A common genetic variant within *SCN10A* modulates cardiac *SCN5A* expression

Malou van den Boogaard, …, Phil Barnett, Ivan P. Moskowitz


Variants in *SCN10A*, which encodes a voltage-gated sodium channel, are associated with alterations of cardiac conduction parameters and the cardiac rhythm disorder Brugada syndrome; however, it is unclear how *SCN10A* variants promote dysfunctional cardiac conduction. Here we showed by high-resolution 4C-seq analysis of the *Scn10a-Scn5a* locus in murine heart tissue that a cardiac enhancer located in *Scn10a*, encompassing *SCN10A* functional variant rs6801957, interacts with the promoter of *Scn5a*, a sodium channel–encoding gene that is critical for cardiac conduction. We observed that *SCN5A* transcript levels were several orders of magnitude higher than *SCN10A* transcript levels in both adult human and mouse heart tissue. Analysis of BAC transgenic mouse strains harboring an engineered deletion of the enhancer within *Scn10a* revealed that the enhancer was essential for *Scn5a* expression in cardiac tissue. Furthermore, the common *SCN10A* variant rs6801957 modulated *Scn5a* expression in the heart. In humans, the *SCN10A* variant rs6801957, which correlated with slowed conduction, was associated with reduced *SCN5A* expression. These observations establish a genomic mechanism for how a common genetic variation at *SCN10A* influences cardiac physiology and predisposes to arrhythmia.

Find the latest version:

http://jci.me/73140-pdf
A common genetic variant within SCN10A modulates cardiac SCN5A expression

Malou van den Boogaard,1 Scott Smemo,2 Ozanna Burnicka-Turek,3,4 David E. Arnoldsd,3,4 Harmen J.G. van de Werken,5 Petra Klous,6 David McKean,7 Jochen D. Muehlhagem,8 Julia Moosmann,9 Okan Toka,9 Xinan H. Yang,3,4 Tamara T. Koopmann,10 Michiel E. Adriaens,10 Connie R. Bezzina,10 Wouter de Laat,6 Christine Seidman,7 J.G. Seidman,7 Vincent M. Christoffels,1 Marcelo A. Nobrega,2 Phil Barnett,1 and Ivan P. Moskowitz2,3,4

1Department of Anatomy, Embryology and Physiology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands. 2Department of Human Genetics, 3Department of Pediatrics, and 4Department of Pathology, University of Chicago, Chicago, Illinois, USA. 5Department of Cell Biology, Erasmus Medical Center, Rotterdam, The Netherlands. 6Institute-KNAW and University Medical Center Utrecht, Utrecht, The Netherlands. 7Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. 8Department of Anesthesiology, Perioperative and Pain Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, Massachusetts, USA. 9Department of Pediatric Cardiology, Children’s Hospital University of Erlangen-Nuremberg, Erlangen, Germany. 10Department of Clinical and Experimental Cardiology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands.

Variants in SCN10A, which encodes a voltage-gated sodium channel, are associated with alterations of cardiac conduction parameters and the cardiac rhythm disorder Brugada syndrome; however, it is unclear how SCN10A variants promote dysfunctional cardiac conduction. Here we showed by high-resolution 4C-seq analysis of the Scn10a-Scn5a locus in murine heart tissue that a cardiac enhancer located in Scn10a, encompassing SCN10A functional variant rs6801957, interacts with the promoter of Scn5a, a sodium channel–encoding gene that is critical for cardiac conduction. We observed that SCN5A transcript levels were several orders of magnitude higher than SCN10A transcript levels in both adult human and mouse heart tissue. Analysis of BAC transgenic mouse strains harboring an engineered deletion of the enhancer within Scn10a revealed that the enhancer was essential for Scn5a expression in cardiac tissue. Furthermore, the common SCN10A variant rs6801957 modulated Scn5a expression in the heart. In humans, the SCN10A variant rs6801957, which correlated with slowed conduction, was associated with reduced SCN5A expression. These observations establish a genomic mechanism for how a common genetic variation at SCN10A influences cardiac physiology and predisposes to arrhythmia.

