The effects of sequence variants on cardiac function and disease

Rósa Björk Þórólfsdóttir

Thesis for the degree of Philosophiae Doctor

Supervisor:
Davíð O. Arnar

Advisors:
Unnur Þorsteinsdóttir, Hilma Hólm

Doctoral committee:
Daníel F. Guðbjartsson, Karl K. Andersen

September 2019

UNIVERSITY OF ICELAND
SCHOOL OF HEALTH SCIENCES
FACULTY OF MEDICINE

Áhrif erfðabreytileika á starfsemi hjartans og hjartasjúkdóma

Rósa Björk Þórólfsdóttir

Ritgerð til doktorsgráðu

Umsjónarkennari:
Davíð O. Arnar

Leiðbeinendur:
Unnur Þorsteinsdóttir, Hilma Hólm

Doktorsnefnd:
Daníel F. Guðbjartsson, Karl K. Andersen

September 2019

UNIVERSITY OF ICELAND
SCHOOL OF HEALTH SCIENCES
FACULTY OF MEDICINE

ISBN 978-9935-9476-4-2

Printing by Háskólaprent.

Reykjavik, Iceland 2019

Ágrip

Inngangur og markmið: Gáttatif er algengast viðvarandi hjartsláttartruflana en núverandi skilningur á flókinni meingerð sjúkdómsins er takmarkaður. Við upphaf þessarar vinnu höfðu víðtækar erfðamengisleitir sýnt fram á tengsl milli um 30 erfðabreytileika og gáttatifs. Flestir eru algengir og staðsettir utan útraða gena og tengjast því sjúkdómnum með óþekktum hætti. Nýlega var sýnt fram á að tveir sjaldgæfir erfðabreytileikar í útröðum vöðvaliðagenanna *MYH6* og *MYL4* auka áhættu á gáttatifi en slík tenging vöðvaliðunnar við hartsláttartruflanir þegar hjartavöðvakvilli var ekki til staðar var nýmæli. Tilgangur þessa doktorsverkefnis var að framkvæma víðtæka erfðamengisleit með það að markmiði að bera kennsl á áður óþekkt tengsl milli erfðabreytileika og gáttatifs. Jafnframt var markmiðið að öðlast innsýn í með hvaða hætti erfðabreytileikar hafa áhrif með því að kanna tengsl þeirra við aðra hjartasjúkdóma og hjartalínuritsbreytur. Í verkefninu skoðuðum við einnig áhrif *MYH6* breytileikans á aðra hjartasjúkdóma og reyndist hann einnig auka líkur á meðfæddum ósæðarþrengslum. Þess vegna var einnig framkvæmd víðtæk erfðamengisleit á þeim meðfædda hjartagalla. Megintilgangur þess að finna áður óþekkt tengsl milli erfðabreytileika og áhættu á gáttatifi er að auka skilning á meingerð sjúkdómsins og stuðla með því að framförum í meðhöndlun hans.

Aðferðir: Við framkvæmdum fjórar rannsóknir byggðar á víðtækri erfðamengisleit. Í öllum tilvikum byggðust erfðaupplýsingar Íslendinga á heilraðgreiningu 15.220 einstaklinga og örflögugreiningu 151.677 sem framkvæmdar voru hjá Íslenskri erfðagreiningu. Við könnuðum erfðabreytileika sem fundust og tengdust sjúkdómum nánar með því að meta áhrif þeirra á hjartalínuritsbreytur úr 289.297 hjartalínuritum frá 62.974 einstaklingum sem ekki höfðu greinst með gáttatif eða fengið gangráð. Þrjár rannsóknanna fjölluðu fyrst og fremst um gáttatif og sú fjórða um meðfædd ósæðarþrengsli: i) Víðtæk erfðamengisleit meðal 13.471 tilfella gáttatifs og 374.939 viðmiða frá Íslandi. Niðurstöðum var fylgt eftir í bandarískum gögnum. ii) Safnrannsókn víðtækra erfðamengisleita á gáttatifi meðal 29.502 tilfella og 767.760 viðmiða frá Íslandi og Bretlandi (UK Biobank). Niðurstöðum var fylgt eftir í sýnum frá Noregi og Bandaríkjunum. Áhrif tiltekinnar stökkbreytingar á umritun var metin í 167 RNA sýnum úr hægri gátt. iii) Safnrannsókn víðtækra erfðamengisleita meðal >1.000.000 þátttakenda frá

sex rannsóknarverkefnum, þ.m.t. 60.620 tilfelli gáttatifs. iv) Víðtæk erfðamengisleit á meðfæddum ósæðarþrengslum meðal 120 tilfella og 355.166 viðmiða frá Íslandi. Nákvæm svipgerðargreining bera tiltekins erfðabreytileika var framkvæmd.

Niðurstöður: Við fundum áður óþekkt tengsl milli gáttatifs og erfðabreytileika í útröðum þriggja gena sem öll gegna hlutverki í viðhaldi eðlilegrar byggingar hjartans; *PLEC, RPL3L* og *MYZAP.* Greining RNA sýna úr hjarta leiddi í ljós að erfðabreytileiki í *RPL3L* olli því að ein útröð gensins var ekki umrituð. Stór safnrannsókn meðal rúmlega milljón þátttakenda leiddi í ljós áður óþekkt tengsl á milli 80 erfðabreytileika og gáttatifs og þau gen sem þóttu líklegust til að útskýra tengslin reyndust hafa áhrif á marga ferla, m.a. byggingu hjartans. Erfðabreytileikar sem auka líkur á gáttatifi höfðu fjölbreytt áhrif á hjartalínuritsbreytur. Víðtæk erfðamengisleit á meðfæddum ósæðarþrengslum fann einungis mislestursstökkbreytinguna p.Arg721Trp í *MYH6* en hún ein útskýrir um 20% tilfella hjartagallans á Íslandi. Sami erfðabreytileiki hafði áður verið tengdur aukinni áhættu á sjúkum sínushnút og gáttatifi, auk þess að hafa hér víðtæk og sterk áhrif á hjartalínuritsbreytur.

Ályktanir: Niðurstöðurnar benda til mikilvægis byggingareininga hjartans í meingerð gáttatifs og styðja kenningar um að hartavöðvakvilli í gáttum hafi þýðingu í þróun sjúkdómsins. Erfðabreytileikar sem auka líkur gáttatifs höfðu fjölbreytt áhrif á rafleiðni í hjartanu sem bendir til að þeir auki líkur á gáttatifi með ólíkum hætti. Kortlagning á tengslum þeirra við hjartalínuritsbreytur gæti gagnast við frekari rannsóknir á með hvaða hætti einstakir erfðabreytileikar tengjast gáttatifi. Fjölbreytt og víðtæk áhrif p.Arg721Trp í *MYH6* á hjartasjúkdóma og starfsemi hjartans benda til að þessi breytileiki tengist gáttatifi ekki eingöngu gegnum sterk tengsl við sjúkan sínushnút. Tengsl meðfæddra ósæðarþrengsla við samdráttareiningu hjartans samræmast kenningum um að skert blóðflæði gegnum ósæðina á fósturstigi stuðli að myndun hjartagallans. Erfðabreytileikinn í *MYH6* útskýrir stóran hluta tilfella meðfæddra ósæðarþrengsla á Íslandi og gæti sú vitneskja komið að notum í erfðaráðgjöf. Samanlagt veita niðurstöðurnar aukinn skilning á meingerð bæði gáttatifs og meðfæddra ósæðarþrengsla, en slík innsýn er nauðsynlegt skref í átt að bættum meðferðarúrræðum.

Lykilorð:

Gáttatif, meðfædd ósæðarþrengsl, víðtæk erfðamengisleit, hjartalínuritsbreytur, byggingareiningar hjartans.

Abstract

Introduction and aims: Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia and its complex pathophysiology is incompletely understood. At the time of beginning of this work, genome-wide association studies (GWAS) on AF had yielded about 30 associations, mostly with common variants in the non-coding genome that affect AF risk through unknown mechanisms. Recently, rare coding variants in the sarcomere genes *MYH6* and *MYL4* associated with increased risk of AF, representing a novel implication of the sarcomere in arrhythmogenesis in the absence of cardiomyopathy. The aim of this doctoral research was to perform GWASs on AF and search for novel associations. Furthermore, to gain insights into the mechanisms by which novel and published variants affect AF risk by assessing their effects on other cardiovascular diseases and on normal cardiac conduction, using electrocardiogram (ECG) measurements. Through association analysis we observed that one of the strongest AF variants, in *MYH6,* accounts for 20% of Icelandic cases of coarctation of the aorta (CoA), thus we decided to explore this variant and its effects on structural heart disease in detail resulting in a separate paper. The overall purpose of identifying novel associations of sequence variants with AF is to further the understanding of underlying pathogenesis and facilitate advances in treatment.

Methods: We conducted four GWASs. In all cases, genotype information of Icelandic participants was based on whole-genome sequencing of 15,220 and chip-typing of 151,677 individuals performed at deCODE genetics. We tested AF variants for association with measurements in 289,297 sinus rhythm ECGs from 62,974 individuals. Three GWASs on AF and one on CoA were performed: i) A GWAS among 13,471 AF cases and 374,939 controls from Iceland with additional follow up in samples from the US. ii) A meta-analysis of GWAS on AF among 29,502 cases and 767,760 controls from Iceland and the UK Biobank with follow-up in samples from Norway and the US, focusing on low-frequency coding and splice variants. Analysis of 167 RNA samples from the right atrium were used to assess the effects of a novel AF variant at the transcript level. iii) A multi-center meta-analysis of GWASs on AF including >1,000,000 participants from six contributing cohorts, including 60,620 AF cases. iv) A GWAS on CoA among 120 cases and 355,166 controls in Iceland. Detailed phenotyping of carriers of a novel variant was performed.

Results: We identified novel AF associations with coding variants in three genes which have a role in maintaining cardiac structure; *PLEC*, *RPL3L* and *MYZAP*. Analysis of RNA samples from the right atrium revealed that a splice-donor variant in *RPL3L* results in exon skipping. Through a meta-analysis involving >1,000,000 participants from six contributing cohorts, 80 novel AF loci were identified. Pathway and functional enrichment analysis suggested variable mechanisms underlying the associations, including effects on structural remodeling. Assessing variants associations with ECG measurements showed diverse effects of AF variants on normal cardiac conduction. We found that one of the strongest AF variants, a missense variant, p.Arg721Trp, in the sarcomere gene *MYH6* was found to explain approximately 20% of CoA cases in Iceland. This variant was previously reported to associate with sick sinus syndrome through GWAS and has strong effects on AF risk and ECG measurements in the current study.

Conclusion: The results highlight the important role of myocardial structure in the pathogenesis of AF. Furthermore, they support speculations of a role for subclinical atrial cardiomyopathy in the development of arrhythmia. The diverse associations between AF variants and ECG measurements suggest fundamentally different categories of mechanisms contributing to the development of AF and might help inform on underlying biology of specific AF loci in future studies. We demonstrate a pleiotropic effect of p.Arg721Trp in *MYH6* on cardiac function and disease. The variant explains a large proportion of CoA cases in Iceland which may be a relevant fact for genetic counseling. The association with *MYH6*, that encodes a part of the cardiac contractile unit, complies with the hemodynamic theory of CoA development. Combined the results increase understanding of the pathophysiology of both AF and CoA, thus providing knowledge that will hopefully translate into advances in treatment.

Keywords:

Atrial fibrillation, coarctation of the aorta, genome-wide association study, electrocardiogram measurements, cardiac structural remodeling.

Acknowledgements

This doctoral study was conducted at deCODE genetics, Reykjavík, Iceland. Many have contributed to this work and I want to express my sincere gratitude to them all. In particular, I wish to mention the following people.

I thank my supervisors/advisors, Davíð O. Arnar, Unnur Þorsteinsdóttir and Hilma Hólm for their excellent scientific guidance and mentorship. I especially thank Hilma for believing in me and taking me on her team at deCODE genetics and for always responding well to any request for assistance or discussions. I also want to thank Davíð for the inspiring enthusiasm for science, encouragement and advice and I thank Unnur for her sharp-minded opinion, positive attitude and good advice.

I thank the two remaining members of my doctoral committee, Daníel F. Guðbjartsson and Karl K. Andersen, for their contribution to the thesis. I especially thank Daníel for statistical advice and instructive cooperation. I am very thankful to Karl for engaging me in scientific work as a medical student and providing good advice and support ever since.

I want to express my sincere gratitude to Kári Stefánsson, whose commendable contribution to the field of human genetics is undisputable. The resources and vision he has built up at deCODE genetics are unique and I am extremely grateful for having had the opportunity to be, if only, a small part of that project. I also thank Kári for his supervision of all studies included in this thesis, for always taking the time to consider every detail in pursuit of the best outcome and for kind words and encouragement.

I thank all the coauthors for their contribution to this work. Special thanks go to Garðar Sveinbjörnsson. His help and teaching when I was starting my work at deCODE have been invaluable and our cooperation since has been very enjoyable. I also want to thank Þorsteinn Björnsson for his hard work and great cooperation. Furthermore, I would like to give special thanks to cooperators who have contributed samples to the study, in particular I want to name Bjarni Torfason and Atli S. Valgarðsson who gathered and handled samples from cardiac atria during cardiothoracic surgery at Landspítali.

I had the fortune of being assigned an office with Þórhildur Júlíusdóttir. Not only have I benefitted from her willingness to share her bioinformatics skills and vast knowledge, but I also gained a dear friend.

I am very privileged to have been a part of the amazing team of scientists at deCODE genetics. Since my first day there I have been met with positive attitude and encouragement and many of the more experienced researchers have kindly and selflessly shared their time and knowledge with me. I thank my colleagues at deCODE genetics for the excellent teamwork, stimulating conversations on science (and other things) and most of all for making going to work always a matter of anticipation.

Participants in research projects at deCODE genetics have a very important role and deserve special thanks. Without their valuable contribution this scientific work would not have been possible.

I want to thank the following people for help with issues regarding writing the thesis, including for example reproduction of tables/images and proofreading: Garðar Sveinbjörnsson, Gísli H. Halldórsson, Hildur M. Ægisdóttir, Jonas B. Nielsen, Júlíus Kristjánsson, Margrét Baldursdóttir, Stefán Jónsson and Þórhildur Júlíusdóttir.

I thank Björgvin Ragnarsson at Háskólaprent for is valuable help in setting up and printing the thesis.

I am very fortunate to be surrounded by good friends and family members who continuously support and encourage me. To all of them I am very thankful. I especially thank my parents for their unconditional love and support and the good values and work ethic they have promoted. Finally, I thank my husband Júlíus, my greatest supporter and best friend, for always being there for me.

Contents

List of abbreviations

ACM	Arrhythmogenic right ventricular cardiomyopathy
AF	Atrial fibrillation
ASD	Atrial septal defect
ATP	Adenosine triphosphate
AVN	Atrioventricular node
CHD	Congenital heart defect
CI	95% confidence interval
CoA	Coarctation of the aorta
CVD	Cardiovascular disease
EBS	Epidermolysis bullosa simplex
ECG	Electrocardiogram
eQTL	Expression quantitative trait locus
FDR	False discovery rate
FOURIER	Further Cardiovascular Outcomes Research with PCSK9 Inhibition in Subjects with Elevated Risk
gnomAD	The Genome Aggregation Database
GTEx	Genotype-Tissue Expression
GWAS	Genome-wide association study
HLHS	Hypoplastic left heart syndrome
ICD	International Classification of Diseases
LD	Linkage-disequilibrium
LUH	Landspitali University Hospital
MGI	Michigan Genomics Initiative
OR	Odds ratio
PRM	Plakin repeat motif
SA node	Sinoatrial node
SD	Standard deviation

SNP	Single nucleotide polymorphism
SRF	Serum response factor
SSS	Sick sinus syndrome
TIA	Transient ischemic attack
VSD	Ventricular septal defect
αMHC	alpha-myosin heavy chain subunit
βMHC	beta-myosin heavy chain subunit

List of figures

List of tables

List of original papers

This thesis is based on the following original publications, which are referred to in the text by their Roman numerals:

Paper I. Thorolfsdottir, R. B., Sveinbjornsson, G., Sulem, P., Helgadottir, A., Gretarsdottir, S., Benonisdottir, S., Magnusdottir, A., Davidsson, O. B., Rajamani, S., Roden, D. M., Darbar, D., Pedersen, T. R., Sabatine, M. S., Jonsdottir, I., Arnar, D. O., Thorsteinsdottir, U., Gudbjartsson, D. F., Holm, H., & Stefansson, K. (2017). A Missense Variant in *PLEC* Increases Risk of Atrial Fibrillation. *J Am Coll Cardiol*, 70(17), 2157-2168.

Paper II. Thorolfsdottir, R. B., Sveinbjornsson, G., Sulem, P., Nielsen, J. B., Jonsson, S., Halldorsson, G. H., Melsted, P., Ivarsdottir, E. V., Davidsson, O. B., Kristjansson, R. P., Thorleifsson, G., Helgadottir, A., Gretarsdottir, S., Norddahl, G., Rajamani, S., Torfason, B., Valgardsson, A. S., Sverrisson, J. T., Tragante, V., Holmen, O. L., Asselbergs, F. W., Roden, D. M., Darbar, D., Pedersen, T. R., Sabatine, M. S., Willer, C. J., Lochen, M. L., Halldorsson, B. V., Jonsdottir, I., Hveem, K., Arnar, D. O., Thorsteinsdottir, U., Gudbjartsson, D. F., Holm, H., & Stefansson, K. (2018). Coding variants in *RPL3L* and *MYZAP* increase risk of atrial fibrillation. *Commun Biol*, 1, 68.

Paper III. Nielsen, J. B., Thorolfsdottir, R. B., Fritsche, L. G., Zhou, W., Skov, M. W., Graham, S. E., Herron, T. J., McCarthy, S., Schmidt, E. M., Sveinbjornsson, G., Surakka, I., Mathis, M. R., Yamazaki, M., Crawford, R. D., Gabrielsen, M. E., Skogholt, A. H., Holmen, O. L., Lin, M., Wolford, B. N., Dey, R., Dalen, H., Sulem, P., Chung, J. H., Backman, J. D., Arnar, D. O., Thorsteinsdottir, U., Baras, A., O'Dushlaine, C., Holst, A. G., Wen, X., Hornsby, W., Dewey, F. E., Boehnke, M., Kheterpal, S., Mukherjee, B., Lee, S., Kang, H. M., Holm, H., Kitzman, J., Shavit, J. A., Jalife, J., Brummett, C. M., Teslovich, T. M., Carey, D. J., Gudbjartsson, D. F., Stefansson, K., Abecasis, G. R., Hveem, K., & Willer, C. J. (2018). Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat Genet*, 50(9), 1234-1239.

Paper IV. Bjornsson, T., Thorolfsdottir, R. B., Sveinbjornsson, G., Sulem, P., Norddahl, G. L., Helgadottir, A., Gretarsdottir, S., Magnusdottir, A., Danielsen, R., Sigurdsson, E. L., Adalsteinsdottir, B., Gunnarsson, S. I., Jonsdottir, I., Arnar, D. O., Helgason, H., Gudbjartsson, T., Gudbjartsson, D. F., Thorsteinsdottir, U., Holm, H., & Stefansson, K. (2018). A rare missense mutation in *MYH6* associates with non-syndromic coarctation of the aorta. *Eur Heart J*, 39(34), 3243-3249.

