td>
<td>3.2</td>
<td>$6.60 \times 10^{-3}$</td>
</tr>
<tr>
<td>Ion channel activity</td>
<td>11</td>
<td>4.4</td>
<td>$1.10 \times 10^{-2}$</td>
</tr>
</tbody>
</table>

GO term analysis was performed using DAVID. See Supplemental Tables 2–4 for additional GO analyses.
tested 11 of these 21 regions for their response to TBX3, TBX5, NKX2-5, and GATA4 in transient transfection assays (Figure 1E). All fragments were stimulated to varying degrees by NKX2-5 and GATA4. The effect of addition of TBX3 or TBX5 was variable, probably reflecting the complexity of enhancer function. Taken together, these results confirmed interaction of TBX3, TBX5, NKX2-5, and GATA4 with regions identified by ChIP-seq and that these regions acted as enhancers responding to the transcription factors.

Overlapping peak regions and gene expression arrays functionally classify TBX3 targets as conduction genes. TBX3 plays a critical role in the suppression of gene expression in the pacemaker and conduction system components (18, 22, 23). We hypothesized that genes expressed at lower levels in the AVC/AV node compared with the working myocardium are expected to be enriched in genes regulated directly by TBX3. We analyzed the overlap between the gene set containing TBX3 peaks and a set of genes differentially expressed between the AVC/AV node and working myocardium (Supplemental Table 1 and ref. 22). The set of genes reduced in the AVC displayed the greatest enrichment (Supplemental Figure 4B), which indicates that TBX3 is involved in the direct suppression of this subgroup of genes. Next, we correlated TBX3 binding and TBX3-mediated gene regulation by combining the gene set containing TBX3 peaks with a set of genes differentially expressed between the AVC/AV node and working myocardium (Supplemental Figure 4B). The enrichment of ion channel genes in the overlap analysis of these gene sets indicated that TBX3 is involved in the coordinated regulation of conduction genes. Quantitative RT-PCR analysis showed downregulation of 10 representative ion channel genes in atria of induced TBX3–expressing compared with control mice. These genes were significantly reduced in the AVC (P < 0.05) and contained ChIP-seq peaks for TBX3, GATA4, and NKX2-5. *P < 0.05.

Regulation of the Scn5a/Scn10a sodium channel gene locus by TBX3. SCN5A encodes Na\textsubscript{v}1.5, the main sodium channel responsible for rapid depolarization. Mutations in SCN5A underlie multiple inherited arrhythmia syndromes (26, 28). SCN10A (encoding Na\textsubscript{v}1.8) has only recently been implicated in cardiac conduction in a series of GWAS (30–33, 54). Several observations suggested that Scn5a and Scn10a are regulated in a coordinated fashion by TBX3 and TBX5. Scn5a and Scn10a were positioned adjacent in the genome, both contained sites occupied by TBX3/TBX5 and other factors (Figure 3A), and both were readily downregulated by TBX3. Both genes had similar relative spatial and temporal expression profiles (for description of the Scn5a profile, see refs. 55, 56), although Scn5a expression was stronger than that of Scn10a, and its level increased in the ventricles toward the end of gestation (Supplemental Figure 7A). In situ hybridization showed that both genes had very similar expression patterns: expressed in the atria; strongest in the mediastinal atrial walls (body of atria); and weaker in the auricles and in the trabeculae and crest of the interventricular septum, the precursors of the ventricular conduction system components (Figure 3B). This pattern resembled that of TBX5, a transcriptional activator expressed in atria and conduction system. Moreover, expression was absent from the TBX3-positive sinus node and AVC/AV node (Figure 3C), consistent with the notion that TBX3 suppresses these genes in its own expression domain. The genes flanking these 2 sodium channels, Scn11a and Exog, showed no binding regions for TBX3 and were not significantly downregulated by TBX3 (Supplemental Figure 7B), which indicates that TBX3-mediated regulation of this locus is likely to be confined to Scn5a and Scn10a. Taken together, these data indicate that Scn5a and Scn10a are coregulated, and that this process involves TBX3.