Introduction

Genome-wide association studies (GWAS) on ECG measures in diverse populations have consistently associated noncoding genetic variants within introns of the voltage-gated sodium channel gene SCN10A with cardiac conduction system (CCS) function (1–9). Recent work associated noncoding variants at SCN10A with Brugada syndrome, a cardiac rhythm disorder with a high risk of sudden death (10). The implication of SCN10A, which had not been previously studied in the heart, in cardiac rhythm parameters and Brugada syndrome by GWAS (Figure 1A) was unexpected. Given the position of the associated variants within the gene, including one whose minor allele caused a nonsynonymous substitution in SCN10A, it was assumed that the variants were functionally connected with Na+(1.8), the protein product of SCN10A. However, functional follow-up analysis of SCN10A provided conflicting evidence about whether Na+(1.8) accelerates (5) or slows (2) conduction velocity, and the magnitude of the effects of blocking Scn10a in mice was small (2, 5). Therefore, the role of SCN10A in cardiac physiology remained controversial (11–13).

Interestingly, SCN10A maps immediately adjacent to SCN5A (encoding Na+(1.5)) on chromosome 3p22.2. In contrast to SCN10A, SCN5A has a well-described role in cardiac physiology and pathophysiology (14–18). CCS function is exquisitely sensitive to Na+(1.5) dose, and heterozygous mutations in SCN5A underlie numerous human conduction system diseases, including Brugada syndrome, long QT syndrome, atrial fibrillation, progressive cardiac conduction disease, and sudden cardiac death (14, 19, 20). Furthermore, noncoding variants within introns and downstream of SCN5A have also been associated with ECG parameters in GWAS.

We recently demonstrated the importance of the T-box transcription factors Tbx5 and Tbx3 in the regulation of Scn5a in mice (21, 22). The rapid conducting portions of the CCS require high levels of Na+(1.5) for function (23). Strong Scn5a expression is driven in the fast conduction system by Tbx5, a transcriptional activator (22). Conversely, repression of Scn5a is a key feature ensuring the slow propagation of the cardiac electrical impulse across the sinus node and atrioventricular node (17, 23). Tbx3, a transcriptional repressor, plays a major role in the repression of Scn5a expression in slowly conducting nodal regions of the developing heart (24, 25). We previously investigated the gene regulatory landscape at the Scn10a-Scn5a cluster and identified 2 T-box–regulated enhancers, one positioned downstream of Scn5a and the other in an intron of Scn10a. A SNP, rs6801957, associated with QRS prolongation and in linkage disequilibrium (LD) with variants associated with Brugada syndrome (5, 10) is located in the Scn10a intronic enhancer. This SNP modulates T-box factor binding and activity of the isolated enhancer fragment (21).
We investigated the novel hypothesis that genetic variation at SCN10A associated with cardiac rhythm disease and conduction mechanistically functions by modulating SCN5A expression. We also tested the hypothesis that the regulatory landscape of SCN5A may be regulated by multiple cis-regulatory elements spread over long-range distances, including cis-regulatory elements within introns of SCN10A. In such a scenario, functional SNPs at SCN10A associated with conduction parameters and disease may reflect disruption of these regulatory elements causing altered expression of SCN5A. Using high-resolution chromatin conformation capture (4C), we demonstrated the intricate association of the Scn10a intronic enhancer with the promoters of Scn5a and Scn10a and the downstream Scn5a enhancer. We found that SCN5A was expressed at high levels, whereas SCN10A was expressed at background levels, in the adult human and mouse heart by RNA-seq. We found that the cis-regulatory element located in Scn10a regulated the pattern of Scn5a expression in the heart, and that the rs6801957 SNP affected cardiac expression of Scn5a using a BAC reporter strategy. Finally, we found a direct correlation of SCN5A expression in humans with the presence of the rs6801957 risk-associated SNP in the SCN10A intronic enhancer. Together, our data provided a genomic mechanism explaining how common genetic variants at SCN10A influence cardiac physiology and predispose to arrhythmia.