Declaration of contribution

Paper I: I took part in planning this study using data from deCODE genetics phenotype/genotype database which has been established over the last two decades. I categorized and updated lists of ICD diagnoses codes and selected and extracted appropriate ECG measurements for individuals that are part of the deCODE database. I conducted statistical analysis and interpreted the results in cooperation with my supervisors and numerous colleagues at deCODE genetics. I wrote the paper and handled all revisions under the guidance of my supervisors.

Paper II: I took part in planning the study. I updated phenotypic data and conducted statistical analysis in cooperation with my supervisors and other colleagues at deCODE genetics. With regards to RNA sequencing analysis, I planned the analysis and prepared the graphical and written presentation of the results with the assistance of deCODE scientists specialized in the subject. I wrote the paper and handled revisions in collaboration with my supervisors.

Paper III: This study was a large meta-analysis in which six study groups collaborated and contributed data. I was involved in handling and mediating data from deCODE genetics. This involved in particular handling of new data on AF diagnoses and ECG measurements, as described for paper I and II. I participated in performing association analysis on AF and ECG data from deCODE. I wrote parts of the manuscript concerning the deCODE data and its analysis and was one of the first authors. Along with a group of colleagues from deCODE genetics I participated in the overall interpretation of results and revisions of the manuscript. In this thesis, emphasis will be placed on my main contribution to the study and other parts will be covered briefly or not covered. The pathway and functional enrichment analysis (using GREGOR and DEPICT) and genetic risk score calculations will be covered briefly. Although I did not perform these calculations, they are an important part of the overall interpretation of the GWAS results, in which I participated. Neither ischemic heart failure model in rabbits nor drug target analysis are covered in the thesis as I was not involved in performing or interpreting these analyses.

Paper IV: I took part in planning the study with my supervisors and conducted the work in collaboration with Þorsteinn Björnsson. In accordance with data protection laws there was a clear division of tasks with regards to handling of phenotype and genotype data so that the same person would not both review phenotype (not encrypted) and genotype (encrypted) data. ÞB

handled unencrypted phenotype data (detailed reviewing of medical records) and I handled the genotype data and encrypted phenotype data (including ICD diagnoses codes, ECG and echocardiogram data). We wrote the paper together as shared first authors and I participated in all revisions.

1 Introduction

1.1 General introduction

The fundamental aim of human genetic research is to explore how variation in the genome contributes to human diversity, including risk of disease. The ultimate hope is that the resulting knowledge will facilitate discoveries of therapeutic targets and aid in personalizing diagnosis and treatment. Technological advances in the past few decades, including the advent of high-throughput genotyping and sequencing, has provided extensive information on human sequence diversity (Frazer et al., 2007). This has enormously widened the horizon of genetic studies both for rare mendelian like disease and common disorders with complex pattern of inheritance (Hindorff et al., 2009). A unique opportunity to conduct this manner of scientific research is provided by the deCODE genetics database which combines genotype information for a large proportion of Icelanders with well documented genealogy and high quality registration of phenotypic data (Gudbjartsson et al., 2015). Cardiovascular diseases (CVDs), the subject of this thesis, encompass a diverse group of disorders which are a leading cause of morbidity and mortality worldwide (Benjamin et al., 2019). Many CVDs have a large heritable component and some share genetic risk factors (Kathiresan & Srivastava, 2012), including the two CVDs that are the focus of this thesis, the common cardiac arrhythmia atrial fibrillation (AF) and the congenital heart defect (CHD) coarctation of the aorta (CoA).

1.2 Genetic research on human traits

1.2.1 Genetic variation

Genetic variation is defined as difference in DNA sequences between individuals within a population. Its main sources are mutations and meiotic recombination (Strachan & Read, 2011). The variation ranges from single nucleotide polymorphisms (SNPs) to small insertions or deletions (INDELs) spanning two to hundreds of base pairs and finally structural variants (Iafrate et al., 2004; Strachan & Read, 2011). The structural variants involve larger insertions or deletions or chromosomal rearrangements and are often referred to as copy number variants (Freeman et al., 2006). The frequency of genetic variation and its effect on biological functions and human traits varies greatly, ranging from having no effects to being incompatible with life.

Variants located in the coding regions of the genome, i.e. exons of genes, affect protein function to a different degree. Exonic variants represent only 1-2% of the variants found in the human genome (Auton et al., 2015; McCarthy et al., 2016). Some of them are synonymous, meaning they do not affect the amino acid sequence and have no effect on protein function. Other variants directly affect the protein product, including nonsense variants that introduce a premature stop codon resulting in either a truncated protein or absence of protein through the process of nonsense mediated decay and missense variants that result in the substitution of one amino acid for another (Strachan & Read, 2011). Variants that have severe effects on protein function and result in serious pathophysiology are usually subject to negative selection and are therefore rare (Reich & Lander, 2001).

In general, interpreting the functional effects of sequence variants, including coding variants, can be challenging (Shameer et al., 2016). Algorithms predicting the effects of amino acid changes on protein function are commonly applied (Adzhubei et al., 2013; Sim et al., 2012). They are based on factors such as protein conservation and chemical difference between amino acids. Importantly, they involve predictions, not functional validation of the effects. Even more challenging is the annotation of the non-coding genome, i.e. in introns or intergenic regions containing regulatory elements that can influence biological functions such as gene expression and transcript isoform repertoire (Cheung & Spielman, 2009; Pagani & Baralle, 2004). These regulatory mechanisms are not well documented and often complex. For example, the same variant can have diverse effects on gene expression in different tissues and developmental stages. The interpretation of functional effects of non-coding sequence variants is therefore a major challenge (Ritchie et al., 2014).

1.2.2 Linkage analysis and candidate gene methods

Earliest studies on the genetics of human disease applied the approaches of linkage analysis and candidate gene studies. Linkage analysis is a statistical method based on the observation that variants that reside physically close on a chromosome remain linked during meiosis. Therefore, identifying variants that show correlated segregation (linkage) with a trait through families can identify the chromosomal locus harboring the disease gene (Altshuler et al., 2008). Linkage analysis has thus proven a successful method to identify causative genes in mendelian disorders where highly penetrant variants segregate with disease status (Jimenez-Sanchez et al., 2001). However, the method has shown limited success when applied to common traits, most

notably because linkage studies lack power to identify common variants as they have low to moderate effects on disease risk (Risch & Merikangas, 1996). Candidate gene studies are based on searching for risk variants through association analysis in specific biologically plausible genes (Tabor et al., 2002). The results of candidate gene studies on common traits often lack replication and the method has been criticized for being hypothesis-driven and therefore only able to identify a fraction of genetic risk factors (Hirschhorn & Daly, 2005; Tabor et al., 2002).

1.2.3 Genome-wide association studies

In the past decade genome-wide association studies (GWASs) have become the most applied approach in research on the genetics of common traits with complex pattern of inheritance. GWAS is an unbiased and comprehensive method that tests for association between sequence variants throughout the genome and a trait (Hirschhorn & Gajdos, 2011). In contrast to candidate gene studies, no assumptions are made beforehand on which genes or variants may affect the trait in question. The earliest GWASs tested common SNPs on SNP chip arrays and resulted in multiple discoveries of common variants with mild or moderate effect sizes affecting a wide range of traits. The discoveries of such variants have increased in number with increasing sample sizes of GWASs (Hindorff et al., 2009). In recent years, advances in genotyping techniques have made large-scale whole-genome sequencing possible, allowing nearly complete coverage of the sequence diversity of populations under study (Frazer et al., 2007). These advances have facilitated the transition to GWASs that include rare variants identified by whole-genome sequencing, including, in particular, coding variants (Gudbjartsson et al., 2015).

Several additional concepts are important for the implementation of GWASs. First, when testing millions of variants for association with a trait, correction for multiple hypothesis testing is necessary. A widely used *P* value threshold of 5×10^{-8} is equivalent to a Bonferroni correction for one million independent tests considered (Auton et al., 2015). With extended coverage of sequence diversity and inclusion of rare sequence variants, the multiple-testing burden increases. Therefore, Sveinbjornsson et al. have suggested a method using as weights the enrichment of sequence annotations among association signals when determining significance thresholds. For example, an intergenic variant is less likely to be causal than a missense variant and therefore its association requires a lower *P* value to be declared genome-wide significant (Sveinbjornsson et al., 2016).

Another important term with regards to GWAS is linkage disequilibrium (LD). It refers to the non-random association of alleles at two (or more) loci after recombination in meiosis. Variants can therefore be inherited together in LD-blocks, also termed haplotypes (Wall & Pritchard, 2003). GWASs involve the application of statistical methods that use knowledge on LD between variants and the haplotype blocks in populations under study to predict or impute genotypes that are not directly assayed in a sample of individuals (Marchini et al., 2007). Thus, variants identified by whole-genome sequencing and not found on chip arrays can therefore be imputed (inferred) for individuals that have been genotyped on SNP chip array platforms.

Performing GWASs in the Icelandic population has several advantages. For the past two decades a large part of the Icelandic nation has been genotyped at deCODE genetics, around 160,000 individuals. The extended sampling and the well documented genealogy of the Icelandic population has allowed for assigning genotypes to parent of origin by long-range phasing (Kong et al., 2008). This has increased imputation accuracy, allowing for accurate imputation of variants identified through whole-genome sequencing down to an allele frequency of 0.1% (Gudbjartsson et al., 2015). The well documented genealogy further allows the incorporation of close relatives of genotyped individuals into the analysis. Furthermore, the small size of the population can be an advantage for studying the effects of rare deleterious variants. Such variants can reach higher frequencies in small, more isolated, populations where a small number of ancestors account for a relatively large proportion of the population, termed the founder effect. Finally, a tradition for high quality registration of phenotypic data in Iceland is a crucial factor for identifying sequence variants associating with disease risk (Gudbjartsson et al., 2015).

1.3 Atrial fibrillation

1.3.1 Definition, presentation and complications

AF is the most common sustained cardiac arrhythmia and a significant cause of morbidity and mortality (Benjamin et al., 1998; Go et al., 2001). It is characterized by an irregular and often rapid heartbeat caused by uncoordinated electrical activation of the atria, the upper chambers of the heart (Fuster et al., 2001). In contrast, normal cardiac conduction is characterized by regular initiation of electrical activation by the sinoatrial (SA) node, the hearts natural pacemaker, followed by synchronized atrial and subsequently ventricular activation (Wakili et al., 2011). Figure 1 shows the

difference between depolarization pathways in normal sinus rhythm (Figure 1A) and AF (Figure 1B). The figure also shows how AF is reflected in the electrocardiogram (ECG), a recording of cardiac electrical activity over time and a diagnostic tool for AF and other arrhythmias. In AF the P waves, representing synchronized atrial depolarization, are replaced by an undulating baseline, reflecting the continuous, rapid and unsynchronized atrial activation. The resulting irregular ventricular activation is represented by irregularly occurring QRS complexes on the ECG (Figure 1) (Wakili et al., 2011).

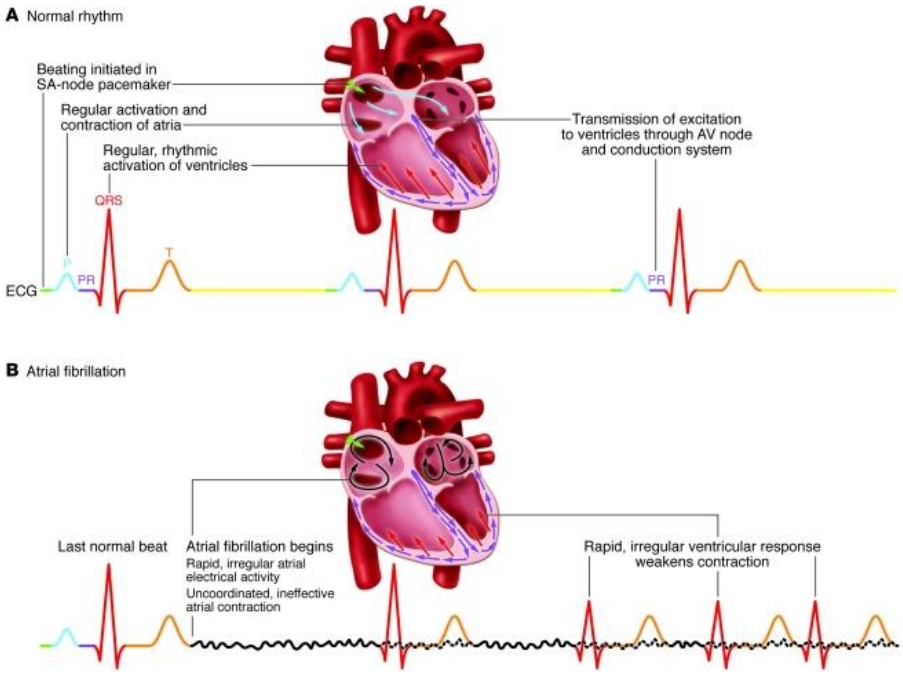


Figure 1. (a) A schematic figure of major events in cardiac conduction in normal sinus rhythm and their reflection on an ECG recording. (b) A schematic figure of rapid and irregular atrial activation in AF and its reflection on the ECG. Reused from Wakili et al., 2011 with permission from the publisher and Copyright Clearance Center.

AF usually presents with symptoms that include palpitations, dyspnea, chest pain, fatigue and lightheadedness but can also be asymptomatic (Fuster et al., 2001). Furthermore, the altered hemodynamics and impaired cardiac mechanics can have serious consequences, most notably stroke and heart failure (Calenda et al., 2016; Stewart et al., 2002). The increased stroke risk in AF is likely the result of a combination of factors that contribute to a prothrombic state in this arrhythmia (Watson et al., 2009). One of the key factors is blood stagnation or stasis due to ineffective atrial contraction (Koizumi et al., 2015) but there is also evidence for the role of additional factors, including endocardial dysfunction (Cai et al., 2002), platelet activation (Akar et al., 2008) and possibly even inflammation and atrial fibrosis (Daccarett et al., 2011; Harada et al., 2015; Watson et al., 2009). AF causes heart failure through shorter diastolic filling time, loss of atrial contractile function and elevated filling pressures leading to diminished cardiac output. Tachycardia induced myocardial dysfunction, as a consequence of AF, can also contribute to heart failure development (Houmsse et al., 2011).

1.3.2 Epidemiology, risk factors and classification

Recent studies have estimated the lifetime risk for AF in individuals of European ancestry to be approximately 1 in 3 (Magnussen et al., 2017; Weng et al., 2018). AF is more common in men and age is the most powerful predictor of incident AF, the cumulative incidence increases markedly after 50 in men and 60 in women (Magnussen et al., 2017). It has been proposed that more than half of the AF burden is attributed to modifiable risk factors, the strongest one being hypertension (Huxley et al., 2011). Other established risk factors include smoking (Zhu et al., 2016), heart failure (Maisel & Stevenson, 2003), obesity (Asad et al., 2018), diabetes (Aune et al., 2018), alcohol consumption (Larsson et al., 2014) and obstructive sleep apnea (Tung et al., 2017).

In addition to causing substantial morbidity and impairment in quality of life (Thrall et al., 2006), AF is an independent predictor of mortality (Benjamin et al., 1998). Furthermore, it is a major source of health care costs (Stewart et al., 2004). It was estimated that AF affected at least 33 million people worldwide in the year 2010 (Chugh et al., 2014). This number is expected to increase substantially in coming decades due to population ageing, making AF a growing public health concern (Chugh et al., 2014; Miyasaka et al., 2006).

1.3.3 Pathophysiology

The molecular pathophysiology of AF is complex and not completely understood (Andrade et al., 2014; Wakili et al., 2011). As shown in Figure 1, normal cardiac conduction is characterized by regular electrical activation by the SA node and subsequently the electrical impulse travels uniformly through well defined conduction pathways in the heart. The electrical activation, or action potential, is created by a sequence of ion fluxes through specialized channels in the membrane (sarcolemma) of cardiomyocytes. In the healthy state, atrial cells have a negative intracellular membrane potential (resting potential) untill they are fired by the SA node pacemaking system and the membrane potential of the cell is depolarized to a positive value. This triggers calcium release from the sarcoplasmic reticulum which in turn triggers cardiac contraction. The cell is then repolarized to the resting potential before the next depolariziation wave reaches it. AF can be initiated and maintained by two fundamental mechanisms that interrupt this process, spontaneous ectopic firing of atrial cells and impulse reentry through atrial tissue. Ectopic firing occurs when cells outside the SA node are depolarized to threshold potential before a normal sinus beat reaches them. Impulse reentry refers to circular movement of the depolarizing wave, as opposed to normal unidirectional conduction. It can occur as one primary circuit (single-circuit reentry) or in multiple locations in the heart simultaneously (multiple-circuit reentry) (Iwasaki et al., 2011; Wakili et al., 2011).

The factors thought to contribute to ectopic firing and reentry are summarized in Figure 2. For example, abnormal intracellular calcium handling can cause ectopic firing. Abnormal leak of calcium from the sarcoplasmic reticulum during repolarization, can be caused by calcium overload or defective calcium channels. It leads to excessive positive intracellular potential before the next normal conduction from the SA node reaches the cell. If the positive charge is large enough to reach the threshold for depolarization it causes ectopic atrial firing (Wakili et al., 2011). Ectopic firing often arises in the pulmonary veins (Haissaguerre et al., 1998). In many cases, atrial tissue with poorly developed cell to cell connections extends up to the pulmonary veins (Sánchez-Quintana et al., 2012).

As shown in Figure 2, reentry is usually initiated by an ectopic beat and maintained in atrial tissue that has been subject to structural or electrical remodeling. Electrical remodeling refers to altered expression and/or function of cardiac ion channels (Nattel et al., 2008). The reentry circuit usually forms around unexcitable myocardial tissue, such as necrotic myocardium or a

scar. The reentry is sustained if components of the circuit have variable conduction velocity and/or refractoriness. Refractoriness refers to the period where a cardiomyocyte is inexcitable following electrical activation. As shown in Figure 2, structural remodeling can maintain reentry because it alters conduction velocities. Furthermore, ion channel dysfunction can affect the refractoriness, for example by shortening action potential duration which in turn shortens the refractory period (Wakili et al., 2011).

Each of the contributing mechanisms summarized in Figure 2 have been implicated through associations of risk factors with AF. For example, heart failure is associated with structural and electrical remodeling (Maisel & Stevenson, 2003) as well as interrupting intracellular calcium handling (Yeh et al., 2008) and both hyperthyroidism and alcohol consumption affect autonomic neural control, thus possibly affecting AF risk by increasing automaticity (Chen et al., 2002; Mandyam et al., 2012). In most cases the development of AF involves a complex interaction between ectopic firing, reentry and factors promoting them. Furthermore, AF itself begets AF, as the mechanisms known to contribute to AF also occur as a result of AF-induced structural and electrical remodeling (Schotten et al., 2011).