Identification of enhancers in the Scn5a/Scn10a locus. To gain insight into the mechanism of Scn5a/Scn10a regulation, we selected 9 fragments (F1–F9) in the Scn5a/Scn10a locus occupied by TBX3 for functional analysis using luciferase reporter assays (Figure 4A). Of these regions, 4 (F1, F2, F4, and F9) also...
contained NKX2-5, GATA4, and p300 peaks (Figure 3A) and overlapped with the TBX5 datasets (35). F1 and F2, located in a noncoding part of Scn10a, were strongly activated by NKX2-5 and GATA4. F9, located downstream of Scn5a, showed strong constitutive activity in H10 cells (Figure 4B). This effect was also observed for F4, located in a noncoding part of Scn5a that contains a region previously reported to affect Scn5a expression and thus modulate sodium channel function in vivo (57). Using an independent PCR-based assay, we confirmed the association of TBX3, TBX5, NKX2-5, and GATA4 and the mapped ChIP-seq positions to variable degrees of enrichment relative to Hprt gene control region in the adult heart (Figure 4C). In E10.5 mouse hearts, F1 and F2 were not found to be occupied, whereas F9 was occupied by endogenous TBX3 (Figure 4C).

Next, we determined the response of F1, F2, and F9 to T-box factors by means of luciferase assays. F1 and F2 were stimulated by TBX5 (activator) and repressed by both TBX3 and TBX2 (repressors). F9 was not stimulated by TBX5, but responded to TBX2 and TBX3 (Supplemental Figure 8B). The lack of response of F9 to transcription factor stimulation prompted us to analyze this region further (Supplemental Figure 8, A and C). Deletion analysis revealed a small, highly responsive fragment harboring binding sites for all transcription factors, indicative of the presence of sequences in F9 that act as suppressors.

To evaluate the in vivo activity pattern of the identified regulatory fragments, we cloned the mouse fragments into an Hsp68/LacZ reporter vector. In this assay, F1 and F2 were grouped as a single fragment, because of their close spatial association and broad

Figure 3
Scn5a and Scn10a are regulated by TBX3. (A) Overview of the ChIP-seq datasets within the Scn5a/Scn10a locus including flanking genes Scn11a and Exog. Top: TBX3, GATA4, NKX2-5, and p300 datasets (present study). Bottom: published ChIP-seq datasets showing tracks for biotinylated TBX5, GATA4, and NKX2-5 expressed in HL-1 cells (35); P300 in ED11.5 embryonic mouse heart (34); and P300 and POL2 in adult mouse heart (37). (B) In situ hybridization of Scn5a and Scn10a in E14.5 WT hearts showing expression pattern overlap, and the resemblance of the Scn5a downstream enhancer fragment F9 (F9 LacZ) with these expression patterns. Arrows denote areas of highest activity, in the atrial body and future ventricular conduction system components. san, sinoatrial node; ra, right atrium; la, left atrium; rv, right ventricle; lv, left ventricle; ivs, interventricular septum; drg, dorsal root ganglion. (C) In situ hybridization of Scn5a and Scn10a in the E17.5 sinoatrial node region showing complementary patterns to Tbx3 and partial overlap with Tbx5. Dotted outlines denote the sinus node region. Original magnification, ×5 (B); ×10 (B, insets, and C).
base coverage by p300 association sites (Figure 5A). The F1–F2 construct showed cardiac activity in 7 of 9 embryos. The activity was weakest in the atria and in the ventricles and strongest in the region of the interventricular septum (Figure 4D). For F9, cardiac activity was observed in 6 of 6 embryos. Activity was present in the atria and ventricles, yet once again strongest in the interventricular region (Figure 4E). The activity pattern of this regulatory fragment closely resembled that of both Scn5a and Scn10a, with highest activity in the atrial body and future ventricular conduction system components (Figure 3B). The TBX3-positive sinus node was devoid of lacZ expression, consistent with the notion that TBX3 represses this fragment. Recent evidence indicated that heart enhancers are not well conserved in evolution (34, 58). Therefore, it is quite possible that the human orthologous
regions do not function as enhancers. To test this possibility, the human positional orthologs of the mouse fragments were used in the same in vivo transgenic reporter assay. Human F1–F2 and F9 both showed activity patterns highly similar to their mouse counterparts (Figure 4, F and G), which indicates that their function is evolutionarily conserved. Furthermore, we tested whether transcription factors bind to the human fragments. ChIP-qPCR on human ventricular samples revealed that NKX2-5 indeed occupied these orthologous fragments in vivo (Figure 5C), thereby conserving their function (Figure 5B).