Results
High-resolution 4C-seq reveals interactions between an enhancer in Scn10a and the Scn5a promoters. We previously identified 2 regions with conserved enhancer activity able to autonomously drive...
expression of a reporter construct in the murine CCS (21, 22). The region termed enhancer A (EnhA) spans 2 adjacent introns and 1 exon in Scn10a, and enhancer B (EnhB) is positioned 15 kbp downstream of Scn5a (Figure 1B). EnhA contains a SNP previously found to be associated with QRS prolongation (5). This SNP disrupts a T-box factor binding site that affects EnhA activity in a zebrafish assay (21). Enhancers regulate gene expression by physically interacting with the promoter of a gene. We used the newly developed tool high-resolution 4C-seq (26) to investigate on a genome-wide scale which gene promoters are contacted by EnhA and EnhB (Figure 1C). We aligned 4C-seq with cardiac ChIP-seq data for TBX3 (21) as well as for Pol2 and p300 (27–29) to visualize interactions for EnhA and EnhB (Figure 1C). By setting 4C bait regions around 2 known promoter regions for Scn5a, we found that both EnhA and EnhB possessed clear cis interactions with the Scn5a and Scn10a promoters (Figure 1B, C, and Supplemental Figure 1; supplemental material available online with this article; doi:10.1172/JCI73140DS1). Moreover, contacts between Scn5a promoters A and B and the Scn10a promoter were observed. When observed from bait sets in the enhancers themselves, EnhA was found to associate with the Scn5a and Scn10a promoters and the EnhB region, and EnhB was found to associate with the Scn5a promoter region and EnhA. These interaction results revealed that EnhA (in Scn10a) and EnhB (downstream of Scn5a) interacted with both promoters and with each other, which indicates that the enhancers and promoters form a complex (Figure 1D). This topology, in which both enhancers and gene promoters are in close contact, suggests that both enhancers are potential regulators of Scn5a and Scn10a. This is consistent with the observation that both genes show similar spatial patterns of expression in the fetal heart (21). Furthermore, these data indicate that the variants identified in cardiac rhythm and disease GWAS in EnhA may affect the regulation of Scn5a, a dosage-sensitive regulator of cardiac conduction (Figure 1D).

Substantial SCN5A transcripts, but not SCN10A transcripts, in the adult mouse and human heart. In the fetal heart, expression of cardiac Scn5a is much higher than that of Scn10a (21), which indicates that the contribution of Scn10a to the sodium current before birth is small. To evaluate the expression levels of Scn5a and Scn10a in the adult, we performed RNA-seq on whole mouse hearts (30). We found that Scn5a expression was 145-fold higher than Scn10a expression (53.81 ± 3.59 versus 0.37 ± 0.08 reads per million [rpm]; Figure 2A). Therefore, Scn10a expression in the heart as a whole was considered to be at background levels. Nevertheless, it remained conceivable that Scn10a expression was present in the limited subset of CCS cells based on the pattern of Scn10a expression before birth (21). We therefore evaluated Scn5a and Scn10a expression in the atroventricular bundle from adult mouse hearts microdissected at 6 weeks of age (Figure 2B). We confirmed the anatomic specificity of isolated tissue, observing robust expression of Gja5 (also known as Connexin40), a gene whose expression in the ventricles is limited to the conduction system (31, 32). In the atroventricular bundle, expression of Scn5a was high, whereas Scn10a expression was undetectable (Figure 2B).

We next evaluated SCN5A and SCN10A expression from human heart samples by RNA-seq (Figure 2C). SCN5A expression was more than 1,000-fold greater than SCN10A expression in the whole heart, atria, and ventricles (whole heart, 326 ± 149 versus 0.24 ± 0.22 rpm, P = 2.5 × 10–27; atria, 205 ± 138 versus 0.12 ± 0.17 rpm, P = 1.0 × 10–3; ventricles, 376 ± 126 versus 0.29 ± 0.21 rpm, P = 1.1 × 10–26; Figure 2C). The absolute magnitude of SCN10A expression, less than 1 part per million, corresponded to a cellular content of less than 1 SCN10A RNA molecule per 2 cells (33–35) and qualified as not expressed in the heart (30). These data indicated that SCN5A but not SCN10A is expressed at appreciable, physiologically relevant levels in the heart. Therefore, we hypothesized that GWAS SNPs in noncoding regions at the SCN10A locus affecting cardiac conduction (2–5) modulate SCN5A expression.

The enhancer in Scn10a is required for expression of Scn5a in vivo. We tested the hypothesis that both EnhA and EnhB are necessary for endogenous Scn5a expression by engineering a BAC reporter system for Scn5a in the mouse. Specifically, we used recombineering strategies (36) to replace the endogenous first Scn5a coding exon with the LacZ reporter gene in mouse BAC RP23-198L19. This Scn5a-LacZ BAC included the entire Scn5a coding region, EnhA, and EnhB (Figure 3A). Independent transgenic lines carrying the WT Scn5a-LacZ reporter BAC RP23-198L19 showed a pattern of LacZ expression highly reminiscent of endogenous Scn5a expression (17, 21), including reproducible expression in structures of the heart (37–39), which indicates that both EnhA and EnhB are necessary for Scn5a expression in vivo.
the ventricular conduction system including the atrioventricular bundle and ventricular myocardium in adult (3 of 3) and embryonic (6 of 6) hearts (Figure 3B, Supplemental Figure 2, A–F, and Supplemental Figure 4A).