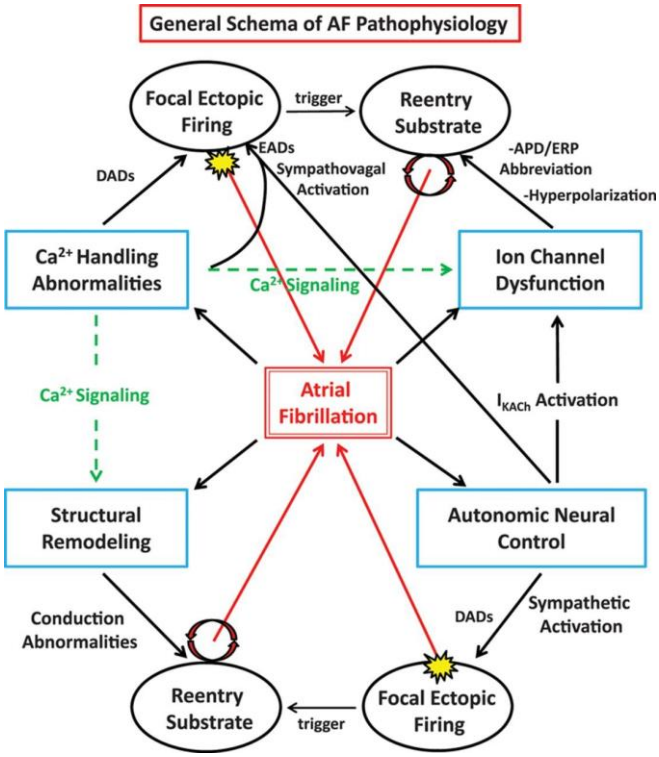


Figure 2. A schematic figure of mechanisms contributing to the development of AF. APD: action potential duration; DAD: delayed afterdepolarization; EAD: early afterdepolarization, ERP: effective refractory period. Reused from Andrade et al., 2014 with permission from the publisher and Copyright Clearance Center.

1.3.4 Classification

AF has traditionally been classified based on whether or not it occurs with identifiable etiology, such as hypertensive or valvular heart disease. In that respect, lone AF refers to AF occurring in the absence of risk factors in patients with an apparently structurally normal heart (Markides & Schilling, 2003). However, there is inconsistency in the use of this term since it is dependent on changing knowledge on AF pathogenesis and the currently limited approach to evaluating cardiac structure (Wyse et al., 2014). A classification based on temporal pattern is also commonly applied. Thus, AF can be categorized as paroxysmal episodes that terminate spontaneously, persistent AF that requires intervention for termination and permanent if not successfully terminated or if lasting over one year, independent of attempted treatment (Fuster et al., 2001). It is believed that the natural history of AF often involves evolution from paroxysmal to persistent to permanent form in accordance with the “AF begets AF” theory (de Vos et al., 2010; Nattel, 2002).

1.3.5 Treatment

The main goals of AF management are restoration of sinus rhythm, stroke prevention, ventricular rate control and relief of symptoms. An individualized approach should be undertaken, considering the onset and pattern of AF as well as the patient's comorbidities and symptoms (January et al., 2019; Kirchhof et al., 2016). Rate and rhythm control are applied to maintain left ventricular function and improve AF associated symptoms (Kirchhof, 2017). Pharmacologic rate control, most commonly with beta-adrenergic or calcium channel blockers, is appropriate for most patients with permanent AF (Van Gelder et al., 2016).

Rhythm control should be considered in most individuals presenting with AF and can also be appropriate for patients with a persistent arrhythmia whose symptoms remain despite rate control. The drugs used for rhythm control are most commonly beta-adrenergic blockers or class Ic/class III antiarrhythmic drugs. Their use is associated with considerable risk of adverse events, including ventricular arrhythmia (Piccini & Fauchier, 2016). Catheter ablation of pulmonary vein AF triggers has not been shown to reduce mortality but is frequently indicated for symptom reduction in patients with medically resistant AF, and to reduce the burden of the arrhythmia. It is a particularly effective procedure in younger patients (<65 years) with paroxysmal AF and no structural heart disease (Piccini & Fauchier, 2016; Walfridsson et al., 2015). Finally, options such as atrioventricular nodal (AVN) ablation and pacemaker implantation as well as surgical intervention may be considered for carefully selected patients with persistent symptoms (Kirchhof et al., 2016).

Stroke prevention with oral anticoagulants remains a cornerstone of AF management (Freedman et al., 2016; Kirchhof et al., 2016). Warfarin has traditionally been the most commonly used oral anticoagulant but is now being increasingly replaced by non-vitamin K antagonists which do not require as frequent monitoring (Olesen et al., 2015). Not all individuals with AF carry the same risk of thromboembolism and specific criteria are used to determine who needs anticoagulation. The most commonly used risk score is the CHA_2DS_2-VASc-score that evaluates stroke risk based on age, comorbidities and previous history of stroke (Freedman et al., 2016). However, risk scores do not incorporate direct evaluation of the prothrombotic factors identified in AF. Many of these factors, including platelet and endothelial activation in the atria, are difficult to assess in routine clinical practice (Akar et al., 2008). Risk scores have a somewhat limited predictive ability and a more thorough clinical evaluation has been suggested

for an improved risk stratification (Goldberger et al., 2015; Graves et al., 2018).

In general, the main aspects of AF management have remained unchanged for decades and currently available treatment options have limited efficacy and substantial adverse effects (Dobrev & Nattel, 2010). In light of the inadequacies of currently available management and the growing global burden of AF, better knowledge on its complex pathophysiology is of critical importance. In recent years, exploring the genetics of AF has been increasingly applied for that purpose.

1.4 Genetics of AF

1.4.1 Evidence for a heritable component in AF

Evidence of a genetic contribution to AF first emerged in the 1940s, when Wolff et.al. reported a family with an autosomal dominant form of AF (Wolf, 1943). Since then, many epidemiological studies have revealed a heritable component in AF in the general population, whether the focus is on lone AF or not (Arnar et al., 2006; Ellinor et al., 2005; Fox et al., 2004; Lubitz et al., 2010; Oyen et al., 2012; Yang et al., 2010). For example, an Icelandic study on over 5,000 AF cases and their relatives, showed that having a first-degree relative with AF conferred a 77% increased risk of developing the disease. This proportion decreased with increased genetic distance, being 5% among those with a fifth degree relative with AF (Arnar et al., 2006). The heritable component in AF is dependent on age, as younger age at onset in relatives confers higher risk of developing AF (Arnar et al., 2006; Fox et al., 2004; Lubitz et al., 2010). Furthermore, studies on Danish twins have identified both higher incidence of AF (Christophersen et al., 2009) and increased mortality (Christophersen et al., 2013) among individuals who had a co-twin with AF compared to twins without familial AF.

1.4.2 Linkage analysis and candidate gene studies on AF

The first studies on AF genetics focused on identifying genetic loci and sequence variants causing rare familial cases of AF, using linkage and candidate gene approaches. The linkage studies identified several AF loci (Brugada et al., 1997; Ellinor et al., 2003) and in 2003 the first gene, *KCNQ1*, was linked to AF (Chen et al., 2003). *KCNQ1* encodes the delayed-rectifier cardiac potassium channel (IKs) and AF associated mutations in the gene are predicted to reduce action potential duration and effective refractory period in atrial myocytes (Chen et al., 2003; El Harchi et al., 2010). This

complies with the suggested pathophysiological model of reduced atrial refractory period serving as a substrate for reentrant arrhythmia (Figure 2). The implication of *KCNQ1* mutations in familial AF motivated investigators to perform candidate gene studies, searching for causative variants in other genes encoding ion channels and gap junction proteins, identifying numerous mutations in number of genes. Subsequently candidate gene and linkage studies implicated non-ion channel genes in AF, including genes encoding signaling molecules, structural proteins and transcription factors (Bapat et al., 2018; Christophersen & Ellinor, 2016; Darbar & Roden, 2013).

The number of rare coding variants identified in candidate gene and linkage studies on AF has now reached a few hundreds. Most of these variants have not been replicated in other studies. If they truly cause AF they are therefore mostly unique to specific families and likely explain only a small proportion of AF cases (Christophersen & Ellinor, 2016; Tabor et al., 2002). Although the variants themselves have not been replicated, many of them are located in genes within AF GWAS loci. However, some of this overlap is explained by the fact that genes are frequently chosen for candidate gene studies based on implication in GWASs and many candidate gene studies explore large panels of genes with known cardiovascular functions (Macri et al., 2014; Savio-Galimberti et al., 2014). AF in the general population is now considered to be a highly heterogeneous disorder arising from a complex interaction of environmental and genetic factors and the genetic factors are believed to include the combined effects of numerous rare and common variants (Darbar & Roden, 2013).

1.4.3 Genome-wide association studies on AF

The view of AF as a complex polygenic disorder has arisen from the results of GWASs on AF conducted over the last decade. At the time of publication of the first study included in this thesis, these studies had reported approximately 30 loci associating with AF as a primary or secondary trait (Benjamin et al., 2009; Christophersen et al., 2017; den Hoed et al., 2013; Ellinor et al., 2012; Ellinor et al., 2010; Gudbjartsson et al., 2007; Gudbjartsson et al., 2009; Gudbjartsson et al., 2017; Holm et al., 2010; Holm et al., 2011; Low et al., 2017; Lubitz et al., 2014; Pfeufer et al., 2010; Sinner et al., 2014). Most of the previously identified loci are represented by common variants with mild or moderate effects on AF risk. Furthermore, most of them are in the non-coding genome, thus not directly affecting the protein coding sequence of a specific gene (Bapat et al., 2018). It can therefore be challenging to identify the biological effect of the associated variants and their

target genes, as applies to most common variants identified in GWASs. In some instances the underlying biology is completely unknown. However, for many of the loci possible candidate genes have been suggested, implicating biological pathways in AF development. In the first paper included in this thesis, all previously published associations from GWASs on AF are listed and looked up in the deCODE data.

The first association, reported by deCODE genetics in the year 2007, was represented by common intergenic variants at chromosome 4q25 (Gudbjartsson et al., 2007). Since then, this has consistently been the strongest AF associated locus in subsequent GWASs and widely replicated across ethnic groups (Christophersen et al., 2017; Delaney et al., 2012; Shi et al., 2009). The mechanism behind AF susceptibility at 4q25 is thought to be mediated by the closest gene, *PITX2* or its cardiac isoform *PITX2c*. This gene is a plausible candidate based on its biological functions and reports of AF risk variants affecting its expression in model organism (Ye et al., 2016). Pitx2c is a homeobox transcription factor critical for early cardiac development. It has a role in controlling cardiac left-to-right asymmetry (Piedra et al., 1998) and inhibits inappropriate left sided pacemaker specification (Wang et al., 2010). Furthermore, Pitx2c is required for the formation of the pulmonary myocardium (Mommersteeg et al., 2007), which is notable for the fact that ectopic triggers for AF frequently originate in the pulmonary veins (Haissaguerre et al., 1998) and are the target of ablation therapy for AF (Walfridsson et al., 2015). In fact, another AF-implicated gene from GWAS, *PRRX1*, is also a homeobox transcription factor required for pulmonary vascular development (Ellinor et al., 2012; Ihida-Stansbury et al., 2004). However, the mechanism by which non-coding variants at the *PITX2*- and *PRRX1*-loci affect AF risk has not been determined. Through its expressional regulation of, for example, ion transport and intercalated disc genes, *PITX2* is involved in several cardiac mechanisms that might contribute to AF development (Tao et al., 2014).

The importance of cardiac development in the pathogenesis of AF has been implicated by several additional associations with variants in or close to transcription factor genes, such as *ZFHX3*, *SOX5*, and *TBX5* (Benjamin et al., 2009; Christophersen et al., 2017; Gudbjartsson et al., 2009). The *ZFHX3* gene is required for striated muscle and neural development (Berry et al., 2001; Jung et al., 2005). Interestingly, *ZFHX3* and *PITX2* are connected, as *ZFHX3* affects the expression of the *POU1F1* gene which modulates transcriptional activity of *PITX2* (Qi et al., 2008). Several of the reported AF loci are located in or close to genes encoding ion channel proteins. For

example, *KCNN2* and *KCNN3* encode the calcium-dependent potassium channels SK2 and SK3 respectively, which are known to form heteromeric channel complexes with each other (Tuteja et al., 2010). They are involved in the maintenance of atrial action potential (Li et al., 2009; Tuteja et al., 2010), the shortening of which promotes reentry, as has been previously discussed (Figure 2) (Wakili et al., 2011). Other ion channel genes implicated in AF through GWAS include *HCN4, KCNJ5* and *SCNA10* (Christophersen et al., 2017).

Some of the AF loci point to genes with a structural role in the heart. For example, *CAV1* (Ellinor et al., 2012) encodes caveolin-1 which possibly prevents atrial fibrosis. Myozenin-1, encoded by *MYOZ1* (Ellinor et al., 2012) is involved in stabilizing the sarcomere (Frey et al., 2008) and nesprin-2, encoded by *SYNE2* (Ellinor et al., 2012), participates in anchoring the nuclear lamina to the peripheral cytoskeleton (Stewart-Hutchinson et al., 2008). All of the above examples are AF loci represented by non-coding variants and speculations are involved in the theories on underlying mechanisms. Furthermore, several reported AF associations from GWASs, such as the ones with intronic variants in *C9orf3* and *SH3PXD2A* do not point to biological plausible candidate genes or mechanisms (Christophersen et al., 2017).

The large-scale whole-genome sequencing of the Icelandic population conducted at deCODE genetics, has facilitated recent discoveries of rare coding variants associating with risk of AF through GWASs. These variants are located in, and directly affect, the protein products of the sarcomere genes *MYH6* (Holm et al., 2011) and *MYL4* (Gudbjartsson et al., 2017). Their discoveries implicate a part of the cardiac contractile system in the development of arrhythmias in the absence of apparent cardiomyopathy, which is a novel concept. Furthermore, their discovery is in line with the notion that uncommon variants with large effect sizes might contribute to the heritability of complex traits such as AF (Manolio et al., 2009).

1.4.4 Electrocardiogram measurements and the genetics of arrhythmia

The ECG recording of cardiac electrical activity over time is widely used in clinical practice and is a key diagnostic tool for AF and other arrhythmias. It is extremely useful to evaluate the function of the conduction system as well as the health and structure of the myocardium (Kadish et al., 2001). ECG parameters represent events in the cardiac cycle and some of them have prognostic value for the development of arrhythmia and mortality, including

heart rate, the QRS complex, QT interval and PR interval (Cheng et al., 2009; Desai et al., 2006). The PR interval and P wave indices are of special interest with regards to AF. The PR interval reflects the time required for the electrical impulse to travel through the atria and the AVN to the ventricles (Saksena & Camm, 2012). Both its prolongation and shortening have been associated with increased risk of AF (Cheng et al., 2009; Nielsen et al., 2013). P wave duration, which quantifies the time of atrial depolarization, has also been shown to have a U-shaped relationship with incident AF (Nielsen et al., 2015), as both its prolongation and shortening increase risk of AF. P wave indices are reflective of the conduction in the atria specifically, as they characterize atrial electrical activity during depolarization without assessing the delay through the AVN (Magnani et al., 2009).

Due to their close association with arrhythmia development, ECG measurements have been used in genetic studies as an endophenotype for identification of sequence variants that affect both electrophysiological function and the risk of arrhythmia (den Hoed et al., 2013; Holm et al., 2010; Pfeufer et al., 2010). The term endophenotype, originates from the field of psychiatric genetics and refers to a quantitative biological trait that is closely related to a specific phenotype (Gottesman & Shields, 1967). The idea is that the endophenotype is measurable independent of disease status and reflects a more direct expression of genetic effects than the phenotype itself, thus improving effect sizes and facilitating the search for causal genes (Hall & Smoller, 2010).

Indeed, the use of ECG measurements as an endophenotype for AF has proven valuable and has led to the discovery of variants that affect both cardiac ECG measurements and the risk of AF. Variants in or close to the genes *TBX5*, *CAV1*, *SCN5A*, *SCN10A* and *SOX5* were first reported in GWAS on the PR interval and associated secondarily with AF (Holm et al., 2010; Pfeufer et al., 2010). These loci have since been associated genome-wide significantly with AF in other cohorts, validating them as true AF variants (Christophersen et al., 2017). Furthermore, variants in or near *GJA1* and *SLC35F1* were first identified as heart rate loci (den Hoed et al., 2013) and subsequently associated with AF through GWASs (Christophersen et al., 2017). Variants near *NKX2-5*, have associated secondarily with AF in GWASs on both the PR interval (Pfeufer et al., 2010) and heart rate (den Hoed et al., 2013). Although some variants that affect ECG measurements also affect the risk of AF, there is inconsistency in the direction of their effects on ECG measurements and AF risk. The AF risk alleles can either prolong or shorten heart rate or PR interval, consistent with the reported bidirectional

association of the PR interval with AF risk (Cheng et al., 2009; Nielsen et al., 2013). Furthermore, strong and widely replicated AF loci, such as the *PITX2*- and *ZFHX3*-loci, have not been reported in GWASs on ECG measurements. These differences between AF variants with regards to their ECG effects raise the question if there is a detectable pattern in how AF variants affect cardiac conduction in the absence of AF. Exploring ECG associations of all AF variants might therefore give insight into the mechanisms behind their increase in AF risk.

1.5 Coarctation of the Aorta

Following the publication of the association of a missense variant in *MYH6* with SSS and AF (Holm et al., 2011; Thorolfsdottir et al., 2017) we tested the association of the variant with congenital heart diseases, for which we had acquired information on. In that analysis a strong association with CoA was observed. To investigate this further a GWAS on CoA was performed and is part of this dissertation. Unlike AF, which is a cardiac arrhythmia primarily affecting the elderly, CoA is a congenital defect in cardiovascular structure. Genetic studies published hitherto have in fact suggested common elements in the pathogenesis of these two seemingly different diseases.

1.5.1 Definition and pathogenesis of CoA

CoA is a congenital heart defect characterized by a narrowing of the descending aorta and/or aortic arch (Allen et al., 2016). The most common form is a discrete stenosis located in the proximal descending aorta (aortic isthmus), the area where the ductus arteriosus (ligamentum arteriosum after regression) inserts (Figure 3). However, CoA may also involve long segment narrowing, hypoplasia of the transverse aortic arch, or stenosis of the abdominal aorta (Ho & Anderson, 1979). The stenosis in CoA is usually formed by a shelf of tissue extending from the postero lateral aortic wall towards the ductus arteriosus (Ho & Anderson, 1979; Russell et al., 1991).

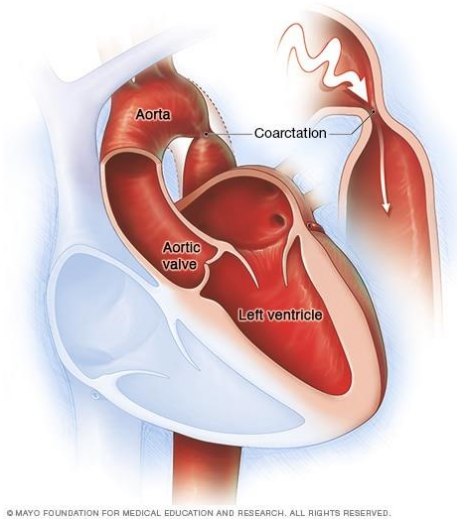


Figure 3. Illustration of the most common form of CoA, a discrete narrowing of the aortic isthmus, the part of the aorta just distal to the origin of the left subclavian artery at the site of the ductus arteriosus/ligamentum arteriosum. Used with permission of Mayo Foundation for Medical Education and Research, all rights reserved.