A genetic variant within the enhancer in SCN10A functionally affects a binding element for TBX3/TBX5. Recent GWAS studies have identified genetic loci influencing conduction system function, but the underlying mechanisms are not known (30–33, 54). The reported GWAS SNP rs6801957 is positioned in the conserved enhancer F1 in SCN10A (Figure 5B) and may affect SCN5A/SCN10A expression. This SNP is positioned directly under a TBX3 ChIP-seq peak and is linked to a G-to-A substitution in the highly conserved portion of the T-box binding consensus sequence (Figure 5D). Enhancer fragments cloned from carriers of the major allele (i.e., G) could be stimulated by TBX5 and repressed by TBX3 in the presence of NKX2-5 and GATA4 (Figure 6A). In contrast, the fragment cloned from a minor allele (i.e., A) carrier did not respond to these T-box factors. Specific mutations generated in the fragments revealed that this SNP, but not other SNPs in this fragment that are in linkage disequilibrium (LD), was involved in the T-box factor response (Figure 6A). Therefore, rs6801957 appeared to functionally affect gene expression of SCN10A/SCN5A. Furthermore, analysis of all other SNPs (n = 25) in LD (r² > 0.6) revealed that rs6801957 was uniquely associated with peaks in our ChIP-seq datasets (Supplemental Table 11). Gel mobility shift assays using DNA probes mimicking both major and minor human alleles demonstrated a significant reduction of T-box association with the minor allele compared with the major allele (Figure 6B and Supplemental Figure 9).

Next, we tested the activity of the 2.2-kb human F1–F2 enhancer fragment in zebrafish using the ZED vector, which allows simultaneous screening for transgenesis and enhancer-driven reporter activity in the heart (59); this analysis revealed that 70% of red fluorescent protein–expressing transgenic fish displayed enhancer-driven green fluorescence in the heart (Figure 6C). When the minor allele for rs6801957 was introduced in the enhancer fragment, this fraction dropped to about 40% (Figure 6D). When the minor allele for the SNP rs6795970 — which presumably affects SCN10A splicing and conduction (31) — was introduced, the fraction of fish displaying enhancer-driven green fluorescence in the heart did not change. However, when rs6801957 was additionally introduced, the fraction again dropped to about 40% (Figure 6D). Taken together, these in vivo data confirmed that rs6801957 affects enhancer function, which indicates that each allele carrying this SNP may express reduced levels of SCN10A, SCN5A, or both.