To investigate the necessity of EnhA and EnhB for Scn5a-LacZ reporter expression, we engineered deletions of EnhA (chr9:119,540,800–119,544,032), EnhB (chr9:119,378,051–119,379,479), or both from the BAC reporter and then evaluated Scn5a-LacZ expression in vivo. Removal of EnhA alone significantly abrogated Scn5a-LacZ; expression of LacZ was absent from ventricular myocardium in all independent adult and embryonic EnhA deletion BAC transgenics (Figure 3C, Supplemental Figure 2, G–O, and Supplemental Figure 4B). LacZ expression was disrupted in the CCS in the majority of EnhA deletion BAC transgenics. When present, LacZ expression was confined to the distal ventricular septum (Figure 3C, Supplemental Figure 2, G–O, and Supplemental Figure 4B). Removal of EnhB alone also substantially altered and reduced LacZ expression from the Scn5a locus (Figure 3D, Supplemental Figure 2, P–V, and Supplemental Figure 4C). Interestingly, the region of ventricular CCS Scn5a-LacZ expression that was maintained in EnhB deletion BAC transgenics (Figure 3D, Supplemental Figure 2, R–T, and Supplemental Figure 4C) was reciprocal to that maintained in EnhA deletion BAC transgenics (Figure 3C, Supplemental Figure 2, H–J, and Supplemental Figure 4B). Removal of both EnhA and EnhB together completely eliminated Scn5a-LacZ expression in all independent adult and embryonic hearts. When present, LacZ expression was confined to the distal ventricular septum (Figure 3C, Supplemental Figure 2, G–O, and Supplemental Figure 4B).
embryonic transgenics (Figure 3E, Supplemental Figure 2, W–AC, and Supplemental Figure 4D). Taken together, these observations demonstrated that EnhA and EnhB are necessary and sufficient for in vivo Scn5a expression.

The SNP rs6801957, associated with conduction slowing in GWAS, leads to lower SCN5A expression in mice and in humans. The observation that EnhA, which is located in Scn10a, was required for Scn5a expression suggested that common genetic variation affecting conduction system physiology and disease at SCN10A may modify SCN5A expression. A sentinel SNP located in EnhA (Figure 1A and Figure 3A). The major allele for rs6801957, G, establishes a conserved canonical T-box binding site (GGTGACAG) and promotes enhancer activity, whereas the minor allele A, associated with conduction slowing, disrupts a core nt in the T-box site (GGTAACAG) and decreases enhancer activity (21). This SNP is in strong LD with the nonsynonymous Scn5a SNP evaluated in prior studies (rs6795970, HapMap 2 CEU population, D’ = 1.0, r2 = 0.96, refs. 2, 3), with 2 other SNPs that implicated SCN10A in cardiac physiology in independent studies (rs6800541, HapMap 2 CEU population, D’ = 1.0, r2 = 0.96, ref. 4; rs6798015, HapMap 2 CEU population, D’ = 0.955, r2 = 0.84, ref. 1), and with the SNP that implicated SCN10A in Brugada syndrome (rs10428132, HapMap 2 CEU population, D’ = 1.0, r2 = 0.97, ref. 10) (Figure 1A).

We tested the allele-specific effects of the murine nt orthologous to SNP rs6801957 on Scn5a expression. The rs6801957 minor allele (A) was recombineered into the orthologous region of the Scn5a-LacZ BAC, and LacZ expression was analyzed in adult murine hearts (Figure 3F, Supplemental Figure 3, and Supplemental Figure 4E). The minor allele at rs6801957 substantially altered Scn5a-LacZ expression: expression of LacZ was markedly disrupted in the compact myocardium of the majority of independent transgenics (Figure 3F, Supplemental Figure 3, and Supplemental Figure 4E). LacZ expression was also significantly disrupted in the CCS of the majority of independent adult (4 of 6) and half of embryonic (8 of 16) transgenics (P = 0.04; Figure 3F, Supplemental Figure 3, and Supplemental Figure 4E). We concluded that the risk allele orthologous to SNP rs6801957 substantially decreased Scn5a expression and, conversely, that the major allele orthologous to SNP rs6801957 is necessary for normal CCS Scn5a expression.

We next tested whether rs6801957 affects the sufficiency of human EnhA sequences to drive CCS expression in transgenic mice in vivo (Figure 4). Human EnhA with the major allele of rs6801957 was sufficient to drive Scn5a-LacZ expression in myocardium and the CCS in the majority of transient transgenic mouse embryos (Figure 4A). However, EnhA with the minor allele was insufficient for cardiac expression (P = 0.05; Figure 4B). We concluded that the major allele of SNP rs6801957 is necessary for normal CCS activity of EnhA.