The pathogenesis of CoA is not well understood, but two main models have been proposed for its development. First, the haemodynamic theory suggests that the formation of coarctation is a product of decreased blood flow through the aortic isthmus during fetal development. This theory is supported by the common co-occurrence of CoA and congenital heart defects with decreased antegrade aortic flow in utero (Rudolph et al., 1972). Second, the Skodaic hypothesis, supported by histological studies (Ho & Anderson, 1979; Russell et al., 1991), proposes a central role of ectopic ductal tissue situated in the aortic wall that constricts the aortic lumen (Elseed et al., 1974).

1.5.2 Epidemiology

CoA is the most common congenital defect affecting the aorta, with an incidence of approximately 1 per 2,500-3,000 live births (Hoffman & Kaplan, 2002; van der Linde et al., 2011) and is slightly more common in males than females (Engelfriet & Mulder, 2009). It is part of a mechanistically defined subgroup of congenital heart defects, the left-ventricular outflow tract obstruction (LVOTO) malformations. LVOTOs also include bicuspid aortic valve (BAV), congenital aortic stenosis, and hypoplastic left heart syndrome (HLHS) (Allen et al., 2016). CoA may occur in isolation or accompanied by other LVOTOs, most commonly BAV which exists in about 50% of CoA

cases. Furthermore, it can occur with additional cardiac defects such as ventricular septal defect (VSD), patent ductus arteriosus, transposition of the great arteries and atrioventricular canal defect (Anderson et al., 1983; Shinebourne et al., 1976; Teo et al., 2011).

1.5.3 Presentation and treatment

CoA can present at any age but the majority (about 60%) of patients present with critical CoA in the neonatal period when closing of the ductus arteriosus after birth causes sudden obstruction of blood flow to the distal aorta (Hoffman, 2018). This results in systemic congestion and often circulatory failure and shock (inadequate blood flow to the tissues). Without prompt medical resuscitation and surgical intervention, death may occur rapidly (Chang et al., 2008). The diagnosis of CoA in the neonate is challenging. Severe symptoms usually present a few days after birth and screening methods, including pulse oximetry and assessment of femoral pulses, have limited efficacy (Ward et al., 1990). About 30-50% of cases are therefore discharged undiagnosed from the hospital (Lannering et al., 2015; Liberman et al., 2014). Prenatal screening also has a low sensitivity for diagnosis due to the presence of the ductus arteriosus and limited blood flow across the aortic isthmus in utero (Lannering et al., 2015). Furthermore, identifying CoA as an underlying cause in the neonate with shock is challenging (Hoffman, 2018).

CoA can also present later in childhood or in adults when the stenosis is milder and collateral vessels have developed to supply blood flow to the descending aorta (Steffens et al., 1994; Strafford et al., 1982). Individuals with late presenting CoA invariably have upper body hypertension and differential blood pressure between upper and lower extremities, since lower body perfusion is achieved by hypertension in the upper aorta. Heart murmurs and absent femoral pulses are also common and some experience claudication (Strafford et al., 1982). These symptoms and signs can be subtle and are often first noticed as incidental findings on physical examination (Hoffman, 2018; Strafford et al., 1982).

The treatment for CoA is correction by surgical repair or transcatheter techniques (Stout et al., 2019) and in the absence of concomitant cardiac defects, short- to medium-term prognosis is good (Hamdan et al., 2001; Kaushal et al., 2009; Wright et al., 2005). However, despite appropriate treatment, CoA is a lifelong condition involving considerable risk of long-term morbidity and mortality (Cohen et al., 1989; Presbitero et al., 1987). Even if no residual coarctation exists after surgical correction, patients often continue

to have systemic hypertension, which increases risk of heart failure, coronary artery disease and stroke (Morgan et al., 2013). Further increasing the risk of cerebrovascular events is the occurrence of cerebral aneurysms in 10% of CoA cases (Cohen et al., 1989; Presbitero et al., 1987; Singh et al., 2010). Furthermore, CoA is associated with aortic aneurysms that can occur years after surgical correction and can cause death because of dissection or aneurysmal rupture (Cohen et al., 1989; Heikkinen et al., 1991). The complex mechanisms underlying long-term complications in repaired CoA are not completely understood (Kenny et al., 2011; Singh et al., 2010).

1.5.4 Genetics of CoA

CoA is primarily a non-familial and sporadic disease. However, it can be a part of a recognized genetic syndrome, for example 45 X (Turner) (Ho et al., 2004) and Kabuki syndrome (Digilio et al., 2017), although the mechanisms behind CoA development in these syndromes is not well understood. Furthermore, CoA has been shown to cosegregate in families with LVOTO malformations and as a group the LVOTO malformations are markedly heritable (0.71 - 0.90) (McBride et al., 2005).

Several studies have reported mutations in families with LVOTO malformations and a few instances of sporadic CoA. The gene that has been most strongly implicated is *NOTCH1* (Freylikhman et al., 2014; Kerstjens-Frederikse et al., 2016; McBride et al., 2008), encoding a transmembrane receptor that regulates cell fate during development (Kopan, 2012). A candidate gene study focusing on CoA, with and without concomitant congenital heart defects, concluded that a missense variant in *NOTCH1* “may represent a disease-susceptibility allele” for CoA (Freylikhman et al., 2014). Other genes, including *MYH6* (Arrington et al., 2012; Tomita-Mitchell et al., 2016), *SMAD6* (Tan et al., 2012), *NKX2-5* (McElhinney et al., 2003) and *GATA5* (Bonachea et al., 2014), have been implicated through mutations found in one or few individuals with CoA. As mentioned previously here, variants in two of those genes, *MYH6* and *NKX2-5* have been associated with AF and specific ECG measurements through GWAS (den Hoed et al., 2013; Holm et al., 2010; Holm et al., 2011; Pfeufer et al., 2010). With regards to CoA, the *MYH6* mutations were found in two families, one with predisposition to atrial septal defect (ASD) (Arrington et al., 2012), and the other to HLHS (Tomita-Mitchell et al., 2016). Some individuals in these families presented with CoA. Many of the mutations implicated in CoA have been identified through a candidate gene approach and sometimes found in only one patient, including for example the missense mutation in *NKX2-5*

(McElhinney et al., 2003). Thus, not much is known about the genetics of sporadic CoA and GWAS has not been performed.

2 Aims

The overall aim of this doctoral dissertation was to explore the genetics of cardiovascular diseases in order to further our understanding of their pathophysiology and facilitate improvements in treatment options. The focus of the work is on exploring the genetics of the cardiac arrhythmia AF. However, through studies on the genetics of AF we observed that one of the strongest AF variants accounts for 20% of Icelandic cases of the congenital heart defect CoA. Thus, we decided to explore this variant and its effects on structural heart disease in detail resulting in a separate paper (paper IV). Using deCODE genetics extensive genotype/phenotype database we searched for novel associations of sequence variants with AF and CoA and explored the biology behind both novel and previously reported variants effects on disease risk. This was done by assessing their effects on other cardiovascular diseases and traits representing cardiac function, namely ECG and echocardiogram measurements. The specific aims of each of the four published papers are listed below.

2.1 Paper I: A GWAS on AF and assessment of the effects of sequence variants on ECG measurements

The aim of this study was to conduct a GWAS on AF in Iceland and search for novel associations. Furthermore, to gain insights into the mechanisms by which novel and published sequence variants affect AF risk by assessing their effects on other cardiovascular diseases and on normal cardiac conduction, using ECG measurements.

2.2 Paper II: GWAS on AF with focus on rare and low frequency coding and splice variants

The aim was to conduct a GWAS on AF using data from Iceland and the UK biobank, with focus on rare and low frequency coding and splice variants. The mechanisms by which the variants affect cardiac function and disease were explored using deCODE genetics extensive phenotype database, including ECG and electrocardiogram measurements. Furthermore, cardiac samples obtained during cardiothoracic surgery were used to perform RNA sequencing and assessing the effects of novel AF variants on expression at the transcript level.

2.3 Paper III: Meta-analysis of GWAS on AF

The aim was to perform a meta-analysis and substantially increase the number of variants associated with AF. Furthermore, the aim was to identify potential mechanisms by which novel variants affect AF risk. This was done by pathway and functional enrichment analysis and by assessing ECG associations, as was done in paper I and II.

2.4 Paper IV: A genome-wide association study of CoA

The aim of the study was to further explore the observed association of one of the strongest AF associated variants with CoA. To do this, we performed a GWAS on non-syndromic CoA. Furthermore we explored the effects of the variant on various CVDs, ECG traits and echocardiogram measurements in order to further our understanding of the overall effects of the variant on cardiac function and to increase understanding of the pathogenesis of CoA.

3 Materials and methods

3.1 Paper I

3.1.1 Ethical approval

The study was approved by the Data Protection Authority and National Bioethics Committee in Iceland (no. VSNb2015030021/03.01 with amendments, VSNb2015030024/03.01 with amendments, VSNb2015010009/03.12 with amendments, VSNb2015030022/03.01 with amendments). All participating subjects who donated blood signed informed consent. Personal identities of the participants and biological samples were encrypted by a third party system approved and monitored by the Icelandic Data Protection Authority. The study complies with the declaration of Helsinki.

3.1.2 The Icelandic AF study population

The AF study population consisted of 13,471 individuals diagnosed with AF (International Classification of Diseases (ICD) 10 code I.48 and ICD 9 code 427.3) at Landspitali University Hospital (LUH) in Reykjavik, the only tertiary referral center in Iceland, and at Akureyri Hospital, the second largest hospital in Iceland, from 1987 to 2015. Thereby, it may be assumed that the study includes the majority of people diagnosed with AF in Iceland at or before this period. Controls comprised 374,939 individuals recruited through different genetic research projects at deCODE genetics. All individuals diagnosed with AF were excluded from the control group. Basic characteristics of AF cases and controls are summarized in Table 1.

Table 1. AF subject and control characteristics (paper I).

	Genotyped AF cases	**Controls**
N	13,471	374,939
% Male	58.6%	50.8%
Mean age at onset (SD)	72.1 (13.6)	-

3.1.3 AF replication cohorts

Novel AF variants were genotyped in two separate AF sample sets, from the Vanderbilt AF registry (764 cases and 762 controls) (Darbar et al., 2007) and from the Further Cardiovascular Outcomes Research with PCSK9 Inhibition in Subjects with Elevated Risk (FOURIER) trial (1,238 AF cases and 11,562 controls) (Sabatine et al., 2016).

The Vanderbilt AF registry is a clinical and genetic registry at the Vanderbilt University Medical Center in Nashville, Tennessee (Darbar et al., 2007). At enrollment into the registry, patients complete a symptom questionnaire and a detailed medical and drug history is obtained. Patients with history of AF only associated with cardiac surgery were excluded from this study. Written informed consent was obtained from all patients under a protocol approved by the Vanderbilt University Institutional Review Board.

The FOURIER trial is a randomized placebo-controlled, double-blind, parallel-group, multinational trial. It tests the hypothesis that adding the drug evolocumab to statin therapy will reduce the incidence of major adverse cardiovascular events in patients with clinically evident cardiovascular disease. The whole study group consisted of 27,564 patients recruited at 1,242 cities in 49 countries from 2013 to 2015. Eligible patients were between 40 and 85 years of age and had clinically evident atherosclerotic cardiovascular disease. Detailed phenotypic information was gathered on all FOURIER study participants, including AF disease status. The design of the trial has been described in detail elsewhere (Sabatine et al., 2016). A subset of over 12,000 participants of European descent from the FOURIER trial have been genotyped at deCODE by whole exome sequencing, chip-typing and imputation. After exclusion of all Icelandic participants the AF sample set from the FOURIER trial consisted of 1,238 cases and 11,562 controls,

3.1.4 ECG data

ECG data included all ECGs obtained and digitally stored at LUH in Reykjavik from 1998 to 2015. These were ECGs obtained from both inpatients and outpatients in all hospital departments, a total of 434,000 ECGs from 88,217 individuals. Out of these, a total of 289,297 ECGs were sinus rhythm (heart rate 50-100 beats per minute) ECGs of 62,974 individuals without the diagnosis of AF. The ECGs were digitally recorded with the Philips PageWriter Trim III, PageWriter 200, Philips Page Writer 50 and Phillips Page Writer 70 cardiographs and stored in the Philips TraceMasterVue ECG Management System. Digitally measured ECG

waveforms and parameters were extracted from the database for analysis. The Philips PageWriter Trim III QT interval measurement algorithm has been previously described and shown to fulfill industrial ECG measurement accuracy standards (Zhou et al., 2009). The Philips PR interval and QRS complex measurements have been shown to fulfill industrial accuracy standards (Gregg et al., 2005).

3.1.5 The deCODE genetics phenotype database

After identifying novel associations with AF, we searched for secondary associations with relevant phenotypes in deCODE genetics phenotype database. This database contains extensive medical information on various diseases and other traits. A special focus was on reporting the associations of AF variants with the known AF related traits of pacemaker implantation, SSS, ischemic stroke and cardioembolic stroke. The pacemaker population sample set included 3,578 individuals who received a pacemaker implantation (NCSP surgical codes FPE and FPF) at LUH between 1997 and 2015. The SSS sample set included 3,310 individuals who received the diagnosis of SSS (ICD 10 code I49.5, ICD 9 code 427.8) at LUH in Reykjavik between 1987 and 2015. Ischemic stroke cases were identified from a registry of individuals diagnosed with ischemic stroke or transient ischemic attack (TIA) at LUH during the years 1993 to 2014 (n = 5,626). The ischemic stroke or TIA diagnoses were based on standard WHO criteria and imaging evidence (either computed tomography or magnetic resonance imaging) and were clinically confirmed by neurologists. A total of 1,369 individuals with ischemic stroke were classified as having cardioembolic stroke based on a neurologist review of medical records and classification according to the Trial of Org 10172 in Acute Stroke Treatment (TOAST) (Adams et al., 1993). The controls used in the various case-control analyses of this study consisted of disease-free individuals from the Icelandic genealogical database at deCODE genetics.

3.1.6 Generation of genotype data

The DNA samples used in this study have been collected through various genetic studies at deCODE genetics since 1996. The GWAS performed in paper I was based on whole-genome sequence data from the whole blood of 15,220 Icelanders. In addition, 151,677 Icelanders have been genotyped using Illumina SNP chips and genotype probabilities for untyped relatives calculated based on Icelandic genealogy. Sequencing was carried out using Illumina standard TruSeq methodology to a mean depth of 35 (SD 8).

Autosomal SNPs and INDEL's were identified using the Genome Analysis Toolkit version 3.4.0 (McKenna et al., 2010). According to GATK best practices, variants that did not pass quality control were excluded from the analysis (Gudbjartsson et al., 2015). Information about haplotype sharing was used to improve variant genotyping. This was possible since all sequenced individuals had also been chip typed and long-range phased (Kong et al., 2008).

The imputation information (informativeness of genotype imputation) was calculated using the ratio of the variance of imputed expected allele counts and the variance of the actual allele counts:

$$\frac{\mathrm{Var}(\mathrm{E}(\theta|\text{chip data}))}{\mathrm{Var}(\theta)},$$

where θ is the allele count. Here, *Var(E(θ│chip data))* was estimated by the observed variance in the imputed expected counts and *Var(θ)* was estimated by *p*(1-*p*), where *p* is the allele frequency.

Variants were annotated using Variant Effect Predictor (VEP) version 2.8 (McLaren et al., 2010) and Ensembl release 80 (Ruffier et al., 2017). A total of 32.5 million variants passed the quality threshold and were imputed into 151,677 Icelanders who had been genotyped using Illumina chips.

For genotyping of single variants in foreign AF sample sets, we used the Centaurus (Nanogen) or KASP platforms. For quality control, the genotypes of one of the variants, rs72700114 at *LINC01142/METTL11B* were confirmed by sanger sequencing in 245/764 cases and 233/762 controls in the Vanderbilt samples.

3.1.7 Association analysis

Logistic regression was used to test for association between 32.5 million variants and AF or secondary traits, treating disease status as the response and allele counts from direct genotyping or expected genotype counts from imputation as covariates. Gender, age and county were included as covariates. Calculations where made under additive and recessive genotype models. The additive model for the minor allele assumes that there is a uniform, linear increase in risk for each copy of the minor allele. In other words, if **A** is the major allele and **a** is the minor allele then the additive model assumes that if the risk for **Aa** genotype is *k* the risk for **aa** genotype is *2k*.

The recessive model (for **a**) assumes that two copies of the **a** allele are required to alter risk. It therefore compares individuals with the **aa** genotype to everyone else (**AA** and **Aa**) (Lewis, 2002).

To account for inflation in test statistics due to cryptic relatedness and stratification, the method of LD score regression was applied (Bulik-Sullivan et al., 2015). With a set of 1.1 million variants the $\chi 2$ statistics from the GWAS scan were regressed against LD score. The intercept was then used as a correction factor. The LD scores were downloaded from a LD score database (ftp://atguftp.mgh.harvard-.edu/brendan/1k_eur_r2_hm3snps_se_weights.RDS; accessed 23.06.2015). The estimated correction factor for AF was 1.38 for the additive model.

The threshold for genome-wide significance was corrected for multiple testing with a weighted Bonferroni adjustment. The method uses as weights the enrichment of variant classes with predicted functional impact among association signals (Sveinbjornsson et al., 2016). With 32,463,443 sequence variants being tested the weights given in Sveinbjornsson et al. were rescaled to control the family-wise error rate (FWER). The resulting significance thresholds were 2.6×10^{-7} for high-impact variants (N = 8,464), 5.1×10^{-8} for moderate-impact variants (N = 149,983), 4.6×10^{-9} for low-impact variants (N = 2,283,889), 2.3×10^{-9} for other variants in DNase I hypersensitivity sites (N = 3,913,058) and 7.9×10^{-10} for other variants (N = 26,108,039). Conditional analysis of the region around the *PLEC* variant, p.Gly4098Ser, was performed by adding the p.Gly4098Ser as a covariate while testing every SNP in the region for association with AF.

In paper I, two novel and 29 replicated AF variants were tested for association with 122 ECG measurements (Table 2) using linear regression. The ECG measurement was treated as the response and the genotype as the covariate. All measures except heart rate and QT corrected are presented for all 12 ECG leads. For this analysis, 289,297 sinus rhythm ECGs (heart rate 50-100 beats per minute) from 62,974 individuals where used. People who had been diagnosed with AF were excluded from this analysis in order to assess the effect of the AF variants on cardiac electrical function in the absence of AF. Individuals with pacemakers were also excluded. The analysis was also done using all ECGs irrespective of rhythm and history of AF, a total of 434,000 ECGs from 88,217 individuals. The ECG measurements were adjusted for sex, year of birth and age at measurement and were subsequently standardized to have a normal distribution. For individuals with multiple ECG measurements, the mean standardized value

was used. It was assumed that the quantitative measurements follow a normal distribution and that their mean depends linearly on the expected allele at the variant and a variance-covariance matrix proportional to the kinship matrix (Benonisdottir et al., 2016). The Benjamini-Hochberg false discovery rate (FDR) procedure controlling the FDR at 0.05 at each marker was used to account for multiple testing, since 122 traits were tested.

Table 2. List of ECG measures available for analysis of correlation with AF risk variants and sample sizes (rounded to hundred) in two groups (paper I).