Figure 5
In vitro analysis of the homologous human genomic region F1–F2 and SNP associated with conduction. (A) Relative position of SNP rs6801957, located under overlapping peaks for ChIP-seq datasets TBX3, GATA4, NKX2-5, and p300. (B) Sequence conservation between species around rs6801957, located in a T-box binding motif. Gaps are denoted by asterisks. (C) ChIP-qPCR assay with NKX2-5 on human (h-) ventricular tissue for F1, F2, and F9 showed enrichment for all 3 fragments. (D) TBX3 binding motif found through TBX3 ChIP-seq.
Regulation of the cardiac electrical pattern by a network of transcriptional regulators and noncoding regulatory sequences of ion channel genes. GWAS identified common variants that modulated cardiac impulse conduction and repolarization. Several of these variants were positioned within noncoding DNA in the vicinity of genes encoding key cardiac transcription factors (e.g., TBX5 and TBX3) and ion channels (e.g., SCN5A and SCN10A), which suggested the presence of a regulatory network for the electrophysiological properties of the heart (29–33). Tissue-specific genes are regulated by enhancers, usually found in noncoding DNA, each driving a subset of its gene's entire expression profile. These enhancers thus represent ideal candidates for involvement in complex diseases, such as arrhythmias, as variants in an individual enhancer result in cell type-specific changes in gene expression (60, 61). The genome-wide occupancy of several factors, including TBX5, NKX2-5, and GATA4, has previously been assessed in vitro in HL-1 cells (35), the factors being first biotinylated to enable their efficient immunoprecipitation. Furthermore, the occupancy of GFP-fused TBX20 was also recently assessed in the heart (48). Here, we derived a ChIP-seq protocol that allowed us to identify — to our knowledge for the first time — the occupancy of the native forms of cardiac transcription factors in the heart in vivo. Our data, together with the existing data, represent an invaluable resource providing a map for the location of tissue-specific enhancers in the adult heart that regulate genes involved in the local electrophysiological properties of the heart.

Analysis of the genes interacting with and differentially regulated by TBX3 indicated that TBX3 functions mainly as a transcriptional repressor in vivo, consistent with previous observations using reporter assays (18, 62, 63). The gene set occupied by TBX3, repressed by TBX3, and reduced in the TBX3-positive AVC was highly enriched in genes encoding ion channels or involved in ion transport functions. This finding suggests that an important function of TBX3 in the conduction system is the direct regulation of electrophysiological properties. Both Scn10a and Scn5a were among the channel genes suppressed by TBX3 (and stimulated by TBX5), revealing functional parts of the regulatory network indicated by the GWAS studies.

Discussion

Regulation of the cardiac electrical pattern by a network of transcriptional regulators and noncoding regulatory sequences of ion channel genes. GWAS identified common variants that modulated cardiac impulse conduction and repolarization. Several of these variants were positioned within noncoding DNA in the vicinity of genes encoding key cardiac transcription factors (e.g., TBX5 and TBX3) and ion channels (e.g., SCN5A and SCN10A), which suggested the presence of a regulatory network for the electrophysiological properties of the heart (29–33). Tissue-specific genes are regulated by enhancers, usually found in noncoding DNA, each driving a subset of its gene's entire expression profile. These enhancers thus represent ideal candidates for involvement in complex diseases, such as arrhythmias, as variants in an individual enhancer result in cell type-specific changes in gene expression (60, 61). The genome-wide occupancy of several factors, including TBX5, NKX2-5, and GATA4, has previously been assessed in vitro in HL-1 cells (35), the factors being first biotinylated to enable their efficient immunoprecipitation. Furthermore, the occupancy of GFP-fused TBX20 was also recently assessed in the heart (48). Here, we derived a ChIP-seq protocol that allowed us to identify — to our knowledge for the first time — the occupancy of the native forms of cardiac transcription factors in the heart in vivo. Our data, together with the existing data, represent an invaluable resource providing a map for the location of tissue-specific enhancers in the adult heart that regulate genes involved in the local electrophysiological properties of the heart.

Analysis of the genes interacting with and differentially regulated by TBX3 indicated that TBX3 functions mainly as a transcriptional repressor in vivo, consistent with previous observations using reporter assays (18, 62, 63). The gene set occupied by TBX3, repressed by TBX3, and reduced in the TBX3-positive AVC was highly enriched in genes encoding ion channels or involved in ion transport functions. This finding suggests that an important function of TBX3 in the conduction system is the direct regulation of electrophysiological properties. Both Scn10a and Scn5a were among the channel genes suppressed by TBX3 (and stimulated by TBX5), revealing functional parts of the regulatory network indicated by the GWAS studies.

