Transcriptional control of ion channel genes in arrhythmogenesis

van den Boogaard, M.

Publication date
2017

Document Version
Other version

License
Other

Citation for published version (APA):
van den Boogaard, M. (2017). Transcriptional control of ion channel genes in arrhythmogenesis. [Thesis, fully internal, Universiteit van Amsterdam].
FROM GWAS TO FUNCTION
GENETIC VARIATION IN SODIUM CHANNEL GENE
ENHANCER INFLUENCES ELECTRICAL PATTERNING

Malou van den Boogaard, Phil Barnett, Vincent M. Christoffels

*Trends in Cardiovascular Medicine* 2014 Apr;24(3):99-104
ABSTRACT
The electrical activity of the heart depends on the correct interplay between key transcription factors and cis-regulatory elements, which together regulate the proper heterogeneous expression of genes encoding for ion channels and other proteins. Genome-wide association studies of ECG parameters implicated genetic variants in the genes for these factors and ion channels modulating conduction and depolarization. Here, we review recent insights into the regulation of localized expression of ion channel genes and the mechanism by which a single-nucleotide polymorphism (SNP) associated with alterations in cardiac conduction patterns in humans affects the transcriptional regulation of the sodium channel genes SCN5A and SCN10A. The identification of regulatory elements of electrical activity genes helps to explain the impact of genetic variants in noncoding regulatory DNA sequences on regulation of cardiac conduction and the predisposition for cardiac arrhythmias.

INTRODUCTION
The development and function of organ systems requires a precise spatio-temporal gene expression pattern, regulated by a distinct transcriptional architecture of the genome. Together with protein transcription complexes, genomic cis-regulatory elements, such as promoters and enhancers, play a pivotal role in the cell-type specific expression of genes. Currently, approximately 400,000 putative enhancers have been mapped in the human genome, the function of which largely remains to be elucidated. This emphasizes the enormous complexity of the regulatory mechanisms that guide the development and proper function of a myriad of different cell types (Figure 1). Considering the high density of regulatory elements in the genome, it is tempting to postulate that part of the phenotypic variation between individuals exists through genetic variation within these elements. In fact, several recent publications have implicated alterations in regulatory elements as the cause for disease. By linking genome-wide maps of genetic variants to phenotypic traits and exploring the chromatin state in different cell types, we are now finally beginning to grasp the effects of genetic variation on transcriptional regulation on a larger scale.

EFFECTS OF NONCODING VARIANTS ON TRANSCRIPTION
Many complex common diseases lack a clear Mendelian pattern of inheritance, but instead display a more polygenic etiology. In order to investigate the underlying risk factors for these diseases, researchers developed a method that evaluates the association between genetic variants and a phenotypic trait of interest in a large number of individuals. These genome-wide association studies (GWAS) have been deployed
for many different diseases and traits, thereby identifying common variants (single nucleotide polymorphisms; SNPs) that potentially affect disease susceptibility. However, GWAS on a specific disease may not identify variants that predispose to this disease. As diseases are often only diagnosed when a clinical threshold is surpassed, a large group of carriers of predisposing (common) variants will be present in the healthy control population. Therefore, as an alternative approach for the identification of genetic variation underlying complex disease, GWAS use intermediate phenotypes, which are traits associated with, or predisposing to particular diseases. For example, instead of collecting subjects with atrial fibrillation, GWAS have been performed for ECG indices in healthy subjects.4–6 Using these data, variants

---

**Figure 1**
Enhancers regulate spatial and temporal gene expression. (A) Schematic overview of the regulatory elements in the genome. Transcriptional activity is absent without specific enhancer activity. Silencers (repressive elements) have not been depicted in this overview. (B) RNA polymerase II and other basal transcription factors (TFs) are recruited to the promoter of Gene B (grey). Through binding of specific TFs to Enhancer Y (green) and recruitment of general co-activator proteins (e.g. p300) to this complex, a chromatin loop is formed between the enhancer and the promoter of Gene B, resulting in specific expression of Gene B in the adult atria. Enhancer Y does not have an effect on Gene A, because of an active intermediate insulator (red). (C) Binding of TFs to the neighboring enhancer X (blue) results in the same process, but drives specific expression of Gene B in the adult ventricles. Again, enhancer X does not have an effect on Gene A, because of the active intermediate insulator.
associated with prolonged PR-interval, a trait associated with increased risk of atrial fibrillation, can be isolated and investigated for their contribution to disease. As such, these GWAS pinpoint genetic variants that ultimately affect biological processes, which contribute to clinical symptoms. To date, the majority of identified variants have been found in noncoding regions of the genome. Recent evidence suggests that a subset of these variants affect transcriptional mechanisms through modulation of regulatory elements.