	N: all ECGs included	**N: sinus rhythm, excluding AF cases and individuals w pacemaker**	**Number of measurements in each category**
Heart rate	88,200	63,000	1
P amplitude	88,100	62,900	12
P area	87,800	62,900	12
P duration	88,000	62,800	12
PR segment	88,200	62,900	12
PR interval	88,100	62,900	12
QRS duration	88,200	63,000	12
R amplitude	88,000	62,900	12
QT interval	88,200	63,000	1
T amplitude	88,100	62,900	12
T area	88,100	62,900	12
T duration	88,100	62,900	12

3.2 Paper II

Paper II describes a meta-analysis of GWAS on AF in Iceland and the publicly available UK Biobank dataset, focusing on rare and low frequency coding and splice variants. The most significant variants were tested in samples from Norway and the US.

3.2.1 Ethical approval

The study was approved by the Data Protection Authority and National Bioethics Committee in Iceland (no. VSNb2015030021/03.01 with amendments, VSNb2015030024/03.01 with amendments, VSNb2015010009/03.12 with amendments, VSNb2015030022/03.01 with amendments). All participating subjects who donated blood signed informed consent. Personal identities of the participants and biological samples were encrypted by a third party system approved and monitored by the Icelandic Data Protection Authority. The study complies with the declaration of Helsinki.

3.2.2 Study populations

For this study, the Icelandic AF sample had been expanded and consisted of 14,710 Icelanders diagnosed with AF (ICD 10 code I.48 and ICD 9 code 427.3) according to electronic medical records at LUH and Akureyri Hospital, between 1987 and 2017. Controls were 373,897 individuals without the diagnosis of AF.

The UK Biobank project is a large prospective cohort study of ~500,000 individuals from across the United Kingdom. Participants were aged 40-69 at recruitment and their extensive phenotypic and genotypic information has been collected (Sudlow et al., 2015). The AF population from the UK Biobank consisted of 14,792 cases and 393,863 controls. All were individuals of European ancestry that were recruited between 2006 and 2010 (Sudlow et al., 2015). AF was ascertained based on primary or secondary ICD 9 or ICD 10 diagnoses codes from both inpatient and outpatient hospital episodes. Self-reported diagnoses were not included in the analysis. Further details on the recruitment and variables collected in the UK Biobank study can be found in previous publications (Elliott & Peakman, 2008; Sudlow et al., 2015).

The Tromsø Study is a population-based prospective study in the municipality of Tromsø, Norway (Jacobsen et al., 2012). The study involves repeated health surveys and examinations including 40,000 individuals thus far. The population is being followed-up on an individual level with registration and validation of diseases and death. Discharge diagnosis lists of CVDs have been retrieved from the University Hospital of North Norway in Tromsø, and medical records of all individuals with a CVD discharge diagnosis have been reviewed. AF has been registered based on ICD 9 and ICD 10 codes since 1986 as part of the ongoing CVD endpoint registration. People with postoperative AF only (<= 28 days after the procedure) were not included as cases. For the current project, one sex- and age matched control for each case of AF was drawn from the population based Tromsø 4 survey, resulting in 714 cases and 698 controls.

The Nord-Trøndelag Health Study (HUNT) is a population-based health survey conducted in the county of Nord-Trøndelag, Norway. Individuals have been recruited at three different time points during approximately 20 years (HUNT1 [1984-1986], HUNT2 [1995-1997] and HUNT3 [2006-2008]]) (Krokstad et al., 2013). The entire adult population (≥ 20 years) was invited to participate at each time point. Taken together, the health studies included information from over 120,000 different individuals. For the current study AF was defined based on ICD 10 codes collected from local hospitals and

outpatient clinics between 1999 and 2016. Cases were 6,493 individuals with one or more ICD 9 or ICD 10 codes specific for AF ("I48" or "427.3") whereas controls comprised 63,142 individuals without a code specific for AF.

Novel AF associations were also assessed in 764 AF cases and 762 controls from the Vanderbilt Atrial Fibrillation Registry (Darbar et al., 2007) and 1,238 cases and 11,562 controls from the FOURIER trial (Sabatine et al., 2016). These cohorts were also included in paper I and are described in detail in 3.1.3.

3.2.3 Electrocardiogram data

The ECG dataset was the same as described in paper I (see 3.1.4.).

3.2.4 Secondary phenotypes

As was done in paper I, novel AF variants in paper II were tested for association with other phenotypes in deCODE genetics extensive phenotype database (see 3.1.5 for details). Novel AF variants were also tested for association with SSS among 403 cases and 403,181 controls in the UK Biobank.

3.2.5 Generation of genotype data

The GWAS performed in paper II was based on the same numbers of whole-genome sequenced and chip typed individuals as in paper II, (15,220 and 151,677 respectively). The generation of genotype data in Iceland was as described in detail in 3.1.6. In the UK Biobank, genotyping was performed using a custom-made Affymetrix chip, UK BiLEVE Axiom (Wain et al., 2015), in the first 50,000 participants, and with Affymetrix UK Biobank Axiom array in the remaining participants (Welsh et al., 2017); 95% of the signals are on both chips. Imputation was performed by Wellcome Trust Centre for Human Genetics using a combination of 1000 Genomes phase 3 (Auton et al., 2015), UK10K (Walter et al., 2015) and *Haplotype Reference Consortium* (HRC) reference panels (McCarthy et al., 2016) for up to 92,693,895 SNPs (Bycroft et al., 2017). For genotyping of single variants in AF sample sets from the Vanderbilt registry, FOURIER trial, HUNT and the Tromsø study Centaurus (Nanogen) or KASP platforms were used.

3.2.6 Statistical analysis

A meta-analysis was performed on 14,710 AF cases and 373,897 controls from Iceland and 14,792 cases and 393,868 controls from the UK Biobank.

Logistic regression was used to test for association between SNPs and AF and other phenotypes in the Icelandic study, treating phenotype status as the response and allele count as a covariate. The allele counts were either from genotyping or integrated over possible genotype counts based on imputation. Other available individual characteristics that correlate with phenotype status were also included in the model as nuisance variables. In Iceland these covariates were: sex, county of birth, current age or age at death (first and second order terms included), blood sample availability for the individual and an indicator function for the overlap of the lifetime of the individual with the time span of phenotype collection. In the UK biobank study age and sex were included as covariates in the logistic regression model and 40 principal components were used to adjust for population stratification. A fixed-effects inverse variance method was used for the meta-analysis (Mantel & Haenszel, 1959). The method was based on effect estimates and standard errors from the Icelandic and the UK Biobank study. Only sequence variants from the Haplotype Reference Consortium panel (McCarthy et al., 2016) were included in the meta analysis and variants from deCODE and the UK Biobank imputation were matched on position and alleles. Standard errors were calculated in the following way:

For a P value smaller than 1 the standard error was calculated as:

$$P = 2\Phi(z) = 2\Phi\left(\frac{\beta}{\sigma}\right).$$

Solving for σ gives

$$\sigma = \frac{\beta}{\Phi^{-1}\left(\frac{P}{2}\right)}$$

If $P = 1$, then $\Phi^{-1}\left(\frac{P}{2}\right) = 0$ and the above method breaks down. In this case data from other markers was used to estimate the relationship between allele frequency (f) and imputation information (I) and σ:

$$Var(\beta) = \sigma^2 \propto \frac{1}{N} f(1-f) \propto \frac{1}{I} f(1-f)$$

Sample size (N) is proportional to imputation information (I) if the analysis is always based on the same set of individuals. Therefore, if we fit the following linear model:

$$\log(\sigma^2) = \gamma_1 + \gamma_I \log(I) + \gamma_f \log(f(1-f))$$

for a subset of 100,000 markers spread over the genome with MAF ranging close to uniformly between 0.1% and 50% and info between 0.9 and 1 and pick the subset of markers with $P < 0.9$ then we can predict σ for a marker with P close to 1.

The threshold for genome-wide significance was corrected for multiple testing with a weighted Bonferroni adjustment as described in 3.1.7. Conditional analysis of the region around novel AF variants was performed in the Icelandic data by adding the top variant or variants as a covariate while testing every SNP in the region for association with AF.

Novel AF variants were tested for association with 122 ECG measurements using linear regression, treating the ECG measurement as the response and the genotype as the covariate. This was also done in paper I and described in detail in 3.1.7. In this study (paper II), the ECG associations of the novel AF variants were plotted next to the associations of the 31 variants from paper I for comparison.

3.2.7 RNA sequencing in cardiac samples

Samples from cardiac right atrium of 167 Icelandic subjects were used for RNA sequencing. The mean age at biopsy was 68.8 years (SD 12.2) and the majority were male (140/167). The samples were obtained during cardiothoracic surgery at LUH. The RNA samples from cardiac atria were used to identify a novel isoform resulting from the splice-donor variant in *RPL3L* (c.1167+1G>A) and quantify expression at the transcript level. RNA sequencing libraries were inspected for sequencing and alignment integrity using parameters retrieved from RNA-SeQC (DeLuca et al., 2012), Picard CollectRnaSeqMetrics (http://broadinstitute.github.io/picard/) and FastQC (http://www.bioinformatics.babraham.ac.uk/projects-/fastqc). Genotype concordance was determined by comparing imputed genotypes to those derived from RNA-seq. Genome alignments were found using STAR (Dobin et al., 2013) aligning to GRCh38 with ensemble v87 (Yates et al., 2016) gene annotations.

Alignments of RNA-seq reads of carriers of the variant contained several reads that spliced over exon 9 in transcript ENST00000268661 of *RPL3L*. Neither of the other two annotated transcripts of *RPL3L* (ENST00000565426 and ENST00000566484 (Yates et al., 2016)) showed expression in any of the samples. To assess quantitatively the effect of the variant on the isoform usage we created the transcript sequence for the novel isoform, added it to the ensemble v87 transcriptome, and re-quantified all samples using kallisto (Bray et al., 2016) and the modified transcriptome. The expression of the annotated and novel transcript was corrected w.r.t. the size factor computed from the gene expression analysis. Finally, the proportion of novel isoform usage was computed by dividing the estimated expression of the novel isoform with the sum of the expression of both isoforms.

Two samples out of 167 were from carriers of the splice-donor variant in *RPL3L* in question. Due to the small number of carriers, we opted for a conservative test for computing the significance of the observed event, that the carriers have a ratio of 50% vs near 0% for non-carriers. The test used was the two-sided Mann-Whitney U test, which only takes the relative ranks of the samples into account and not the underlying values. The *P* value computed was $P = 0.0052$, the lowest possible *P* value that can be obtained using this statistical test with $n_1 = 165$ and $n_2 = 2$.

3.3 Paper III

Paper III describes a multi-center meta-analysis of GWASs on AF in which the AF cohort from deCODE was combined with five other large sample sets. The number of participants was over one million, including 60,620 AF cases.

3.3.1 Study cohorts

The Icelandic AF population from deCODE genetics was described in 3.1.2. and consisted of 13,471 patients diagnosed with AF (ICD 10 code I.48 and ICD 9 code 427.3) at LUH and Akureyri Hospital from 1987 to 2015 and 358,161 controls.

A total of 6,493 AF cases and 63,142 AF-free controls came from the HUNT study, described in 3.2.2. A combination of hospital, outpatient, and emergency room discharge diagnoses (ICD 9 and ICD 10) was used to identify cases and controls.

The Michigan Genomics Initiative (MGI) is a hospital-based cohort collected at Michigan Medicine, USA. AF cases were 1,226, defined as patients with ICD 9 code 427.31. Controls were individuals without AF, atrial

flutter, or related phenotypes (ICD-9 426-427.99). MGI was reviewed and approved by the Institutional Review Board of the University of Michigan Medical School.

The DiscovEHR collaboration cohort is a hospital-based cohort including 58,124 genotyped individuals of European ancestry from the ongoing MyCode Community Health Initiative of the Geisinger Health System, USA (Carey et al., 2016). AF cases were 6,679 individuals defined as DiscovEHR participants with at least one electronic health record problem list entry or at least two diagnosis code entries for two separate clinical encounters on separate calendar days for ICD 10 I.48. Corresponding controls were 41,803 individuals with no electronic health record diagnosis code entries for ICD 10 I.48. The Study was approved by the Geisinger Institutional Review Board.

The UK Biobank cohort is described in 3.2.2 as is was also included in the meta-analysis in paper II (Sudlow et al., 2015). Cases of AF were selected using ICD 9 and ICD 10 codes for AF or atrial flutter (ICD 9 427.3 and ICD 10 I48). Controls were participants without any ICD 9 or ICD 10 codes specific for AF, atrial flutter, other cardiac arrhythmias, or conduction disorders.

A total of 17,931 AF cases and 115,142 controls were obtained from published AF association summary statistics from 31 cohorts in the AFGen Consortium (Christophersen et al., 2017).

3.3.2 Genotyping array, imputation and association analysis

Genotyping and association analysis in the deCODE data has been described here in detail in 3.1.6 and 3.1.7. Variants for the meta-analysis were selected based on matching with either the 1000 Genomes phase 3 (Auton et al., 2015) or the Haplotype Reference Consortium reference panel (McCarthy et al., 2016) (based on allele, frequency and correlation matching). Genoyping in the other cohorts is described in detail in paper III (Nielsen et al., 2018b).

3.3.3 Meta-analysis

All markers that were available for analysis in the six contributing studies were included in the meta-analysis. For the DiscoverEHR that applied the BOLT-LMM mixed model, an approximation of the allelic log-odds ratio and corresponding variance from the linear model was optained as described previously (Cook et al., 2017). Subsequently, a meta-analysis was performed using the inverse variance method implemented in the software package METAL (http://genome.sph.umich.edu/wiki/METAL_Documentation) (Willer et

al., 2010). When estimating the cross-cohort allele frequencies, only the studies sampling individuals independent of AF status were included (HUNT, deCODE, MGI, DiscoverEHR, UK Biobank). This was done to avoid sampling bias. Heterogeneity tests were performed as implemented in METAL (Willer et al., 2010).

Independent loci were defined as genetic markers > 1Mb and > 0.25 cM apart in physical and genomic distance, respectively, with at least one genetic variant associated with AF. The genome-wide significance threshold was set at a P value < 5×10^{-8}. The borders of loci were defined as the highest and lowest genomic positions within the locus reaching genome-wide significance plus an additional 1Mb on either side. The software PLINK1.9 (https://www.cog-genomics.org/plink/1.9) was used to calculate LD r^2 based on 5,000 unrelated individuals that were randomly sampled among the HUNT study participants. We also used the 1000 Genomes phase 3 European sample for LD estimation (Auton et al., 2015).

To identify independent risk variants within the identified AF-associated loci, the COJO-GCTA software (http://cnsgenomics.com/software/gcta/) was used. It performed approximate, stepwise conditional analyses based on summary statistics from the meta-analyses and a LD-matrix obtained from 5,000 unrelated individuals randomly sampled from the HUNT study (Yang et al., 2012). Only variants with MAF > 0.01 were included in the analyses and variants were only considered truly independent if they were not in LD (r^2 < 0.05) with the locus index variant and any of the other independent risk variants.

3.3.4 Estimation of heritability

The genome-wide heritability explained by all markers was estimated based on LD-score regression, GWAS summary statistics and European-ancestry LD information from the 1000 Genomes Project (Bulik-Sullivan et al., 2015). The heritability explained by AF-associated index variants and additional independent risk variants was calculated on the basis of odds ratios and risk allele frequencies as described previously (So et al., 2011).

3.3.5 Pathway and functional enrichment analysis

As explained in “Declaration of contribution”, the methods of this part of the work will be covered here briefly. For a more detailed description see paper III (Nielsen et al., 2018b).

The software GREGOR (Genomic Regulatory Elements and Gwas

Overlap algorithm, http://csg.sph.umich.edu/GREGOR/) was used to test for enrichment of index variants with functional domains (Schmidt et al., 2015). This method tests if the AF-associated variants, or their LD proxies, overlap with experimentally annotated regulatory domains more often than expected by chance and in which tissues. The regulatory features included DNase I hypersensitive sites (DHS), histone methylation marks, transcription factor binding sites, and chromatin states in a variety of cell and tissue types available from Roadmap Epigenomics (Kundaje et al., 2015).

The DEPICT tool (https://data.broadinstitute.org/mpg/depict/) was employed to identify the most likely causal gene at associated loci, reconstructed gene sets enriched for genes at AF loci and tissues and cell types in which genes at associated loci are preferentially expressed (Pers et al., 2015). To do this DEPICT employs gene expression data and predicted gene functions. The expression data is derived from a panel of 77,840 mRNA expression arrays (Fehrmann et al., 2015). Predicted gene functions are on the form of 14,461 gene sets which are defined based on molecular pathways derived from experimentally verified protein-protein interactions (Lage et al., 2007), genotype-phenotype relationships from the Mouse Genetics Initiative (Blake et al., 2014), Reactome pathways (Croft et al., 2011), KEGG pathways (Kanehisa et al., 2012), and Gene Ontology (GO) terms (Ashburner et al., 2000). DEPICT reconstructs the gene sets so that each gene in the genome is functionally characterized by its membership probabilities across the 14,461 reconstituted gene sets. Finally, DEPICT relies on 37,427 human gene expression microarrays to identify tissues and cell types in which genes from associated loci are highly expressed (Pers et al., 2015).

A systematic approach was applied to identify functional candidate genes possibly explaining the associations of the 111 AF loci. There were 3,048 genes or transcripts for which the transcribed region overlapped at least one genome-wide significant variant in the 111 loci. Out of these, biological candidate genes were prioritized if they i) harbored a protein-altering index variant themselves or in high LD ($r^2 > 0.80$); ii) had expression levels that were associated and colocalized with AF-associated variants (P value < 1.14×10^{-9} in GTEx consortium data) ("Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans," 2015); iii) were highlighted by DEPICT (FDR < 0.05); or iv) were nearest to the index variant in a locus.

3.3.6 Testing AF variants for association with ECG measurements

deCODE genetics provided ECG data to assess the effects of AF variants on cardiac conduction in the absence of AF. The dataset is described in 3.1.4. The 111 genome-wide significant AF index variants were tested for association with 123 ECG parameters using a linear mixed effects model implemented in the Bolt software package (Loh et al., 2015), treating the ECG measurement as the response and the genotype as the covariate. All measures except heart rate and QT interval were presented for all 12 ECG leads. For this analysis, we used 289,297 sinus rhythm ECGs (heart rate 50-100 beats per minute) from 62,974 individuals who had not been diagnosed with AF nor had an implanted pacemaker according to our databases. Further details regarding the association analysis are described in 3.1.7.

3.3.7 Genetic risk scores

An inverse normal-transformed genetic risk score for AF was constructed for each study participant in the UK Biobank and in the HUNT study. This was done using summarized dosage-weighted risk estimates from all 142 independent genome-wide statistically significant risk variants. For the UK Biobank risk score, risk estimates (beta coefficients) were obtained by meta-analyzing the risk variants across all contributing studies excluding the UK Biobank. To explore the association between the genetic burden of AF and the age-of-onset of AF, which we assumed was independent of the case status used for obtaining the risk estimates, we obtained risk estimates from meta-analyses of the full sample size.