Figure 6

In vitro and in vivo analysis of genomic region F1–F2 and SNPs associated with conduction. (A) Luciferase assays performed in Cos7 showed that rs6801957 was involved in T-box factor response. Representative human major (G) alleles Maj1 and Maj2 responded to stimulation by TBX5 and repression by TBX3, whereas this effect on the minor (A) allele Min1 was significantly less. Mutating rs6801957 in these alleles showed a similar effect. *P < 0.01. Error bars represent SD. See Supplemental Table 10 for information on variants per allele. (B) EMSA showing relative oligonucleotide binding of MBP-TBX3 (+) compared with MBP only (−). MBP-TBX3 fusion associated well with both Nppa probe and human Maj1, but less with Min1 and Mut (containing a 3-bp mutation in the proposed T-box binding element). See Supplemental Figure 9 for quantification and statistical analysis. (C) Representative image of the zebrafish heart (dotted outline) containing ZED-Maj1, showing expression in the ventricle (V). A, atrium. Scale bar: 100 μm. (D) Percent GFP expression in hearts of zebrafish containing the major or minor allele for rs6801957. ZED represents the vector without genomic region F1–F2. Representative human major alleles Maj1 and Maj3 showed GFP expression in 60%–70% of the hearts of zebrafish containing these constructs. When major alleles were mutated into minor alleles (Maj1Mut and Maj3Mut), a significant reduction in GFP expression was found. The presence of rs6795970 in Maj3, also linked to conduction disease, had no significant effect on GFP expression in the heart.
Identification of specific T-box factor target sites. TBX3-positive cardiomyocytes of the adult conduction system were limited in number and not sufficient for the ChIP-seq assay. To overcome this problem, we used a model in which Tbx3 was ectopically activated in all cardiomyocytes, assuming that this allows identification of binding sites occupied by TBX3 in the conduction system cardiomyocytes. Although the levels of Tbx3 expression in this model are moderate, and tolerated by the atria for a prolonged period (18), it is possible that TBX3 associates with sites normally occupied by other cardiac T-box proteins. Of the T-box proteins expressed in the adult myocardium (i.e., TBX5 and TBX20), TBX5 most closely resembles TBX3 (46, 47), and its binding sites are likely to be detected by ectopically expressed TBX3. Indeed, the consensus DNA binding site we identified for TBX3 binding resembled those identified for TBX5 (35), both of which diverge from previously defined T-box binding motifs (44, 45). The recently described TBX20 motif derived from ChIP-seq data (48) differs from the TBX3/TBX5 motif and known consensus sites by 1 position with no nucleotide preference in the middle of the sequence. To rule out possible differences in the computational prediction approach, we confirmed the TBX20 motif as published using MEME software. Thus, the observed difference in binding consensus may genuinely reflect a substantial in vivo functional difference in TBX20-specific gene target recognition. Assuming that the TBX20-GFP fusion protein used in this study mimics native TBX20, this finding may reflect the closer phylogenetic relationship of TBX3 and TBX5, which belong to the TBX2 subfamily, than to TBX20, which belongs to the TBX1 subfamily (41, 64). Preferential binding of the different cardiac T-box T-box subfamilies, each playing important roles in the heart, may greatly influence their target specificity, affecting the mechanism underlying their function in heart development and gene regulation. In this respect, it will also be of great interest to fully unravel how T-box proteins interact with their target genes in the context of varying protein complexes to regulate gene expression.

Identification of functional enhancers for the cardiac sodium channels. SCN5A, encoding the cardiac sodium channel, is essential for heart function (26, 65). Modest reduction in SCN5A expression levels slows cardiac conduction and predisposes to arrhythmias under certain conditions (i.e., monogenic disease and drug therapy) (57, 66). GWAS have implicated SCN10A in cardiac conduction. Its function in the heart remains unclear. Na,1.8 (encoded by Scn10a) may function in the myocardium to either decrease (31) or increase (33) conduction. Alternatively, Na,1.8 may function in the vagal innervation of the heart (67). We here identified key enhancers in Scn5a and Scn10a in mice and humans, providing more detailed insight into the link between the transcription factor network and the regulation of these important sodium channel genes in the heart. We hypothesized that these sequences represent key enhancers, both of which contribute to TBX5/TBX3-regulated expression patterns of Scn5a/Scn10a. More extensive analysis, such as chromatin conformation capture (68) and genetic modification of mice, will be required to investigate the precise contributions of each of the enhancers to the regulation of the 2 sodium channel genes.

