Genomics and epigenetics of atrial fibrillation

van Ouwerkerk, A.F.

Publication date
2021

Introduction

Atrial fibrillation (AF) is the most common cardiac arrhythmia. Showing increasing prevalence with age and being a major contributor to complications such as stroke, this disease represents a significant socio-economic burden.[1] Although co-morbidities have been shown to contribute to AF risk, the study of patient families left no doubt of the heritable component of AF.[2–4] Moreover, recent genome-wide association studies that were able to include increasing numbers of patients, have identified over 100 genetic loci in the general population that are associated with AF.[5] A number of these AF-associated loci were found close to important cardiac transcription factors, hinting at the possibility of the involvement of transcription factor (TF) regulatory networks.

TFs regulate the spatiotemporal expression patterns of target genes underlying structure and function. A TF recognizes motifs in regulatory genomic elements and binds them with a certain affinity depending on availability of the motif, the TF dose and interacting partners.[6] The chromatin containing the bound regulatory element (enhancer or repressor) folds in three-dimensional space to interact with target gene(s) to facilitate or hamper transcription.[7] Thus, TFs, in collaboration with other TFs and other co-factors, regulate transcriptional networks that underlie the correct development and homeostasis of tissues and organs such as the heart. Abnormal dosage during development of such TFs caused by for instance pathogenic DNA variations can lead to congenital heart defects or conduction abnormalities.[8] Similarly, common genetic variation has also been associated with conduction anomalies such as AF,[5] leading to the hypothesis that dosage variation of key cardiac TFs can significantly contribute to AF susceptibility.

The majority of variants or single nucleotide polymorphisms (SNPs) associated with AF are found in intergenic or intronic (i.e. non-coding) regions.[5] Moreover, to complicate matters, each locus contains a lead SNPs with the highest statistical association that is co-inherited due to genetic linkage (i.e. are in linkage disequilibrium) with large numbers of other common SNPs that display statistically significant associations as well.[5,9–13] This leads to the following question that dominates current research in the field of the genomics of AF; which SNP(s) is the causal variant, and how does it add to AF-susceptibility? Disease-associated variants have revealed themselves to be enriched in genomic regions with increased presence of epigenetic signatures including chromatin accessibility, histone modifications and DNA methylation, suggesting that the variants lie in regulatory elements.[14–17] This leads to the current hypothesis that disease-associated variants alter regulatory element function and consequentially variation in target gene expression predisposing to disease. Therefore, a goal that has arisen from this hypothesis is to find the target gene(s) of variant regulatory regions.

In this thesis, we investigate the genomics and epigenetics of atrial fibrillation in the general population and in a patient family. We prioritize variant regulatory elements and the candidate target genes of atrial fibrillation-associated loci. We describe the TF networks that result in altered target gene expression, and investigate in as far as possible how these changes can contribute to AF susceptibility. We dissect *in vivo* models of AF-associated regulatory element deletion and the transcriptional and ultimate electrophysiological effects that arise from the disrupted function of key TFs and their target genes. Finally, we make an in-depth study of an AF-mouse model that describes the role of a human pathogenic missense variant G125R in the TF TBX5 in increased AF susceptibility in a patient family.

The work we present here demonstrates the necessity of combining different molecular approaches to probe spatial and temporal cardiac gene expression if one is trying to understand the complexities of diseases such as AF. We envisage that identifying and linking regulatory variants to target gene expression will bring the field one step closer to the goal of providing effective therapies for complex diseases.

Scope

In **Chapter 1**, we describe the current state of research on epigenetics and transcriptional networks underlying the most common cardiac arrhythmia, AF. Genome-wide association studies have found over 100 loci associated with AF, many of which lie in loci harboring important cardiac transcription factors such as *PITX2*, *ZFHX3*, *PRRX1* and *TBX5*. We discuss how transcriptional networks involving these TFs may influence cardiomyocyte function resulting in increased AF susceptibility.