3.3.8 Phenome-wide association analyses in the UK Biobank

A previously published scheme was used to define disease-specific binary phenotypes by combining hospital ICD 9 codes into hierarchical PheCodes, each representing a particular disease group (Denny et al., 2013). ICD 10 codes were mapped to PheCodes using a combination of available maps through the Unified Medical Language System (https://www.nlm.nih.gov/research/umls/), string matching, and manual review. UK Biobank study participants were labeled with a PheCode if they had one or more of the PheCode-specific ICD codes. Cases were all UK Biobank study participants with the PheCode of interest and controls were all UK Biobank study participants without the PheCode of interest or any related PheCodes. Gender checks were performed, so PheCodes specific for one gender could not mistakenly be assigned to the other gender. The

association between the optimized genetic risk score and each of the defined phenotypes where tested using a logistic regression adjusted for sex and birth year.

3.4 Paper IV

3.4.1 Ethical approval

The study was approved by the National Bioethics Committee of Iceland and the Icelandic Data Protection Authority (VSNb2015030021/03.01 with amendments, VSNb2015030024/03.01 with amendments, VSNb2015010009/03.12 with amendments, VSNb2015030022/03.01 with amendments). Written informed consent was obtained from all study participants. The study complies with the declaration of Helsinki.

3.4.2 The Icelandic CoA study sample

The CoA sample set included individuals who received the discharge diagnosis of CoA at LUH, the only hospital in Iceland with specialized care for patients with congenital heart defects, between 1984 and 2016. The individuals diagnosed with CoA were identified either through diagnosis codes of CoA (ICD 9 code 747.1, ICD 10 code Q25.1) registered between 1990 and 2016 or procedure codes of CoA (WHO codes 1-273, 5-369, 5-382 and 5-387, NOMESCO codes FDJ 00, FDJ 10, FDJ 20, FDJ 30, FDJ 42 and FDJ 96) registered between 1984 and 2016. CoA was defined as a congenital narrowing of aorta, the diagnosis of which was confirmed by echocardiography and/or cardiac catheterization by a cardiologist. The diagnoses of CoA and other CHDs were confirmed and type of CoA established through review of electronic and paper medical records at LUH. Originally 146 Icelanders with CoA were identified, 11 of which were syndromic and excluded from the study. Of the remaining 135 individuals with CoA, genotypes were available for 120 individuals that were included in the analysis. The individuals that served as controls consisted of disease-free individuals randomly drawn from the Icelandic genealogical database and individuals from other genetic studies at deCODE.

Based on severity and/or anatomy of the narrowing, individuals were classified into five different types of CoA: (1) Mild CoA was defined as untreated and hemodynamically insignificant CoA, (2) moderate CoA as treated, hemodynamically significant CoA and (3) critical CoA as severe narrowing presenting as either congestive heart failure or cardiogenic shock during the neonatal period, with or without concomitant large VSD. (4)

Complex CHD was defined as CoA occurring in conjunction with multiple CHDs Finally, (5) we included coarctation of the abdominal aorta and interruption of the aortic arch type A. The age at diagnosis and the presence of associated hypoplasia, CHDs, arrhythmias and other CVDs was also documented. The phenotypic information was used for phenotype-genotype correlation of the individuals with CoA.

3.4.3 Secondary phenotypes in the deCODE genetics phenotype database

A novel CoA variant was tested for association with secondary phenotypes in the deCODE genetics phenotype database, described in 3.1.5. The CHD sample sets included 715 individuals with VSD (ICD 10: Q21.0 or ICD 9: 745.4), 657 individuals with ASD (ICD 10: Q21.1 or ICD 9: 745.5), 594 individuals with patent ductus arteriosus (ICD 10: Q25.0 or ICD 9: 747.0) and 208 individuals with description of bicuspid aortic valve according to cardiologist interpretation of echocardiograms from LUH between 1994 and 2015. The coronary artery disease dataset (N = 37,782) has been described (Helgadottir et al., 2016). The aortic valve stenosis (N = 2,457), AF (N = 13,471), heart failure (N = 10,480), high-degree atrioventricular block (N=1,303) and thoracic aortic aneurysm (N = 353) sample sets were based on discharge diagnoses from LUH from 1987 to 2015. The ischemic stroke (N = 8,948) and SSS (N = 3,310) sample sets have been described in detail elswere (Gretarsdottir et al., 2008; Holm et al., 2011). The hypertension sample set included 54,974 individuals who received the diagnosis of hypertension at LUH or the Primary Health Care Clinics of the Reykjavik area. Identification of patients with HCM has been described elswere (Adalsteinsdottir et al., 2014).

Three phenotypes describing measurements from echocardiograms were used in the analysis, left atrial diameter (N = 19,380), aortic root diameter (19,506) and left ventricular end-diastolic diameter (N = 5,701). These measurements were obtained from a database of 53,122 echocardiograms from 27,460 individuals performed and documented by a cardiologist at LUH between 1994 and 2015. Measurements were adjusted for sex, year of birth and age at measurement and were subsequently standardized to have a normal distribution. Electrocardiogram (ECG) data was collected from LUH in Reykjavik and included all ECGs obtained and digitally stored from 1998 to 2015, a total of 434,000 ECGs from up to 88,217 individuals, analysis was performed on 289,297 sinus rhtyhmi ECGs of 62,974 individuals without the diagnosis of AF, as explained in 3.1.4.

3.4.4 Generation of genotype data

The GWAS on CoA performed in paper IV was based on the same numbers of whole-genome sequenced and chip typed individuals as in paper I and II, (15,220 and 151,677 respectively). Genotyping, variant calling and imputation was described in detail in 3.1.6. Of the individuals diagnosed with CoA, 39 were chip typed and long-range phased and of the individuals that were controls 140,661 were chip typed and long-range phased. The remaining individuals (n = 81 CoA cases and n = 214,412 controls), were not chip typed themselves but were first or second degree relatives of the chip typed individuals and imputed using genealogical imputation as described in (Styrkarsdottir et al., 2013).

3.4.5 Association analysis

Association testing for case-control analysis was performed using logistic regression, adjusting for gender, age and county as is described in 3.1.7. The threshold for genome-wide significance was corrected for multiple testing with a weighted Bonferroni adjustment (Sveinbjornsson et al., 2016). Threshold for functional groups of variants are listed in 3.1.7. The estimated correction factor was 1.04 for the multiplicative model of the CoA association.

3.4.6 Phenotypic differences between carriers and non-carriers of a variant associated with increased risk of CoA

To analyze if carriers of p.Arg721Trp in *MYH6* differed clinically from non-carrier individuals with CoA, we evaluated the frequencies of various clinical characteristics in these two groups of individuals with CoA. Fisher's exact test was used to test for significant difference in the mean frequency of the variants between non-carriers and carriers, and the odds ratio (OR) was calculated as (pa/(1-pa)) / (pc/(1-pc)), where pa and pc are the mean frequencies of the variants in non-carriers and carriers, respectively.

4 Results

4.1 A genome-wide association study on AF in Iceland (paper I)

Paper I reports the results of a GWAS testing 32.5 million sequence variants for association with AF in 13,471 Icelandic cases and 374,939 controls under the additive model. Genotype-information was based on whole-genome sequencing of 15,220 Icelanders with variants imputed into 151,677 chip typed and long-range phased individuals and their close relatives (Gudbjartsson et al., 2015). The study yielded ten genome-wide significant associations, two of which were novel, one with an intergenic variant at the genes *METTL11B* and *LINC01142* and one with a missense variant in the gene *PLEC* (Figure 4, Table 3).

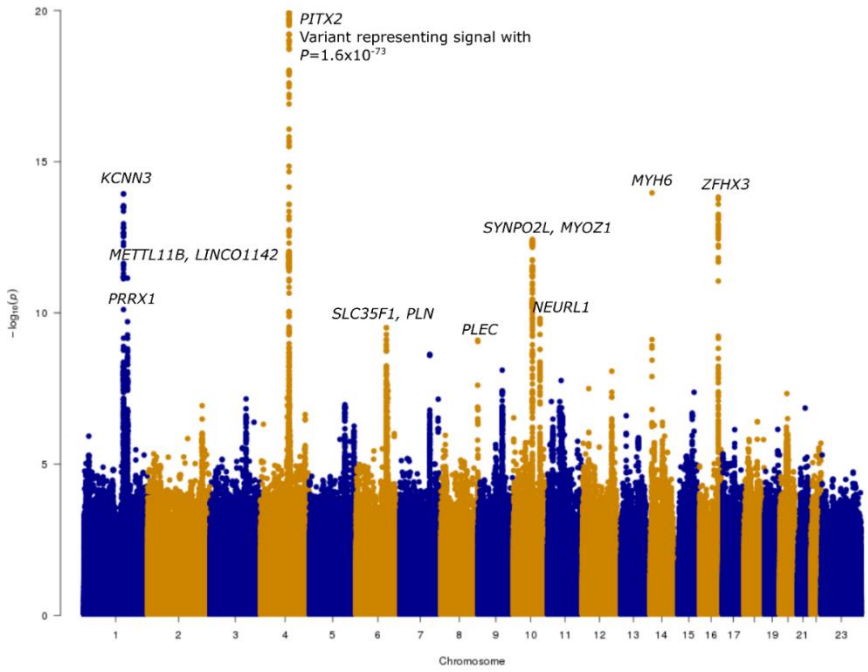


Figure 4. Manhattan plot of the results of the first GWAS on AF (N = 13,471, paper I). The $-\log_{10}(P \text{ value})$ is plotted against the position of each SNP on each chromosome. Genome-wide significant variants are annotated with the corresponding gene name. The strongest signal, close to *PITX2*, is cut at a level of $P = 1 \times 10^{-20}$.

Table 3. Variants associating genome-wide significantly with AF in paper I (N = 13,471)

Reference SNP[a] ID	Chr[b]/Pos[c] (hg38)	Closest genes	Risk allele/ other	RAF[d] (%)	Coding effect	OR[e] (CI[f])	*P* value	*P* value trheshold	Reported SNP ID	r^2 with reported SNP
Novel signals										
rs72700114	1q24/170,224,684	*METTL11B, LINC01142*	C/G	8.1	intergenic	1.22 (1.15-1.29)	7.0×10^{-12}	2.3×10^{-9}	NA	NA
rs373243633	8q24/ 143,917,940	*PLEC*	T/C	1.2	missense	1.55 (1.35-1.78)	8.0×10^{-10}	5.1×10^{-8}	NA	NA
Variants reported to associate with AF										
rs6843082	4q25/110,796,911	*PITX2*	G/A	20.1	intergenic	1.43 (1.38-1.49)	1.6×10^{-73}	2.3×10^{-9}	rs6843082[h]	1.00
rs387906656[g]	14q11/23,396,970	*MYH6*	A/G	0.3	missense	2.43 (1.94-3.04)	1.1×10^{-14}	5.1×10^{-8}	rs387906656[i]	1.00
rs34245846	1q21/154,858,667	*KCNN3*	G/A	32.7	Intronic	1.15 (1.11-1.19)	1.2×10^{-14}	7.9×10^{-10}	rs13376333[h]	0.52
rs2359171	16q22/73,019,123	*ZFHX3*	A/T	18.8	Intronic	1.17 (1.13-1.22)	1.5×10^{-14}	2.3×10^{-9}	rs2106261[j]	1.00
rs148321568	10q22/73,668,538	*SYNPO2L, MYOZ1*	A/T	85.2	intergenic	1.19 (1.14-1.25)	3.7×10^{-13}	7.9×10^{-10}	rs10824026[k]	0.95
rs11598047	10q24/103,582,915	*NEURL1*	G/A	17.9	Intronic	1.15 (1.10-1.19)	1.5×10^{-10}	2.3×10^{-9}	rs12415501[l]	0.91
rs651386	1q24/170,622,169	*PRRX1*	A/T	57.4	intergenic	1.11 (1.08-1.15)	1.9×10^{-10}	7.9×10^{-10}	rs3903239[k]	0.61
rs1572226	6q22/118,313,916	*SLC35F1, PLN*	C/G	42.4	Intronic	1.11 (1.08-1.15)	3.1×10^{-10}	2.3×10^{-9}	rs4946333[m]	0.56

[a]SNP: single nucleotid polymorphism; [b]Chr: chromosome; [c]Pos = position, [d]RAF= risk allele frequency; [e]OR = odds ratio; [f]CI = 95% confidence interval; [g]Locus previously associating with AF but not genome-wide significantly.

References: [h](Ellinor et al., 2010), [i](Holm et al., 2011), [j](Benjamin et al., 2009), [k](Ellinor et al., 2012), [l] (Sinner et al., 2014), [m] (Christophersen et al., 2017)

4.1.1 Novel association of a coding variant in *PLEC* with AF (paper I)

The *PLEC* association was driven by a low frequency (1.2%) missense variant, p.Gly4098Ser (NP_958782.1), that associated with AF with an odds ratio (OR) of 1.55 (P = 8.0 × 10^{-10}, 95% CI 1.35 - 1.78, Table 3, Figure 5). The adjusted significance threshold for moderate impact variants was 5.1 × 10^{-8} (Sveinbjornsson et al., 2016). P.Gly4098Ser was the only variant in the region reaching genome-wide significance (Figure 5) and conditional analysis did not reveal additional signals in the region. Furthermore, no other coding variant in *PLEC* associated independently with AF (Appendix 1).

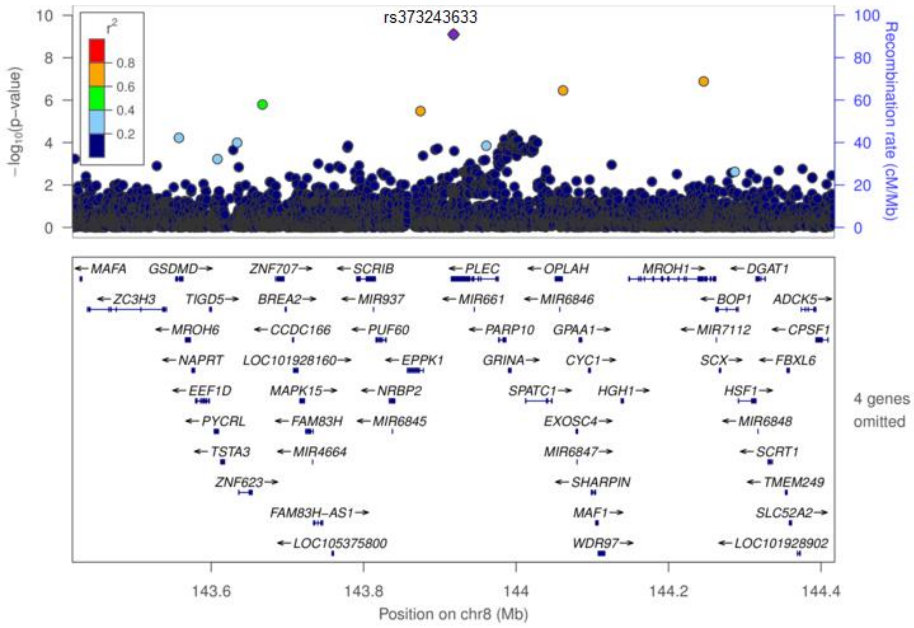


Figure 5. Regional plot of the *PLEC* locus on chromosome 8q24 (paper I). The plot depicts association results (P values) with AF (N = 13,471). The y-axis shows the $-\log_{10} P$ values and x-axis the genomic position (hg38). The leading variant is labelled as a diamond and coloured purple. Other variants are coloured according to correlation (r^2) with the leading marker.

PLEC encodes plectin, a very large (>500 kDa) multidomain cytoskeletal linking protein that can bind to all types of intermediate filaments, actin filaments and microtubules (Castanon et al., 2013). Plectin has a role in maintaining tissue integrity in skin, striated muscle and the heart (Wiche, 1998). It is expressed in many tissues, with equal expression in cardiac atria and ventricles ("Human genomics. The Genotype-Tissue Expression (GTEx)

pilot analysis: multitissue gene regulation in humans," 2015).

The p.Gly4098Ser missense variant (C>T position: 143,917,940) is located in exon 32/32 which encodes the C-terminal intermediate filament binding domain of plectin. The variant results in a glycine to serine substitution in the 5th and last type B plakin repeat motif (PRM) of the C-terminal (figure 6). This motif is homologous to the B-type PRM from desmoplakin whose 3D crystal structure is avaliable (Choi et al., 2002). As shown in Figure 6, the glycine in desmoplakin homologous to Gly4098 in plectin is found on the protein surface.

In general, the *PLEC* gene is tolerant to missense mutations (constrained metric z-score = -0.38 (Lek et al., 2016)), however the region in question is conserved and has low tolerance to mutations (GERP-score = 4.2 (ranges from -12.3 to 6.17, with 6.17 being the most conserved), subRVIS score < 35th percentile (Cooper et al., 2005; Gussow et al., 2016)). Furthermore, according to the PolyPhen-2 and SIFT predicting tools the amino acid substitution is potentially damaging for the protein structure and function (PolyPhen-2: 1 (range: 0 - 1, where 1 is most deleterious), SIFT: 0.02 (range: 0 – 1, cutoff for deleterious classification <= 0.05) (Adzhubei et al., 2010; Kumar et al., 2009)). The closest functional domain to p.Gly4098Ser is a suggested interaction site with intermediate filaments beween the 5th and 6th PRMs (Castanon et al., 2013).

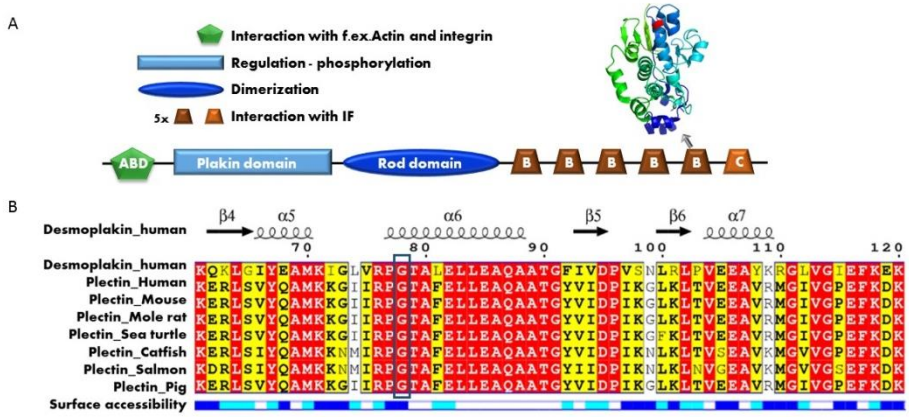


Figure 6. A schematic domain structure of plectin. The C-terminal intermediate filament binding domain is comprised of six plakin repeat motifs (PRM), five of type B and one of type C (Castanon et al., 2013; Huber, 2003). P.Gly4098Ser is located in the 5th and last type B PRM. The 3D crystal structure of the homologous B-type PRM from desmoplakin is used here to show the variants location as a red dot on the protein surface (Choi et al., 2002).

Seventeen of the 151,677 genotyped study participants were homozygous for the p.Gly4098Ser variant in *PLEC,* consistent with what was expected under Hardy-Weinberg equilibrium (22 expected, P = 0.52). Their mean age was 46 (range 11 - 80 years). None of the homozygotes had been diagnosed previously with AF according to available information but this did not deviate from the additive model (P = 0.093). It is possible that some of the homozygotes had undiagnosed AF or will develop AF later in life. P.Gly4098Ser did not associate with AF (P = 0.37) or ECG traits under the recessive model (lowest P value 0.0069 in T amplitude lead V3, beta = -1.25).