We hypothesized that rs6801957 influenced Scn5a expression in humans. Therefore, we evaluated Scn5a mRNA levels as a function of genotype at this SNP in 2 independent heart tissue bio-banks by RNA-seq and quantitative real-time RT-PCR (qRT-PCR). In the first set, GG samples (homozygous for the major allele as associated with shorter PR and QRS intervals) expressed significantly more Scn5a mRNA (428 ± 128 normalized rpm; n = 9) than GA (331 ± 104 rpm; n = 18) or AA (322 ± 66 rpm; n = 6) tissue samples. In this set, each additional copy of the A allele reduced cardiac Scn5a expression by 12% ± 6% (additive genetic model, P = 0.01; dominant genetic model, P = 0.01). Scn10a was not appreciably expressed in any genotype. In the second set, samples homozygous for GG at rs6801957 expressed significantly more Scn5a mRNA (428 ± 128 normalized rpm; n = 9) than GA (331 ± 104 rpm; n = 18) or AA (322 ± 66 rpm; n = 6) tissue samples. In this set, each additional copy of the A allele reduced cardiac Scn5a expression by 12% ± 6% (additive genetic model, P = 0.01; dominant genetic model, P = 0.01). Scn10a was not appreciably expressed in any genotype. In the second set, samples homozygous for GG at rs6801957 expressed significantly more Scn5a mRNA (428 ± 128 normalized rpm; n = 9) than GA (331 ± 104 rpm; n = 18) or AA (322 ± 66 rpm; n = 6) tissue samples. In this set, each additional copy of the A allele reduced cardiac Scn5a expression by 12% ± 6% (additive genetic model, P = 0.01; dominant genetic model, P = 0.01). Scn10a was not appreciably expressed in any genotype. In the second set, samples homozygous for GG at rs6801957 expressed significantly more Scn5a...
RNA (0.060 ± 0.029 normalized expression; n = 33) than GA (0.048 ± 0.017 normalized expression; n = 37) or AA (0.050 ± 0.028 normalized expression; n = 17) tissue samples (additive genetic model, P = 0.11; dominant genetic model, P = 0.03). In this set, although the level of SCN5A mRNA did not increase linearly per copy of G allele, genotype at rs6801957 was significantly associated with SCN5A mRNA levels with the same direction of effect as in the first sample set. SCN10A variant rs6801957 had no detectable effect on SCN10A expression, although power to detect allele specific expression on SCN10A would be very poor, given its low expression in adult cardiac tissue.

Discussion

Our results unveil the genomic regulatory logic controlling Scn5a expression, with implications for interpreting cardiac GWAS and understanding the genomic basis of CCS function and disease. The physical associations between enhancers and promoters at the Scn10a-Scn5a locus were analyzed by 4C, a technique that provides a genome-wide view of all cis interactions (26, 37). Viewpoints from both EnhA and EnhB revealed interactions with the Scn5a and Scn10a promoter regions and, additionally, multiple contact points in and beyond Scn5a, fitting with ChIP-seq data. These contacts were confirmed when taking the promoters as viewpoint. Our findings indicate that the enhancers and promoters at this locus are all in close proximity, allowing the enhancers to simultaneously interact with and potentially regulate the transcription of both Scn5a and Scn10a. A conserved genomic architecture that enables enhancer sharing and coregulation of genes within an evolutionary conserved cluster has also been observed for the Irex and Hox gene clusters (38). Although the enhancers contact the Scn10a promoter, quantification of Scn10a transcript levels indicated that Scn10a transcription was not activated by these enhancers. Therefore, we focused our attention on the functionality of interactions between these enhancers and Scn5a. Using BAC transgenesis, an assay complementary to 4C contact mapping, we demonstrated the functionality of the interactions between EnhA and EnhB and Scn5a. The enhancers functioned in a modular pattern to determine Scn5a expression: EnhA, located in Scn10a, was required for proximal ventricular CCS expression; EnhB, located downstream of Scn5a, was required for distal ventricular CCS expression; and both enhancers were required for ventricular myocardium expression.