In **Chapter 2**, the AF-associated loci are examined in detail. The majority of the associated loci lie in non-coding regions of the genome, regions thought to compose or be part of transcription regulatory elements responsible for target gene expression. Using human transcriptomic, epigenomic and chromatin conformation datasets, we prioritize the candidate target genes as well as sub-threshold variants co-localizing with regulatory elements, for each locus, thus linking genetic variation and target gene regulation. Moreover, we describe the *in vivo* removal of the homologous mouse region of a prioritized distal regulatory elements involved in the regulation of *Gja1* (Cx43), *Kcnn3* and *Zfhx3* target gene expression.

In **Chapter 3**, we investigate 12 highly associated atrial fibrillation loci using self-transcribing active regulator region sequencing (STARRseq), identifying hundreds of potential regulatory elements of which 24 displayed allele-specific regulatory activity. Additionally, we describe the *in vivo* deletion of a mouse orthologue of a regulatory element containing non-coding atrial fibrillation-associated variants close to *Hcn4*, causing sinus node dysfunction and reduced gene expression.

In **Chapter 4**, we describe the *in vivo* modeling of missense pathogenic variant p.G125R of transcription factor TBX5 that was found in an extended pedigree of patients with atypical Holt-Oram syndrome. Using a variety of epigenetic and transcription profiling methods, we show that TBX5-p.G125R has altered DNA binding and interaction properties causing shifts in H3K27 acetylation, chromatin accessibility and transcription. We show evidence that these changes underlie the electrophysiological characteristics such as variable RR interval, atrial extra systoles and susceptibility to atrial fibrillation observed in mice and patients.

References

1. Staerk, L., Sherer, J. A., Ko, D., Benjamin, E. J. & Helm, R. H. Atrial Fibrillation: Epidemiology, Pathophysiology, Clinical Outcomes. *Circulation Research* (2017). doi:10.1161/CIRCRESAHA.117.309732
2. Fox, C. S. Parental Atrial Fibrillation as a Risk Factor for Atrial Fibrillation in Offspring. *Jama* **291**, 2851 (2004).
3. Lubitz, S. A. *et al.* Independent Susceptibility Markers for Atrial Fibrillation on Chromosome 4q25. *Circulation* **122**, 976–984 (2010).
4. Nattel, S. & Dobrev, D. Electrophysiological and molecular mechanisms of paroxysmal atrial fibrillation. *Nat. Rev. Cardiol.* **13**, 575–590 (2016).
5. Roselli, C. *et al.* Multi-ethnic genome-wide association study for atrial fibrillation. (2018).
6. de Laat, W. & Duboule, D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502**, 499–506 (2013).
7. Spitz, F. & Furlong, E. E. M. Transcription factors: From enhancer binding to developmental control. *Nature Reviews Genetics* (2012). doi:10.1038/nrg3207
8. McCulley, D. J. & Black, B. L. Transcription Factor Pathways and Congenital Heart Disease. in *Current Topics in Developmental Biology* (2012). doi:10.1016/B978-0-12-387786-4.00008-7
9. Benjamin, E. J. *et al.* Variants in ZFHX3 are associated with atrial fibrillation in individuals of European ancestry. *Nat. Genet.* **41**, 879–881 (2009).
10. Ellinor, P. T. *et al.* Common variants in KCNN3 are associated with lone atrial fibrillation. *Nat. Genet.* **42**, 240–244 (2010).
11. Ellinor, P. T. *et al.* Meta-analysis identifies six new susceptibility loci for atrial fibrillation. *Nat. Genet.* **44**, 670–675 (2012).
12. Sinner, M. F. *et al.* Integrating genetic, transcriptional, and functional analyses to identify five novel genes for atrial fibrillation. *Circulation* (2014). doi:10.1161/CIRCULATIONAHA.114.009892
13. Christophersen, I. E. *et al.* Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nat. Genet.* **49**, 946–952 (2017).
14. Wasserman, N. F., Aneas, I. & Nobrega, M. A. An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer. *Genome Res.* **20**, 1191–1197 (2010).
15. Maurano, M. T. *et al.* Systematic Localization of Common Disease-Associate Variation in Regulatorty DNA. *Science (80-.).* **337**, 1190–1195 (2012).
16. Pasquali, L. *et al.* Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* **46**, 136–143 (2014).
17. Grundberg, E. *et al.* Global analysis of dna methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. *Am. J. Hum. Genet.* (2013). doi:10.1016/j.ajhg.2013.10.004