Recent ChIP-seq of p300 occupation in mouse and human hearts indicated that cardiac enhancers are not well conserved throughout evolution (34, 69). The sequences we identified by ChIP-seq in mice may not be functionally conserved in humans. Therefore, we performed ChIP for Nkx2-5 on human ventricular samples and found that the orthologous human DNA fragments were occupied by this factor. Furthermore, these elements were able to drive expression in patterns strikingly similar to their mouse counterparts. Together, these data revealed that the enhancers were functionally conserved in humans.

We provide the first functional evidence to our knowledge for the direct involvement of a human SNP in cardiac conduction system gene regulation. This SNP was found to be associated with increased PR/QRS duration, which suggests that the SNP — or a variant in LD with the SNP — decreases SCN5A/SCN10A expression. Our functional analysis indicated that the SNP itself converts a highly conserved nucleotide within the consensus T-box binding site and selectively affects TBX5/TBX3-mediated activity of the enhancer fragment, thereby reducing the cardiac activity of the enhancer. We hypothesize that the SNP affects the TBX5-activated/TBX3-repressed expression of SCN5A/SCN10A in vivo. It is well possible that the enhancer regulates both SCN5A and SCN10A. Given that SCN5A provides a much larger contribution to sodium current and conductivity of the heart than does SCN10A, it seems probable that altered SCN5A expression underlies the observed influence of the SNP on conduction system function. Other genetic variants in noncoding DNA identified by GWAS in this and the other loci (30–33, 54) may influence conduction system function in a similar way. Nevertheless, the effect of the variant on conduction times was very small (33). Furthermore, the GWAS identified heterozygous variants associated with conduction system function, whereas in our functional studies, we studied the effect of the isolated variants, equivalent to homozygosity for the variant. In this respect, it is likely that multiple genetic variants together can cause a cumulative biologically significant modulation of sodium channel gene expression.

In summary, the present study identified a set of enhancers and key binding factors important for cardiac conduction system development and function, and further identified disrupted transcription factor binding as a mechanism likely underlying the effect of a genetic variant at the SCN10A/SCN5A locus on cardiac conduction. Uncovering the molecular network and the noncoding regions regulating conduction system function will help to understand how they affect conduction system function, and may offer tools to predict disease susceptibility and progression among individuals (70).

Methods

ChIP sequencing from adult mouse heart. CAG-CAT-TBX3–heterozygous (CTBX3+/–) mice (18) were crossed with Myb6MCM mice (39). Tamoxifen (catalog no. T5648; Sigma-Aldrich) was injected intraperitoneally (100 mg per mouse per day) at 4-day intervals in adult males. Hearts were removed on day 5, prior to the onset of heart dysfunction (40). Quantitative RT-PCR was performed to determine the efficiency of recombination. For Nkx2-5, Gata4, and p300 ChiPs, hearts were isolated from adult FVB males. Tissues were cross-linked in 1% (p300) or 2% (Gata4, and Scn5a) formaldehyde for 30 minutes at room temperature. Cross-linking was quenched by addition of 0.125M glycine. Tissues were dissociated by IKA Ultra Turrax T5 FU, pelleted, and resuspended in cold lysis buffer (50 mM Tris-HCl, pH 8.1; 10 mM EDTA, pH 8.0; 1% SDS; and 1x Protease Inhibitor Cocktail [Roche]). Nuclei were obtained by use of a tight glass dounce homogenizer. Cross-linked nuclei were sonicated under conditions established to yield an average fragment size of approximately 300 bp. Antibodies TBX3 (A-20, sc-17871, lot no. J0306), TBX5 (C-20, sc-17866, lot no. D1708), Nkx2-5 (N-19, sc-8697, lot nos. A1109 and J2010), Gata4 (C-20, sc-1237, lot no. F3010), and p300 (N-15, sc-583, lot no. K3009) were purchased from Santa Cruz. Immunoprecipitation, washing, elution, and reverse cross-linking were performed as previously described (71). Quality of the ChIP was assessed with primers on locations of known cardiac enhancers. Quality was also previously qualitatively assessed by noting peaks at published T-box respon-
sive enhancers (52). Material was used for library preparation with standard NEBNext library kit protocol for Illumina sequencing (for nos. E6000S/L and E6040S/L; New England Biolabs). ChIP-seq data were deposited in NCBI’s Gene Expression Omnibus (72) under accession no. GSE35151 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE35151).