Our data indicate that the physiological role underlying common variants at SCN10A in cardiac rhythm control is not ascribed to the SCN10A gene product Na1.8, but instead to the SCN5A gene product Na1.5. SNPs identified by GWAS have modest effect sizes, an attribute inherent to their common frequency (39). This suggests that noncoding SNPs associated with human phenotypes by GWAS may cause expression alterations to genes with dosagesensitive physiology. This paradigm held only for SCN5A, not SCN10A. SCN5A has an exclusively dosage-sensitive relationship to ECG parameters and cardiac rhythm control. Heterozygous mutations in human SCN5A underlie numerous human cardiac conduction system diseases, including Brugada syndrome, long QT syndrome, atrial fibrillation, progressive cardiac conduction disease, and sudden cardiac death (14, 19, 20). Furthermore, noncoding variants within introns and downstream of SCN5A have also been associated with ECG parameters in GWAS. The mouse homoygous null phenotype is embryonic lethal, and mice haploinsufficient for Scn5a display a host of profound cardiac rhythm disturbances, including slowed cardiac conduction and a Brugada-like phenotype, similar to those associated with SNPs at neighboring SCN10A (2, 5). In contrast, whether the SCN10A product Na1.8 plays a role in cardiac physiology remains controversial (11–13). The Scn10a knockout mouse displayed subtle cardiac ECG abnormalities, including modest PR interval shortening indicative of more rapid than normal conduction (2). This phenotype is at odds with the role of voltage-gated sodium channels in the rapid phase of myocardial depolarization, in which a decrease in channel function and concomitant sodium current would slow conduction (as observed by lowering Scn5a expression). Moreover, functional analysis of Scn10a indicated that the effects of pharmacologically blocking or genetically inactivating Scn10a are small, and provided conflicting evidence about whether Na1.8 accelerates (5) or slows (2) conduction velocity. Expression analysis also supports a more substantial role for the gene product of Scn5a than that of Scn10a. During embryonic development, low levels of Scn10a expression can be qualitatively observed using in situ hybridization (21, 22) and qRT-PCR, demonstrating expression levels 4–5 times lower than those of Scn5a. However, as development progressed to adult stages, Scn10a expression remained low, whereas Scn5a transcript levels were induced to levels 100-fold greater than those of Scn10a. In humans, the expression discrepancy was 1,000-fold in favor of SCN5A over SCN10A. Although transcripts for both genes could be detected in heart tissue by PCR-based approaches, the confluence of data—including our SCN10A transcript quantification in adult heart tissue, our genotype-SCN5A expression correlation, and physiological observations of Scn10a knockout mice by others (2)—suggest that the contribution of Na1.8 to cardiac physiology is minor. The absence of substantial effects in the Scn10a knockout, combined with the dramatic effects on electrophysiology observed for Scn5a knockout models (16, 18, 40), supports the contention that common genetic variation at the SCN10A locus affects cardiac physiology via dysregulation of SCN5A expression. Our observations do not formally rule out the possibility of other mechanisms affecting conduction physiology at the SCN10A-SCN5A locus. However, the totality of the evidence, including our present observations, strongly supports the relevance of the genomic mechanism we describe to cardiac conduction physiology and arrhythmia risk.

Mechanistically connecting noncoding genetic variations to their associated traits requires a commitment to understanding the functional biology of cis-regulatory elements, given that up to 85% of GWAS SNPs associated with human disease traits are noncoding (41, 42). Here, we presented an integrative and comprehensive experimental strategy to functionally link noncoding variants mapping within long-range cis-regulatory elements to their target genes. Combining high-resolution 4C strategies with engineering of BACs and mouse transgenic reporter assays, we demonstrated that the regulatory landscape of SCN5A extended into SCN10A and that the ECG-associated SNPs within SCN10A modified the function of an SCN5A enhancer. Using genomics and genetics data, we further demonstrated that these SNPs were associated with expression of SCN5A, but not SCN10A, in human hearts. Together, our data established SCN5A as the functional target of the noncoding variants within SCN10A associated with ECG parameters. The present work provides a genomic mechanism for the effect of common genetic variants at SCN10A on cardiac physiology and disease and provides a strong rationale for careful interrogation of the noncoding activity of GWAS variants prior to...
ascribing function to annotated genes. Our findings also suggest a genomic pathophysiological mechanism for Brugada syndrome, with the potential to refine diagnosis and implications for future therapeutic approaches.

Methods

Experimental animals

Generation of Scn5a-LacZ BAC transgenic mice was performed by the University of Chicago Transgenic Core Facility by pronuclear microinjection of DNA. Mice were maintained on a mixed genetic background and analyzed at E13.5, E14.5, and 12 weeks of age. All experiments used age-, gender- and genetic strain-matched controls to account for any variations in data sets compares across experiments. Mice were bred and housed in specific pathogen-free conditions in a 12-hour light/12-hour dark cycle and allowed ad libitum access to standard mouse chow and water. Mice requiring medical attention were provided with appropriate veterinary care by a licensed veterinarian and were excluded from the experiments described. No other exclusion criteria existed.