We looked up the p.Gly4098Ser variant in the Genome Aggregation Database (gnomAD), a large publicly available reference dataset of human genetic variation (Karczewski et al., 2019). The variant is exceedingly rare outside of Iceland and was found in only 13 of the 140,842 individuals in gnomAD (12 European, one South-Asian). The higher frequency of p.Gly4098Ser in Iceland is likely explained by the founder effect, the fact that a small number of ancestors accounts for a relatively large proportion of the population. Acknowledging the limited power to detect association, the variant was genotyped in two AF sample sets, from the Vanderbilt AF registry (764 cases and 762 controls) and the FOURIER trial (1,238 cases and 11,562 controls). Three carriers of p.Gly4098Ser were identified and all had AF.

To further explore the effects of p.Gly4098Ser, it was tested for association with all phenotypes in deCODE genetics extensive phenotype database. The most significant disease associations were with the AF related traits SSS (OR = 1.64, 95% CI = 1.31 - 2.05, $P = 1.7 \times 10^{-5}$), pacemaker implantation (OR = 1.54, 95% CI = 1.24 - 1.92, $P = 9.5 \times 10^{-5}$), ischemic stroke (OR = 1.22, 95% CI = 1.01 - 1.47, P = 0.035) and the ischemic stroke sub-phenotype cardioembolic stroke (OR = 1.53, 95% CI = 1.09 - 2.14, P = 0.013) (Table 4). Table 4 shows the most significant secondary associations of p.Gly4098Ser and associations with relevant cardiovascular phenotypes and phenotypes described in plectinopathies, such as epidermolysis bullosa simplex (Winter & Wiche, 2013). The only other significant association among these ($P < 0.05/16 = 0.003$) was with decreased levels of creatine kinase (beta = -0.09, 95% CI = -0.14 - -0.04, $P = 2.6 \times 10^{-4}$).

Table 4. Association of p.Gly4098Ser in *PLEC* with secondary phenotypes.

Phenotype	**N**	**Effect (OR/beta[a]) (CI[b])**	***P* value**
Sick sinus syndrome	3,310	1.64 (1.31-2.05)	**1.7×10^{-5}**
Pacemaker implantation	3,578	1.54 (1.24-1.91)	**9.5×10^{-5}**
Creatine kinase	71,277	-0.09 (-0.14- -0.04)	**2.6×10^{-4}**
Cardioembolic stroke	1,369	1.53 (1.09-2.14)	0.013
Ischemic stroke	5,626	1.22 (1.01-1.47)	0.035
Coronary artery disease	37,782	1.10 (0.99-1.23)	0.087
Type II diabetes	11,448	0.86 (0.72-1.03)	0.10
Aortic valve stenosis	1,718	0.81 (0.55-1.19)	0.28
Hypertension	25,577	0.94 (0.83-1.06)	0.31
Dilated cardiomyopathy	424	1.38 (0.73-2.60)	0.32
Thyroid stimulating hormone	188,057	0.02 (-0.03-0.07)	0.40
Ventricular tachycardia	945	1.17 (0.75-1.83)	0.49
Congenital heart disease	1,804	0.86 (0.54-1.38)	0.53
Hypertrophic cardiomyopathy	372	1.14 (0.56-2.33)	0.72
Heart failure	15,237	1.03 (0.87-1.22)	0.73
Second and third degree atrio-ventricular block	1,303	0.97 (0.67-1.40)	0.87

P values below 0.0031 (0.05/16) are in **bold**

[a]Estimated odds ratio (OR) or effect in standard deviation; [b]CI = 95% confidence interval

4.1.2 An intergenic variant between *METTL11B* and *LINC01142* increases risk of AF (paper I)

In addition to the *PLEC* association, paper I reported a second novel AF association with an intergenic variant of 8.1% frequency, rs72700114 (g.170224684G>C), located between the genes *LINC01142* and *METTL11B* on chromosome 1q24. This variant associated with AF with OR = 1.22 (P = 7.0 × 10^{-12}, 95% CI 1.15 - 1.29) in Iceland (Table 3). The association replicated well in two replication sample sets from the Vanderbilt AF registry (764 cases and 762 controls) and the FOURIER trial (1,238 cases and 11,562 controls). The direction of effect was consistent in all cohorts and the combined OR was 1.26 (95% CI 1.19 - 1.32) and P value 3.1 × 10^{-18} (Appendix 2).

Rs72700114 is located 400 kb away from a previously reported AF locus, represented by a common variant, rs651386 (frequency = 57.4%), close to *PRRX1* (Figure 7) (Ellinor et al., 2012). According to conditional analysis, the two associations are independent (Appendix 3) and the two variants not strongly correlated (D' = 0.38, r^2 = 0.009). This is not the first discovery of a secondary AF signal at this locus. A recent publication reported an association with rs72700118 (Christophersen et al., 2017) located 1 kb away from rs72700114, but not in LD with it (D' = 1, r^2 = 0.014). None of the three variants (rs72700114, rs651386 and rs72700118) associated with the expression of nearby genes in deCODEs RNA samples from blood (2,528 samples) or adipose tissue (686 samples).

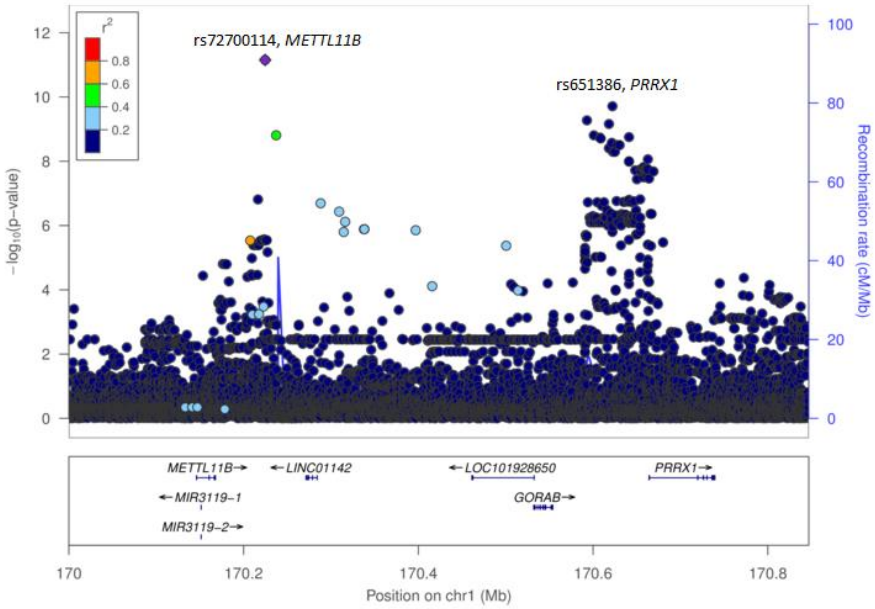


Figure 7. Regional plot of the AF signal on chromosome 1q24, rs72700114 between *METTL11B* and *LINC01142*, and a previously reported signal 400 kb away, rs651386, close to *PRRX1*. The plot depicts association results (*P* values) with AF (N = 13,471). For each plot, the y-axis shows the $-\log_{10}$ *P* values and x-axis the genomic position (hg38). The leading variant (rs72700114) is labelled as diamond and coloured purple. Other variants are coloured according to correlation (r^2) with rs72700114.

To further explore the effects of rs72700114 at *LINC01142*/*METTL11B*, the variant was tested for association with all phenotypes in deCODE genetics phenotype database. After correcting for multiple testing considering relevant phenotypes ($P < 0.05/16 = 0.003$) the only significant secondary association was with risk of heart failure (OR = 1.10, 95% CI = 1.04 - 1.17, $P = 0.0016$) (Table 5).

Table 5. Association of rs72700114 at *LINC01142/METTL11B* with secondary phenotypes.

Phenotype	N	Effect (OR/beta[a]) (CI[b])	*P* value
Heart failure	15,237	1.10 (1.04, 1.17)	**0.0016**
Creatine kinase	71,277	-0.03 (-0.05, -0.01)	0.0058
Pacemaker implantation	3,578	1.13 (1.03, 1.24)	0.0077
Cardioembolic stroke	1,369	1.18 (1.02, 1.36)	0.022
Sick sinus syndrome	3,310	1.08 (0.98, 1.19)	0.11
Aortic valve stenosis	1,718	1.12 (0.97, 1.29)	0.12
Second and third degree atrio-ventricular block	1,303	1.11 (0.96, 1.28)	0.16
Ischemic stroke	5,626	1.05 (0.98, 1.13)	0.18
Hypertension	25,577	1.03 (0.98, 1.09)	0.29
Coronary artery disease	37,782	1.02 (0.98, 1.06)	0.31
Dilated cardiomyopathy	424	1.13 (0.87, 1.47)	0.36
Thyroid stimulating hormone	188,057	-0.00 (-0.02, 0.01)	0.53
Ventricular tachycardia	945	1.04 (0.86, 1.26)	0.69
Type II diabetes	11,448	1.01 (0.96, 1.06)	0.70
Hypertrophic cardiomyopathy	372	0.97 (0.65, 1.44)	0.88
Congenital heart disease	1,804	1.01 (0.83, 1.23)	0.92

P values below 0.0031 (0.05/16) are in **bold**

[a]Estimated odds ratio (OR) or effect in standard deviation; [b]CI = 95% confidence interval

4.1.3 Replication of previously reported AF signals (paper I)

In addition to the two novel loci described in paper I (*PLEC* and *LINC01142*/*METTL11B*) eight previously reported loci reached genome-wide significance (Figure 4, Table 3,). One of those had associated with AF as a secondary trait before, a missense variant in *MYH6* previously reported for association with SSS (Holm et al., 2011). It is now the second strongest AF signal in the Icelandic dataset. A total of 29 previously reported loci (at the time) were replicated either with genome-wide significance (Table 3) or nominal significance (Table 6). These had been associated with AF in GWAS either as a primary trait or secondary trait, following association with SSS, heart rate or PR interval. Two of the variants replicated here (rs12044963 at *KCND3* and rs2047036 at *SH3PXD2A*) were reported in GWAS on AF among Japanese before and nominally associated with AF in Europeans (Low et al., 2017). Additional variants reported in Japanese cohorts were either not replicated (Table 6) or very rare/nonexistent in the Icelandic population (rs17059534 in *HAND2*, rs2540953 in *SLC1A4–CEP68* and rs2296610 in *NEBL*; (Low et al., 2017)). In total four previously reported loci were tested and did not reach nominal significance in this study (Table 6).

Table 6. Replication of loci previously reported to associate with AF at the time of publication of paper I, Oct 24th 2017 (N = 13,471).

Reference SNP[a] ID	Chr[b]/Pos[c] (hg38)	Closest genes	Risk allele/ other	RAF (%)[d]	Coding effect	OR[e] (CI[f])	*P* value
Variants reported to associate genome-wide significantly with AF							
rs3807989[g]	7q31/116,546,187	*CAV1*	G/A	59.5	intronic	1.11 (1.07-1.14)	**2.5×10^{-9}**
rs10821415[g]	9q22/94,951,177	*C9orf3*	A/C	38.5	intronic	1.10 (1.06-1.14)	**3.8×10^{-8}**
rs35176054[h]	10q24/103,720,629	*SH3PXD2A*	A/T	13.6	intronic	1.14 (1.08-1.19)	**9.3×10^{-8}**
rs72700118[h]	1q24/170,225,682	*METTL11B, LINC01142*	A/C	13.6	intergenic	1.12 (1.07-1.17)	**$2,9\times10^{-6}$**
rs10507248[i]	12q24/114,359,288	*TBX5*	T/G	70.2	intronic	1.09 (1.05-1.13)	**3.0×10^{-6}**
rs7164883[g]	15q24/73,359,833	*HCN4*	G/A	14.7	intronic	1.12 (1.07-1.17)	**1.1×10^{-6}**
rs7508[h]	8p22/18,056,461	*ASAH1*	A/G	71.5	3´ UTR	1.09 (1.05-1.13)	**1.2×10^{-5}**
rs1152591[g]	14q23/64,214,130	*SYNE2*	A/G	51.0	upstream	1.07 (1.04-1.11)	**3.5×10^{-5}**
rs13216675[i]	6q22/122,131,183	*GJA1*	T/C	66.2	intergenic	1.08 (1.04-1.11)	**3.5×10^{-5}**
rs11047543[h]	12p12/24,635,405	*SOX5*	G/A	86.7	intergenic	1.11 (1.05-1.16)	**5.3×10^{-5}**
rs2540949[h]	2p14/65,057,097	*CEP68*	A/T	59.3	intronic	1.07 81.03-1.10)	**1.5×10^{-4}**
rs12044963[j]	1p13/111,849,738	*KCND3*	T/G	9.4	intronic	1.11 (1.05-1.17)	**2.1×10^{-4}**
rs4642101[i]	3p25/12,800,724	*CAND2*	G/T	65.8	intronic	1.07 (1.03-1.10)	**3.1×10^{-4}**
rs6800541[h]	3p22/38,733,341	*SCN10A*	T/C	63.6	intronic	1.06 (1.03-1.10)	**4.3×10^{-4}**
rs3771537[h]	2p13/69,811,660	*ANXA4*	A/C	53.1	intronic	1.06 (1.03-1.10)	**4.6×10^{-4}**
rs75190942[h]	11q24/128,894,676	*KCNJ5*	A/C	10.6	down-stream	1.09 (1.03-1.15)	**0.0016**
rs337711[h]	5q22/114,412,874	*KCNN2*	T/C	34.1	intronic	1.05 (1.01-1.08)	**0.0071**
rs2047036[j]	10q24/103,717,405	*SH3PXD2A*	C/T	40.1	intronic	1.04 (1.01-1.08)	**0.012**
rs2288327[h]	2q31/178,546,938	*TTN*	G/A	14.0	intronic	1.06 (1.01-1.11)	**0.017**
rs2967791[h]	5q31/137,677,417	*KLHL3*	T/C	56.7	intronic	1.04 (1.00-1.07)	**0.025**
rs7698692[j]	4q34/173,682,953	*HAND2*	G/A	2.1	intergenic	1.10 (0.99-1.23)	0.089
rs17461925[j]	1q32/203,057,463	*PPFIA4*	A/G	66.2	intronic	1.01 (0.97-1.04)	0.69
rs6490029[i]	12q24/111,260,653	*CUX2*	A/G	29.8	intronic	1.00 (0.98-1.02)	1.00
Variants reported to associate with AF as a secondary trait, following association with heart rate and/or PR interval							
rs6882776[k]	5q35/173,237,160	*NKX2-5*	G/A	71.8	upstream	1.09 (1.05-1.13)	**8.1×10^{-6}**
rs11708996[l]	3p22/38,592,432	*SCN5A*	G/C	88.2	intronic	1.04 (0.99-1.09)	0.14

Significant *P* values (<0.05) are in **bold**

[a]SNP: single nucleotid polymorphism; [b]Chr: chromosome; [c]Pos = position, [d]RAF= risk allele frequency; [e]OR = odds ratio; [f]CI = 95% confidence interval. References: [g](Ellinor et al., 2012), [h](Christophersen et al., 2017), [i](Sinner et al., 2014), [j](Low et al., 2017), [k](den Hoed et al., 2013), [l](Pfeufer et al., 2010)

4.2 A meta-analysis of GWASs on AF in Iceland and the UK Biobank focusing on low frequency coding and splice variants (paper II)

The second paper reports a meta-analysis of 14,710 cases and 373,897 controls from Iceland and 14,792 cases and 393,863 controls from the UK Biobank, with follow-up in samples from Norway and the US. This study focused on low frequency coding and splice variants. Three novel associations of coding variants were discovered, two in the gene *RPL3L* and one in *MYZAP* (Table 7).

Table 7. Results of meta-analysis of GWASs on AF in Iceland (N=14,710) and UK Biobank (N = 14,792) with replication in cohorts from the US and Norway, focusing on rare and low frequency coding and splice variants (paper II).

	***MYZAP* p.Gln254Pro Missense**			***RPL3L* p.Ala75Val Missense**			***RPL3L* c.1167+1G>A splice-donor**		
Rs name	rs147301839			rs140185678			rs140192228		
Position (hg38)	chr15:57632516			chr16:1953015			chr16:1945498		
Dataset (cases/controls)	[a]**RAF (%)**	[b]**OR ([c]CI)**	***P***	[a]**RAF (%)**	[b]**OR ([c]CI)**	***P***	[a]**RAF (%)**	[b]**OR ([c]CI)**	***P***
Iceland (14,710/373,897)	1.08	1.38 (1.20-1.60)	9.0×10^{-6}	3.65	1.18 (1.09-1.28)	6.4×10^{-5}	0.61	1.37 (1.14-1.65)	8.7×10^{-4}
UK Biobank (14,792/393,863)	0.36	1.32 (1.10-1.57)	0.0023	3.37	1.20 (1.13-1.27)	1.2×10^{-8}	0.31	1.71 (1.40-2.07)	6.9×10^{-8}
Meta-analysis: Iceland and UK		**1.36 (1.21-1.52)**	**7.8×10^{-8}**		**1.19 (1.13-1.25)**	**3.4×10^{-12}**		**1.52 (1.33-1.74)**	**8.2×10^{-10}**
Fourier (1,238/11,562)	0.21	1.61 (0.46-5.69)	0.46	4.88	0.93 (0.55-1.59)	0.79	0.12	1.59 (0.13-20.01)	0.72
Vanderbilt (759/759)	0.46	7.03 (1.29-38.23)	0.024	3.00	1.33 (0.88-2.02)	0.18	0.43	1.61 (0.46-5.69)	0.40
Tromsø (714/698)	0.49	1.14 (0.39-3.32)	0.81	3.90	1.29 (0.87-1.90)	0.20	0.073	1.60[d] (0.53-4.88)	0.30
HUNT (6,493/63,142)	0.64	1.47 (1.14-1.89)	0.0027	2.87	1.22 (1.07-1.40)	0.0038	0.15	1.19 (0.71-2.00)	0.51
Replication only (Fourier, Vanterbilt, Tromsø and HUNT)		**1.50 (1.18-1.91)**	**8.1×10^{-4}**		**1.22 (1.08-1.37)**	**0.0011**		**1.33 (0.88-2.00)**	**0.17**
Combined		**1.38 (1.25-1.53)**	**3.3×10^{-10}**		**1.20 (1.14-1.25)**	**1.7×10^{-14}**		**1.50 (1.32-1.70)**	**5.0×10^{-10}**

[a]RAF= risk allele frequency; [b]OR = odds ratio; [c]CI = 95% confidence interval

[d]No control was a carrier of the variant resulting in 2×2 table entry with a zero count. Effect and *P* value was calculated with the R-package „metafor" by adding a small constant of ½ to the cells of the 2×2 table.

4.2.1 A missense variant in *MYZAP* increases risk of AF (paper II)

One of the three novel associations reported in paper II was with a missense variant, p.Gln254Pro, in the *MYZAP* gene on chromosome 15 (allele frequency 1.08% in Iceland, 0.36% in the UK). In the meta-analysis of Icelandic and UK Biobank data the association was suggestive (OR = 1.36, 95% CI = 1.21 - 1.52, $P = 7.8 \times 10^{-8}$, Table 7), as it did not reach the applied threshold for genome-wide significance for moderate impact variants ($P < 5.1 \times 10^{-8}$). To further assess the association, the variant was tested in four additional sample sets of 9,204 cases and 76,161 controls combined, from the Nord-Trøndelag Health Study (HUNT), the FOURIER trial, the Vanderbilt Atrial Fibrillation Registry, and the Tromsø Study. Joint analysis of all datasets yielded a genome-wide significant association of p.Gln254Pro with AF (OR = 1.38, 95% CI = 1.25 - 1.53, $P = 3.3 \times 10^{-10}$) (Table 7). No other coding variant in *MYZAP* associated independently with AF in a conditional analysis conducted with the Icelandic data (Appendix 4).