ChIP-qPCR was performed on a Roche LightCycler 480 System using Sybr Green detection. Fold enrichment indicates the ratio of ChIPed DNA to a negative control region in the Hprt locus, normalized for input DNA. Tested locations were randomly chosen using the Galaxy software interface. Primer sequences are provided in Supplemental Table 6.

ChIP-qPCR on human ventricular samples. The human ventricular sample used for ChIP was from a nondiseased human heart that was technically unsuitable for transplantation based on logistic considerations and was provided by A. Varró (University of Szeged, Szeged, Hungary). Before cardiac explantation, the organ donor did not receive medication except dobutamine, furosemide, and plasma expanders.

Processing of ChIP sequencing data. Sequencing results were run through the standard Illumina GA Pipeline (version 1.3 for GAII runs) to convert images to reads of 36 bp in length (unaligned sequences produced by the Illumina Genome Analyzer) and edit for quality (FIRECREST, Bustard, and GERALD). Reads were then processed and aligned to the mouse reference genome (Mm9, NCBI Build 37) as previously described, allowing a maximum of 2 mismatches and retaining only sequences with unique alignments (73).

Peaks were extracted and defined from the wiggle files containing the coverage per bp using an in-house algorithm that returns peak start and stop coordinates in a BED file format. BED files for TBX3, GATA4, and NKX2-5 are shown in Supplemental Table 12. Locations of known cardiac enhancers were checked for transcription factor binding as a quality control for these datasets. See Supplemental Methods for the exact settings for peak calling.

Gene sets were extracted from UCSC refGene for the mouse using the Galaxy software interface. We allocated promoter regions for the 27,585 mouse genes currently listed under refGene; a promoter was defined as the 1-kb upstream region of an annotated transcriptional start site. Overlaps between gene regions and peak regions were identified using Galaxy. For determining the intersection between gene arrays and TBX3 ChIP-seq peak regions, significantly up- or downregulated genes from the various microarray datasets were assigned coordinates using the BioMart Central Portal. All chromosomal coordinates were subsequently extended both up- and downstream by 1 kb. Coordinates were then intersected using the Galaxy software interface. 2 control gene sets were generated, both for the induced TBX3 array analysis and the AVC/working myocardium array analysis, using 6 sets of equivalent numbers of unique genes, randomly extracted from the nonsignificant genes of the latter study and further treated in exactly the same way as the test gene sets. Statistical analysis of the significance of enrichment between intersects and controls was carried out using a Z test, which uses the normal approximation of the binomial distribution to compare proportions (74, 75). Microarray data were deposited in NCBI’s Gene Expression Omnibus (72) with accession nos. GSE13614 and GSE31969.

Proteins assigned to overlaps were extracted using Galaxy and further analyzed using the DAVID bioinformatics resource gene ontology (GO) analysis software (49, 50). Unique gene names were called using chromosome locations via the BioMart Central Portal. Motif analysis was performed with MEME (42) using 600 random peaks, with an average length of approximately 165 bp, per total dataset. For scanning ChIP-seq peaks with known transcription factor motifs, position weight matrices were obtained from JASPAR (43, 44) and TRANSFAC (version 11.3; ref. 45) databases.