4C template preparation

4C templates were prepared as previously described (43). In short, adult mouse hearts were isolated in ice-cold PBS. Single-cell suspensions were obtained by dissociation of tissue with IKA Ultra Turrax T5 FU, followed by dounce homogenization. Chromatin was cross-linked with 2% formaldehyde in PBS with 10% FCS for 10 minutes at room temperature, nuclei were isolated, and cross-linked DNA was digested with a primary restriction enzyme recognizing a 4-bp restriction site (DpnII), followed by proximity ligation. Cross-links were removed, and a secondary restriction enzyme digestion (Csp6I) was performed, followed again by proximity ligation. For all experiments, 200 ng of the resulting 4C template was used for the subsequent PCR reaction, of which 16 (total, 3.2 μg of 4C template) were pooled and purified for next-generation sequencing. The PCR products were purified using 2 columns per sample of the High Pure PCR Product Purification Kit (catalog no. 11732676001; Roche). The kit separates the PCR products larger than 120 bp from the adaptor-containing primers (~75 and ~40 nt in size). Similar results were obtained with products from a single PCR reaction (200-ng template).

4C-seq primer design

PCR primers were designed based on the following criteria. The size of the viewpoint fragment was at least 500 bp, to allow efficient cross-linking to other DNA fragments. The fragment end (the region between the primary and secondary restriction enzyme) was more than 350 bp, to allow efficient circularization during the second ligation step. Primers were designed to be maximally 20 nt in length. The strategy therefore produces sequencing reads (36mers in this study) composed of the 4C primer sequence (20nt, specific to a given viewpoint) followed by 16 nt that identify a captured sequence. The reading primer always hybridizes to, and ends at, the 3‘ side of the first restriction recognition site. This design ensures analysis of only primary ligation events and provides sufficient sequence information to unambiguously identify most captured sequences. The nonreading primers, 18–20 nt in size, were designed at a distance ≤100 bp from the secondary restriction site. All primers had 35%–65% GC content and an optimal basic temperature of 55°C, ranging 45°C–65°C. Primers were checked against the mouse genome with MegaBLAST23 (settings, -p 88.88, -W 12, -e 1, -F T), which requires primers on the reading side to be matched uniquely in the genome and primers on the nonreading side to have a maximum of 3 perfectly matching BLAST high-scoring segment pairs (HSP). See Supplemental Table 1 for all primers used.

4C data analysis and statistics

4C templates were mixed and sequenced simultaneously in 1 Illumina HiSeq 2000 lane. The sequence tags generated by the procedure were prefixed by the 4C reading primer, which includes the DpnII restriction site sequence (see 4C-seq primer design). The 4C reading primer sequences were separated from multiplexed 4C-seq libraries, and the suffixes were extracted for further processing. Mapping and filtering of the sequence reads was done as previously described (26). The algorithm constructs a background model for remote intra- and interchromosomal contacts to correct for systematic biases that can occur during the 4C-seq experimental protocol. The algorithm is designed to use controls for sequencing errors and nonunique sequences while considering the high coverage (100x–100,000x) of fragment ends that are proximal to the viewpoint fragment. To normalize the interactions in close proximity to the viewpoint, the algorithm was used to calculate the median of normalized coverage for running windows of size 4 kb (depicted as black line) and sliding windows of 2–50 kb of linearly increasing size (depicted as color-coded multiscale diagrams). All median values represent enrichment relative to the maximum attainable 4-kb median value, whereas sliding windows represent enrichment relative to the maximum attainable 12-kb median value. The 20th and 80th percentiles were also computed and depicted as the green area around the 4-kb running windows (Supplemental Figure 1).

mRNA-seq and genotyping rs6801957

Ventricular tissue set 1. Human left ventricular samples (n = 39) were obtained from patients of mixed ancestry who were given left ventricular assist devices. Discarded left ventricular tissue was obtained at the time of insertion of the device. Human right atrial samples (n = 4) were obtained as discarded tissue from patients who had cardiac surgery to repair their congenital heart disease lesion. The expression levels of Scn5a and Scn10a in adult mouse and human cardiac tissue were assessed by RNA-seq as described previously (30, 33–35). Samples were blinded to genotype at the SCN locus. Post-hoc analyses showed that there was no association between lane or batch and expression of SCNSA or SCN10A expression, independent of genotype. The rs6801957 genotype was assessed by dideoxy sequencing of PCR-amplified genomic DNA using standard procedures (44). rs6801957 sequences were amplified using the following primers: primer A, CACCTG-GAGGCTCCCTAAGA; primer B, GAGTCTGTAGCTCTCCCATAG. Primer A was used as sequencing primer. A linear additive regression model was used to determine statistical significance (45).