Based on the high modularity of regulatory elements, each regulating transcription at different time points and in different tissues, and the high frequency of variants in the population, one can expect that single nucleotide variants will have only minor contributions to the trait or disease risk. However, experiments aimed to comprehend the transcriptional control of protein-coding genes have demonstrated that the synergistic activity of different transcription factors is represented in the sequence of the regulatory element itself. Subsequently, a single mutation in a transcription factor binding site (TFBS) can have a significant effect on transcription.

By focusing on disease-associated variants and mutations it may be possible to uncover regulatory elements that are crucial in the development of disease. Lettice et al. demonstrated this concept by identifying an enhancer of the SHH gene in a region harboring mutations causing preaxial polydactyly. Variants in noncoding regulatory sequences will mainly affect transcription, resulting in a quantitative loss of protein. However, as these are still functional proteins, it is unlikely that every trait- or disease-associated variant will have such striking effects on disease development. One hypothesis is that these regulatory variants increase the risk for disease and that a certain combination of variants and mutations will exacerbate a pre-existing condition or tip the balance towards a state in which disease becomes evident.

This was shown in Hirschsprung’s disease, where mutations in an intronic enhancer of RET increase the risk for disease. Both the SHH enhancer and the RET enhancer were found through mapping of causal mutations for human disease. This approach seems fairly straightforward, but in the case of associated variants discovered by GWAS, any SNP that is in linkage disequilibrium (LD) with the sentinel SNP could represent the actual causal variant. This might explain why a highly common variant located in an enhancer upstream of the proto-oncogene MYC and associated with increased risk for both colorectal and prostate cancer, did not correlate with altered MYC transcription levels in patients with colorectal tumors. However, deletion of the entire variant-containing enhancer, resulted in reduced transcript levels of MYC in the colon and a marked reduction in tumor formation in mice. A possible explanation is that nearby SNP alleles tend to be inherited together more often than expected by chance, as they arise through mutational events that occur once in an
ancestral haplotype background. As a result, the sentinel SNP indicated by GWAS marks a region of interest, and not particularly the causal variant. In order to understand and explain the pathological process underlying complex common disease, the relationship between these variants and the increased risk for disease has to be investigated.

**GENETIC VARIATION AND THE CARDIAC CONDUCTION SYSTEM**

The cardiac conduction system (CCS) represents a highly specialized component of the heart that facilitates the initiation and propagation of the cardiac electrical impulse. The cardiac conduction system can be divided into a slowly and a rapidly conducting component. The slow-conducting pacemaker tissues consist of the sinoatrial node (SAN), which generates the impulse, and the atrioventricular node (AVN), which delays and propagates the impulse from the atria to the ventricles. The rapid conducting ventricular conduction system (VCS) is composed of the AV bundle (AVB), right and left bundle branches (BB) and Purkinje network. As such, it coordinates myocardial contraction, which is regulated through the well-balanced activity of different ion channels and gap junction proteins. These proteins define cardiac electrophysiological properties, giving rise to the action potential and intercellular propagation of the cardiac electrical impulse. Through local differences in ion channels and gap junction protein expression, the distinct compartments of the heart (e.g. SAN, atria, AVN, and ventricles) possess different electrophysiological characteristics seen in the ECG. This heterogeneity in gap junction and ion channel expression in the different heart components is regulated through interplay of several activating and repressing transcription factors that bind to regulatory elements of these genes to activate or repress their activity.

Interestingly, using variations in ECG-parameters as intermediate phenotypes for conduction disease and arrhythmia susceptibility, a number of GWAS identified a high correlation between associated variants and noncoding regions nearby ion channels (e.g. SCN10A-SCN5A, KCNQ1, KCNH2) and transcription factors (e.g. TBX3-TBX5, NKX2-5, MEIS1). Abnormalities in ion channel function cause changes in the cardiac electrical impulse, resulting in ECG abnormalities and arrhythmias. Mutations have often, but not always, been identified in the protein-coding domains of ion channels that cause specific forms of heritable arrhythmogenic disorders in the structurally normal heart. In many cases, modulation of the phenotype through environmental or additional genetic factors will play a role in the exacerbation of the pre-existing phenotype. The observation that transcription factors modulate cardiac conduction implies the presence of a transcriptional network involved in regulation of the cardiac conduction system. Misregulation of this transcriptional network could offer an additional explanation for certain conduction disease phenotypes.
Among the most frequently implicated regions in these GWAS are the **TBX3-TBX5** locus and the **NKX2-5** locus. NKX2-5 is crucial for the development of the heart and remains broadly expressed in the adult heart.\(^{26}\) Perinatal loss of **Nkx2-5** in mice leads to rapid conduction and a loss of ventricular ion channel expression.\(^{27}\) **TBX3** and **TBX5** belong to the Tbx2-subfamily of T-box transcription factors and are involved in the development and maintenance of the cardiac conduction system (Figure 2). The expression of **Tbx3** is restricted to the central parts of the conduction system (SAN, AVN, AVB and proximal BB), where it acts as a repressor of ion channels like **Scn5a** and gap junction subunit-encoding genes like **Cx40** (**Gja5**) and **Cx43** (**Gja1**).\(^{28,29}\) During development, **TBX3** dictates a nodal phenotype through stimulation of the pacemaker gene program and repression of the working myocardial gene program in the primary myocardium. Heterozygous mutations in **TBX3** cause ulnar-mammary syndrome in humans,\(^{30,31}\) whereas loss of function of TBX3 is embryonic lethal in mice.\(^{28}\) Reduction of **TBX3** in the developing heart causes prolonged QRS duration and a spectrum of conduction defects, including sinus pauses, bradycardia, pre-excitation, atrioventricular