Transgenic mouse enhancer assay. Enhancer candidate regions of 2–3 kb in the Scn5a/Scn10a locus were cloned into the Hsp68-LacZ reporter vector as previously described (76). Genomic coordinates of the amplified regions are provided in Supplemental Table 9. DNA was injected into the pronucleus of 0.5-day-old fertilized FVB/N eggs, which were subsequently transferred into the oviducts of CD-1 pseudopregnant foster females (Cyagen Inc.). At least 200 injections were performed per construct. Embryos were harvested and stained with X-gal to detect LacZ activity, and yolk sacs were processed for PCR genotyping.

In vivo zebrafish enhancer assay. Enhancer sequences Maj1, Maj1Mut, Maj3, and Maj3Mut were amplified by PhusionTAQ PCR (New England Biolabs) and cloned into the ZED vector (59) by Gateway technology (Life Technologies BV). ZED-Maj1, ZED-Maj1Mut, ZED-Maj3, and ZED-Maj3Mut were injected in WT zebrafish embryos at the 1-cell stage at a final concentration of 15 ng/μl in the presence of 25 ng/μl TOL2 transposase RNA. Embryos were kept at 28.5°C in E3 medium and scored for heart-specific red fluorescent protein, and subsequently heart-specific GFP, fluorescence at 48 hours post fertilization (hpf) on a Leica MZFLIII fluorescence stereomicroscope (Leica Microsystems GmbH) set up with appropriate fluorescence filters.

In vivo imaging of the embryos at 48 hpf was carried out on a Zeiss Axioskop 2 Mot plus fluorescence microscope (Carl Zeiss MicroImaging GmbH) mounted with a Leica DFC490 CCD camera (Leica Microsystems) using appropriate fluorescence filters.

Other methods. Procedures for quantitative expression analysis, luciferase assays, in situ hybridization, and EMSA were performed as described previously (18). See Supplemental Methods and Supplemental Tables 5, 7, and 8 for details.

Statistics. Unless otherwise indicated, a P value less than 0.05 was considered significant. See Supplemental Methods for details.

Study approval. This investigation conformed to NIH guidelines (Guide for the Care and Use of Laboratory Animals. NIH publication no. 85-23. Revised 1996). All animal work was approved by the Animal Experimental Committee of the Academic Medical Center, University of Amsterdam, and carried out in compliance with Dutch government guidelines. Investigations using the human ventricular sample conformed to the principles outlined in the Helsinki Declaration of the World Medical Association. Use of human cardiac material was approved by the Ethical Review Board of the Medical Center of the University of Szeged.

Acknowledgments

The authors thank Jan Ruijter, Bouke de Boer, Corrie de Gier-de Vries, Carol Verhoek-Pocock, Ted Bradley, Frank Baas, Marcel Willemsen, Antoine van Kampen, Kim Beijer, and Henk Buermans for their contributions. We thank András Varró for providing the human cardiac tissue. This work was supported by grants from the Netherlands Heart Foundation (NHS grant 2010B205, to V.M. Christoffels) and by the European Community’s Seventh Framework Programme contract (“CardioGeNet” 223463, to M. van den Berghe and V.M. Christoffels). C.R. Bezzina is supported by the Netherlands Heart Foundation (NHS 2005T024). P.A.C. ’t Hoen is supported by the Center for Medical Systems Biology, in the framework of the Netherlands Genomics Initiative/Netherlands Organization for Scientific Research.

Received for publication December 23, 2011, and accepted in revised form May 10, 2012.

Address correspondence to: Phil Barnett or Vincent M. Christoffels, Department of Anatomy, Embryology, and Physiology, University of Amsterdam, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands. Phone: 3120.5667822; Fax: 3120.6976177; E-mail: v.m.christoffels@amc.uva.nl (P. Barnett). Phone: 3120.5667821; Fax: 3120.6976177; E-mail: v.m.christoffels@amc.uva.nl (V.M. Christoffels).