Ventricular tissue set 2. Human left ventricular samples (n = 129) were collected from nonimplanted human donor hearts that were considered suitable for transplantation, yet not used due to logistical reasons. All heart samples used in this study were from individuals of self-reported European descent and were collected at centers in Szeged (gift from A. Varro, University of Szeged, Szeged, Hungary), Sydney (gift from C. dos Remedios, University of Sydney, Sydney, Australia), Miami (gift from N.H. Bishopric, University of Miami School of Medicine, Miami, Florida, USA), and Vanderbilt (gift from A.L. George, Vanderbilt University, Nashville, Tennessee, USA). Samples were flash-frozen and stored in liquid nitrogen. RNA and DNA were isolated from cardiac tissue following standard protocols. Manhattan distance hierarchical clustering using genome-wide SNP genotypic data confirmed a genetically homogeneous group. Genotype at rs6801957 was determined by means of a Taqman assay (Applied Biosystems). cDNA was prepared using Oligo-dT and Thermoscript First-Strand Synthesis System (Invitrogen) and used for qRT-PCR for determination of SCNSA transcript levels. All qRT-PCR assays were performed in triplicate. Expression values were normalized by dividing by the geometric mean of HPRT1 and TNNJ3 mRNA expression values (46). For each set of triplicates, the mean, SD, and coefficient of
variation were calculated. Sets with a coefficient of variation higher than 0.20 were left out of the analysis, as previously described (47), leaving a total of 87 subjects for analysis. The association between rs6801957 and SCN5A transcript levels (Supplemental Figure S) was tested using an additive and dominant genetic model, taking gender, age, and center of origin as covariates. All calculations were performed in statistical programming language R, version 2.15.3.

**BAC modification**

Mouse BAC RP23-198L19 (mm9; chr9:119,351,100–119,598,458) was engineered in vitro using recombineering kits and protocols from Gene Bridges GmbH (Heidelberg). BAC RP23-198L19 was converted into an Sn5A transcriptional reporter (Sn5A-LacZ) by inserting a LacZ-ampicillin cassette, in frame, replacing the first exon of Sn5A using the Red/ET recombination kit and protocol from Gene Bridges (catalog no. K001). Primers used for generation of the recombineering cassette were as follows: Sn5A-Lai forward, CTCGCCAGGCACCAATGTGCTCCGGCT; Sn5A-Lai reverse, AGCCTCTTCTTCGAGCTGCTGAGCTCCCGAAGGCT; Sn5A-Lai I forward, GCTTGT-CGAAGGCG. Ampicillin-resistant colonies were screened for homologous recombination using primers in the unaffected genomic region flanking the insertion and within the vector cassette, and the insertion junction was verified by sequencing using following the primers: Sn5A pI forward, CATCACACCTCCT-GTTTGTCTC; Sn5A pI reverse, GGTACTATAGAAAGGGTCCAGGTCT; Sn5A pI 3 forward, TGGATGGAATCAGATTAAGTGTGA. Recombinant BACs were confirmed intact by restriction digest fingerprinting with PspXI (NEB).

EnhA and EnhB were deleted in the Sn5A-LacZ BAC through their targeted replacement, again using the Red/ET recombination kit and protocol from Gene Bridges (catalog no. K001). EnhA (chr9:119,540,801–119,544,032) was replaced with a spectinomycin resistance gene amplified from the rpsL-neo plasmid (BAC modification of origin as covariates. All calculations were performed in statistical programming language R, version 2.15.3.

**Statistics**

Values represent mean ± SD of the indicated number of measurements. Statistical significance was determined using 2-tailed Student’s t test or a linear additive regression model. A P value less than 0.05 was considered significant. No statistical method was used to predetermine sample size, and animal experiments were not performed in a blinded fashion. Mice were assigned at random to treatment groups for all mouse studies.

**Acknowledgments**

We thank Nanette H. Bishopric, Alfred L. George, Andras Varro, and Cristobal dos Remedios for providing samples for expression analysis in human donor hearts. The study was funded by grants from the NIH (R01 HL114010, to I.P. Moskowitz and M.A. Nobrega), AHA (10PRE2670012, to D.E. Arnold), the European Community’s Seventh Framework Programme contract ("CardioGeNet" 223463, to V.M. Christoffels), the Cardiovascular Onderzoek Nederland (PRE-DICT, to C.R. Bezzina), the German Foundation for Heart Research (F/09/10, to O. Toka), and the Federal Ministry of Education and Research (BMBF; FKZ01GI0601, to O. Toka).