**Figure 2**

**TBX5** and **TBX3** regulate the expression of ion channels (**Scn5a**). **Tbx5**, together with other cardiac TFs, activates **Scn5a** (blue) in the heart. The expression pattern of **Scn5a** is broader than that of **Tbx5**, showing the additive effect of the other cardiac TFs. **Scn5a** is absent from the Tbx3-positive (green) and Tbx5-positive (pink) region of the sinus node (san), internodal region, atrioventricular node (avn) and atrioventricular canal myocardium. This demonstrates the dominant repressive function of Tbx3 in these parts of the heart. avb, atrioventricular bundle; l/ra, left/right atrium; l/rv, left/right ventricle.
block and an increased risk of sudden death.\textsuperscript{32} Loss of function of TBX5, a more broadly expressed transcriptional activator, results in severely disrupted heart development and embryonic lethality, whereas heterozygous loss of function in humans causes the Holt-Oram syndrome of heart-hand defects.\textsuperscript{26} Loss of \textit{Tbx5} in the mature murine VCS results in loss of \textit{Scn5a} and \textit{Gja5} in this domain. Subsequently, this leads to loss of fast conduction, cardiac arrhythmias and sudden cardiac death.\textsuperscript{33} Despite their adverse effects, TBX3 and TBX5 recognize the same TFBS and compete for the same regulatory elements.\textsuperscript{12} Genome-wide mapping of the binding profile for NKX2-5, TBX3 and TBX5 in murine hearts uncovered that many cardiac enhancers co-localize with ion channels repressed by TBX3.\textsuperscript{12} These findings, together with their presence among the highly associated regions in GWAS for PR interval and QRS duration, confirm the central role for these transcription factors in the transcriptional network regulating the genes that define the development and the function of the cardiac conduction system. The key to a properly functioning conduction system, therefore, lies in the balanced activities of TBX3, TBX5 and NKX2-5.\textsuperscript{12}

**ROLE OF SCN10A IN CARDIAC CONDUCTION DISEASE**

A meta-analysis of 14 GWAS on QRS duration in individuals of European descent implicated an intronic region of the \textit{SCN10A} gene as a major risk region for prolonged QRS duration.\textsuperscript{24} This finding was highly interesting, since \textit{SCN10A} had never been linked to cardiac conduction before. Functional follow-up studies revealed that \textit{SCN10A} is expressed in cardiomyocytes of the human ventricular conduction system and that loss of \textit{Scn10a} has an apparent effect on both PR interval and QRS duration in mice.\textsuperscript{21,24} However, its role in cardiac conduction disease remains to be established.\textsuperscript{34,35} Recent work has associated noncoding variants at \textit{SCN10A} with Brugada Syndrome, an arrhythmia disorder with a high risk of sudden cardiac death.\textsuperscript{26} Intriguingly, \textit{SCN10A} is located next to \textit{SCN5A}, which encodes the alpha-subunit of the cardiac voltage-gated Na\textsuperscript{+} channel. Mutations in \textit{SCN5A} are known to cause several types of heritable arrhythmogenic disorders, including Brugada Syndrome, in which loss-of-function mutations decrease the \textit{l\textsubscript{Na}} sodium current, and Long QT Syndrome (type 3), in which gain-of-function mutations increase the \textit{l\textsubscript{Na}} sodium current. It is therefore tempting to speculate that the presence of a putative regulatory element influenced by genetic variation impacts the expression of \textit{SCN5A} (Figure 3B). Considering the role of TBX3, TBX5 and NKX2-5 in the development and homeostasis of the conduction system and their link to QRS duration, it is very likely that these factors are involved in the regulation of this element (Figure 3A). Close examination of this risk region in \textit{SCN10A}-containing rs6801957, the sentinel SNP frequently found in ECG GWAS, reveals that it is occupied by TBX3, TBX5, NKX2-5 and several
enhancer-associated co-activators, such as p300 and PolII, in the adult mouse heart. The element is responsive to stimulation with GATA4, NKX2-5 and TBX5, and it can be repressed by TBX3 in vitro. In vivo reporter analysis showed the human ortholog of the enhancer is specifically active in the interventricular septum of the heart, the location where the proximal VCS develops. Analysis of the risk allele, with a frequency of 42% in individuals of European descent, shows the SNP alters a TFBS for TBX3/TBX5, which inhibits the response of the enhancer to these factors and decreases the activity of the enhancer in zebrafish in vivo (Figure 3C). Further investigation of this region will be necessary to determine the effect of the variant on transcription of SCN10A-SCN5A and to establish its contribution to conduction disease.

CONCLUSION AND FUTURE PERSPECTIVES
The rapidly growing number of genome-wide screening techniques have allowed for an enormous accumulation of new data on regulatory elements and genetic variation within the genome. The important task now at hand is to provide a mechanistic link between each of the regulatory elements and their associated phenotypic traits, explaining how genetic variation in these regulatory elements influences transcriptional regulation. Here, we have provided a plausible functional explanation for effects of variants in noncoding DNA on cardiac conduction and arrhythmias. In the case we highlight, the functional variant that was statistically associated with the phenotypic trait turned out to be the actual causal variant. However, since GWAS is highly dependent on linkage disequilibrium (LD), future cases may reveal that one or more SNPs in LD with the sentinel SNP are the origin of the causative effect. Possible approaches to overcome this problem include deep sequencing of regions surrounding the sentinel SNP or studying of different ethnic groups. Unfortunately, these approaches are costly and time-consuming. Taken that single variants only have minor effects, it is likely that many phenotypic traits are the result of multiple genetic variants in both coding and noncoding regions. Indeed, recent GWAS of Brugada patients revealed that common SNPs at SCN10A-SCN5A and HEY2 can have a strong cumulative effect on rare diseases susceptibility. Since the ultimate goal of these data is to translate them into clinical applications, many current research projects focus on the identification of coding genes regulated by genetic variants through in vitro and in vivo functional experiments, eQTL studies and exome sequencing. Recent advances in genome engineering by use of transcription activator-like effector nucleases (TALENs) and clustered regulatory interspaced short palindromic repeat (CRISPR) endonucleases are likely to accelerate the translation of GWAS into clinically relevant information, as these techniques can be used to observe the functional consequences of modifying risk regions picked up in GWAS in vivo. These efforts to pinpoint the
regulatory networks through which phenotypic variation arises can eventually aid in setting up a form of personalized medicine by improving diagnostic and prognostic strategies, as well as disease management and prediction. Taken together, these first examples provide a clear mechanistic insight into how a genetic variant can influence the transcriptional network involved in complex common diseases. Although there are still important topics to cover, such as comparative analysis of associated variants between populations of different ethnic origin and determining consequences of environmental effects on transcriptional regulation, these findings could ultimately contribute to personalized therapy.

**Figure 3**
Common variant in the Scn5a/Scn10a locus alters expression through disruption of a T-box binding element. (**A**) ChIP-seq tracks for TBX5 and TBX3 shows many binding events in the Scn5a/Scn10a locus, indicating these TFs play a role in the regulation of these genes. Two enhancers in this locus, indicated with blue and purple, recapitulate the endogenous expression pattern of Scn5a/Scn10a. (**B**) Cartoon depicting the putative role of the enhancer located in an Scn10a intron. The important role of Scn5a in cardiac conduction and the overlap in expression patterns between Scn5a and Scn10a, make it likely that both genes are being regulated by this enhancer. (**C**) Cartoon depicting the potential mechanism by which a variant in Scn10a alters gene expression. The major allele creates a T-box binding element that is recognized by TBX3 and TBX5 and drives expression in the atria, trabeculae and interventricular septum of the heart. The minor allele alters an important base in the recognition site, causing decreased affinity for TBX3 and TBX5. This results in loss of enhancer activity, which is likely to subsequently lead to loss of Scn5a/Scn10a transcription.
ACKNOWLEDGMENTS
We apologize to all researchers whose original articles were not directly cited. We thank Sameer Rana and Wino Wijnen for their contributions to this review. This work was supported by grants from the Netherlands Heart Foundation (NHS grant 2010B205) and by the European Community’s Seventh Frame work Programme contract (“CardioGeNet” 223463).
REFERENCES