MYZAP encodes myozap, myocardial zonula adherens protein, primarily expressed in the heart in man and its homolog in the mouse has been localized to the intercalated discs (Seeger et al., 2010). The p.Gln254Pro variant is located close to the myozap protein region involved in both actin colocalization and activation of serum response factor (SRF)-dependent transcription (amino acids 91-250) (Seeger et al., 2010). The variant was looked up in the PROVEAN web server, a tool for prediction of the functional effect of amino acid substitutions, based on comparisons with related homologous sequences (Choi & Chan, 2015). According to PROVEAN, Gln254 is a conserved amino acid and the variant is predicted to be deleterious (Table 8).

Table 8. Prediction scores for the functional effect of amino acid substitutions and indels from PROVEAN (Choi & Chan, 2015) for missense variants in *MYZAP* and *RPL3L* associating with risk of AF (paper II).

	***MYZAP* p.Gln254Pro missense**	***RPL3L* p.Ala75Val missense**
Number of clusters	30	30
Number of supporting sequences used	166	470
PROVEAN score	-4.037	-3.492
Predicted effect	Deleterious	Deleterious
PROVEAN score in RPL3	-	-3.504

Default cutoff for Deleterious classification = -2.5.

Specificity = sensitivity = 80%.

To further explore the effects of the novel AF variant in *MYZAP* it was tested for association with other phenotypes in deCODE´s genotype/phenotype database under the additive model (Table 9). The most significant of these was with the related atrial arrhythmia SSS (OR = 1.51, 95% CI = 1.20 - 1.89, $P = 3.5 \times 10^{-4}$). The associations with SSS and pacemaker implantation were the only ones passing Bonferroni correction accounting for multiple testing of 15 relevant phenotypes ($P < 0.0033 = 0.05/15$). In a combined SSS sample set from Iceland (N = 3,568, controls = 346,025) and the UK Biobank (N = 403, controls = 403,181), p.Gln254Pro associated with SSS with an OR of 1.65 (95% CI = 1.33 - 2.05) and $P = 5.0 \times 10^{-6}$ (Appendix 5). Since mutations in intercalated disc genes, albeit not *MYZAP*, have been associated with cardiomyopathies in man (Moncayo-Arlandi & Brugada, 2017) the link between the *MYZAP* variant and cardiomyopathies was specifically assessed in the deCODE dataset, but none passed Bonferroni correction (Table 9).

Table 9. Association of p.Gln254Pro in *MYZAP* with secondary phenotypes.

Phenotype	N	Effect (OR/beta[a]) (CI[b])	*P* value
Sick sinus syndrome	3,568	1.51 (1.20-1.89)	**0.00035**
Pacemaker implantation	3,578	1.41 (1.12-1.77)	**0.0031**
Cardioembolic stroke	1,369	1.30 (0.89-1.89)	0.17
Ischemic stroke	5,626	1.09 (0.89-1.34)	0.41
Coronary artery disease	37,782	1.17 (1.04-1.32)	0.0096
Type II diabetes	11,448	0.97 (0.82-1.15)	0.73
Aortic valve stenosis	2,457	1.00 (0.45-2.17)	0.98
Hypertension	54,974	1.00 (0.45-2.17)	0.98
Dilated cardiomyopathy	424	1.28 (0.64-2.58)	0.49
Thyroid stimulating hormone	188,175	-0.01 (-0.04-0.02)	0.45
Ventricular tachycardia	945	1.06 (0.66-1.70)	0.81
Congenital heart disease	2,097	0.88 (0.58-1.32)	0.54
Hypertrophic cardiomyopathy	372	1.28 (0.62-2.62)	0.50
Heart failure	15,237	1.05 (0.91-1.21)	0.49
Second and third degree atrio-ventricular block	1,303	1.26 (0.86-1.85)	0.24

P values below 0.0033 (0.05/15) are in **bold**

[a]Estimated odds ratio (OR) or effect in standard deviation; [b]CI = 95% confidence interval

4.2.2 Two coding variants in *RPL3L* increase risk of AF (paper II)

The other two novel associations reported in paper II were with coding variants in the gene *RPL3L*. Joint analysis of all datasets yielded genome-wide significant associations with AF of the missense variant p.Ala75Val (OR = 1.20, 95% CI = 1.14 - 1.25, $P = 1.7 \times 10^{-14}$) and the splice donor variant c.1167+1G>A (OR = 1.50, 95% CI = 1.32 - 1.70, $P = 5.0 \times 10^{-10}$) (Table 7). The two *RPL3L* variants are not correlated (D' = 1, $r^2 = 0.00024$), and when conditioned on each other in the Icelandic dataset, both associations with AF remained (Appendix 6). All 15 low frequency coding variants found in *RPL3L* in the deCODE data were then tested for association with AF after conditioning on the two novel genome-wide significant variants, p.Ala75Val and c.1167+1G>A (Appendix 7). One variant associated with AF with a *P* value below the Bonferroni-corrected significance threshold of 0.0033 (0.05/15), but the association was not genome-wide significant in the meta-analysis. The *RPL3L* gene encodes a ribosomal protein (ribosomal protein like 3L) that is primarily expressed in skeletal muscle and the heart unlike most ribosomal proteins, that are ubiquitously expressed (Van Raay et al.,

1996).

The *RPL3L* variants did not associate with other phenotypes related to AF (Tables 10 and 11) Furthermore, when combining the SSS sample sets from Iceland (N = 3,568, controls = 346,025) and the UK Biobank (N = 403, controls = 403,181) neither of the two *RPL3L* variants associated with SSS (Appendix 5). Since mutations in ribosomal genes are commonly associated with bone marrow failure (Raiser et al., 2014), the relationship between the *RPL3L* variants and blood cells was specially queried but no associations observed.

Table 10. Association of p.Ala75Val in *RPL3L* with secondary phenotypes.

Phenotype	**N**	**Effect (OR/beta[a]) (CI[b])**	***P* value**
Sick sinus syndrome	3,568	1.02 (0.85-1.22)	0.83
Pacemaker implantation	3,578	1.03 (0.90-1.17)	0.66
Cardioembolic stroke	1,369	1.22 (0.99-1.50)	0.059
Ischemic stroke	5,626	1.03 (0.92-1.15)	0.60
Coronary artery disease	37,782	1.01 (0.96-1.06)	0.71
Type II diabetes	11,448	1.00 (0.90-1.11)	0.94
Aortic valve stenosis	2,457	0.93 (0.77-1.12)	0.45
Hypertension	54,974	0.96 (0.90-1.02)	0.22
Dilated cardiomyopathy	424	0.81 (0.53-1.25)	0.34
Thyroid stimulating hormone	188,175	0.00 (-0.01-0.02)	0.80
Ventricular tachycardia	945	0.84 (0.64-1.10)	0.21
Congenital heart disease	2,097	1.23 (0.98-1.54)	0.071
Hypertrophic cardiomyopathy	372	0.61 (0.35-1.05)	0.074
Heart failure	15,237	1.05 (0.96-1.14)	0.26
Second and third degree atrio-ventricular block	1,303	0.92 (0.72-1.17)	0.50

P value threshold = 0.0033 (0.05/15)

[a]Estimated odds ratio (OR) or effect in standard deviation; [b]CI = 95% confidence interval

Table 11. Association of c.1167+1G>A in *RPL3L* with secondary phenotypes.

Phenotype	N	Effect (OR/beta[a]) (CI[b])	*P* value
Sick sinus syndrome	3,568	0.97 (0.71-1.33)	0.85
Pacemaker implantation	3,578	1.31 (0.97-1.77)	0.078
Cardioembolic stroke	1,369	1.46 (0.94-2.27)	0.093
Ischemic stroke	5,626	1.30 (1.02-1.66)	0.036
Coronary artery disease	37,782	1.00 (0.85-1.17)	0.97
Type II diabetes	11,448	1.01 (0.83-1.23)	0.92
Aortic valve stenosis	2,457	1.28 (0.88-1.87)	0.20
Hypertension	54,974	1.07 (0.93-1.23)	0.33
Dilated cardiomyopathy	424	1.05 (0.41-2.72)	0.92
Thyroid stimulating hormone	188,175	0.00 (-0.03-0.03)	0.95
Ventricular tachycardia	945	0.67 (0.33-1.36)	0.27
Congenital heart disease	2,097	0.94 (0.57-1.56)	0.81
Hypertrophic cardiomyopathy	372	1.00 (0.21-4.77)	1.00
Heart failure	15,237	1.26 (1.06-1.50)	0.0096
Second and third degree atrio-ventricular block	1,303	1.42 (0.88-2.29)	0.15

P value threshold = 0.0033 (0.05/15)

[a]Estimated odds ratio (OR) or effect in standard deviation; [b]CI = 95% confidence interval

Prediction scores for consequences of the p.Ala75Val amino acid substitution are conflicting. The variant is predicted to be deleterious for protein function by PROVEAN (PROVEAN prediction score = -3.492, Table 8) (Choi & Chan, 2015)), but not by CADD score (CADD score = 26, range 1 - 99, 99 being most deleterious) (Kircher et al., 2014). Ala75 is highly conserved over a range of species, in both RPL3L and its homologous protein, RPL3 (GERP conservation score = 5.28, range: -12.3 - 6.17, with 6.17 being the most conserved) (Cooper et al., 2005).

The c.1167+1G>A variant in *RPL3L* is located in a splice donor site by the 9th and second to last exon of *RPL3L*. To assess the variants effects on transcription, RNA samples from cardiac atria of 167 Icelanders were used. Two of the 167 individuals were heterozygous carriers of the variant and their RPL3L transcripts were compared to those of non-carriers. Quantification of the transcripts showed that non-carriers only produced the primary RPL3L isoform, but both carriers also produced an alternative isoform that skips exon 9 (*P* = 0.0052, Figure 8, panel a). It was also evident that carriers expressed the two isoforms in approximately equal abundance. Exon 9 is the second to last exon in RPL3L and is 120 base pairs long and therefore its deletion is in-frame (Figure 8, panel b).

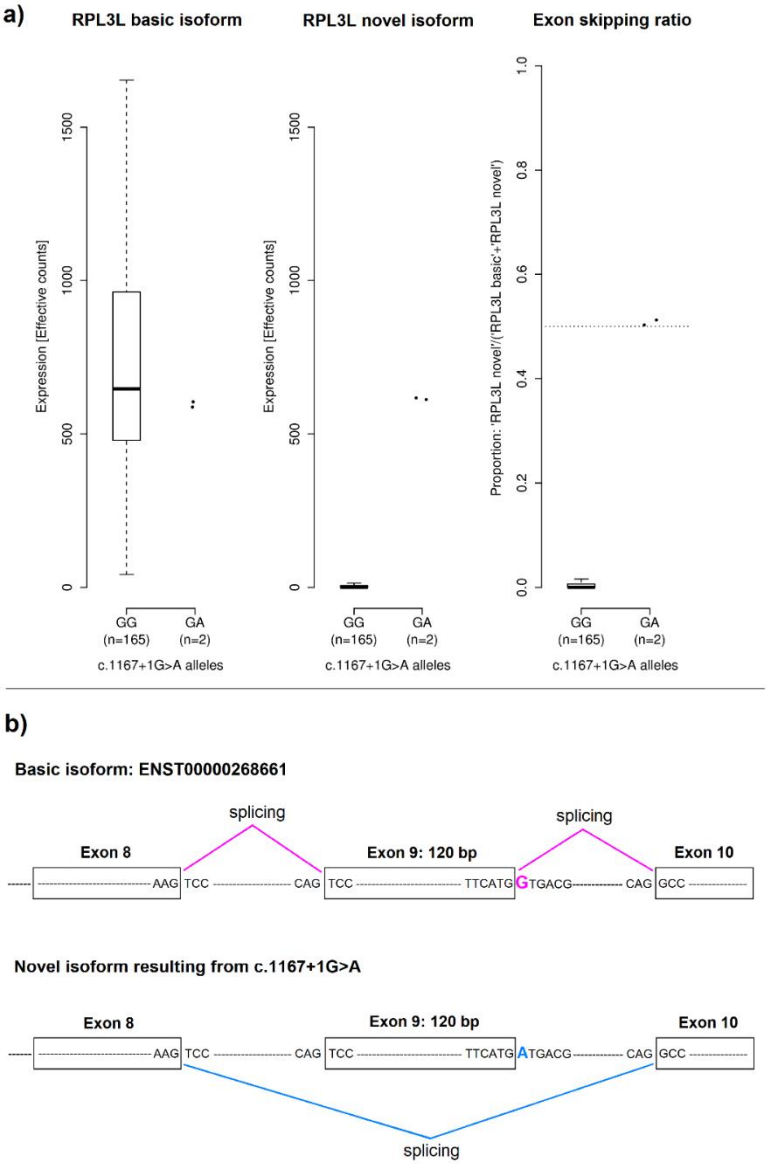


Figure 8. The effects of c.1167+1G>A in *RPL3L* on transcription in cardiac atria. Panel a shows quantification of two forms of RPL3L transcripts; the primary isoform, ENST00000268661, and a novel isoform with skipping of exon 9 resulting from c.1167+1G>A. It also shows the proportion of novel isoform among all transcripts. A total of 167 samples, all from the right atrium, were included in the analysis, two of which came from carriers of c.1167+1G>A. The figure demonstrates that only the two carriers have the novel isoform with skipping of exon 9. Their exon skipping proportion is approximately 0.5 while it is zero in non-carriers. Panel b is a schematic illustration of the splicing of RPL3L among carriers and noncarriers of c.1167+G>A. The variant

is in a splice donor site by the second last exon and results in exon skipping. The skipped exon is 120 base pairs and therefore its deletion is in-frame.

4.3 Associations of novel and reported AF variants with ECG measurements (papers I and II).

In papers I and II the novel and reported AF variants were tested for association with ECG traits that reflect electrophysiologic function of the heart (Figure 9). A total of 289,297 sinus rhythm ECGs from 62,974 individuals without the diagnosis of AF were included in the analysis. All variants were tested for association with 122 ECG variables, some of them correlated (Table 2).

The novel AF variants reported in papers I and II have diverse effects on cardiac electrical function in sinus rhythm. The p.Gly4098Ser variant in *PLEC* associates with many ECG measures, independent of AF diagnoses (Figure 9). The variant affects most components of the cardiac cycle, including the P wave amplitude and area. It also prolongs the PR segment, representing AVN conduction, and lowers the R wave amplitude. The coding variant in *MYH6*, another structural gene, has a similar general effect on cardiac conduction, albeit a relatively stronger effect on the atria and AVN than the ventricles. The *RPL3L* missense variant p.Ala75Val also affects ECG measures in the absence of AF (Figure 9). It associates with measures of atrial conduction, both P wave amplitude and area, and with shorter QRS duration. Neither of the other variants in *RPL3L* and *MYZAP* associate with ECG traits in sinus rhythm. When testing for association with ECG traits using all ECGs irrespective of rhythm and history of AF, both p.Gly4098Ser in *PLEC* and p.Ala75Val in *RPL3L* associate more significantly with ECG measurements and p.Gln254Pro in *MYZAP* associates with various P wave indices, R amplitude and T wave indices (Appendix 8). The novel intergenic AF variant at *LINC01142/METTL11B* did not associate with ECG measurements, with or without AF cases in the analysis (Figure 9, Appendix 8).

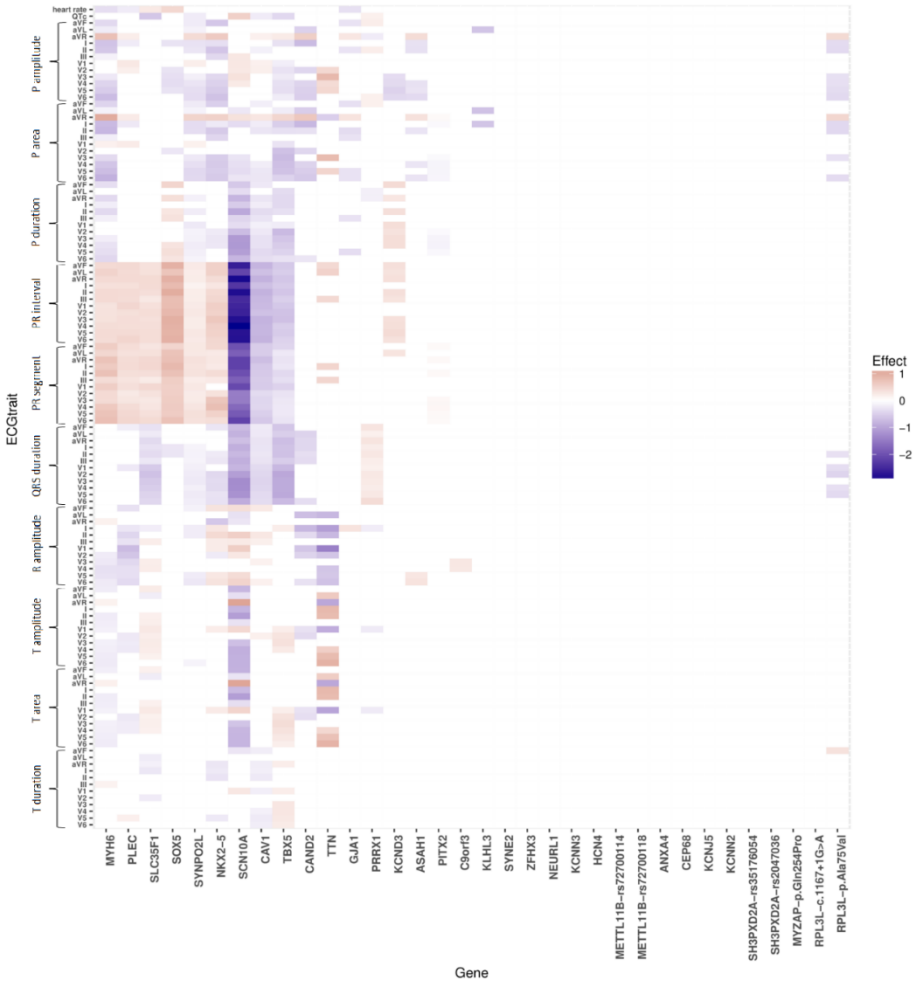


Figure 9. Heatmap showing the effects of AF variants on ECG traits in sinus rhythm ECGs, excluding AF cases (papers I and II). ECG measurements were available for 62,974 individuals without AF. Each column shows the estimated effect of the risk allele of an AF variant on various ECG traits. The effect of each variant, annotated with the corresponding gene name, is scaled with the $\log_{10}$-AF odds ratio. Red color represents a positive effect on the ECG variable and blue color a negative effect. The effect is shown only for significant associations after adjusting for multiple testing with a false discovery rate procedure for each variant. Non-significant associations are white in the heatmap.

In general, AF variants have highly diverse effects on ECG traits in individuals unaffected by AF, ranging from none to extensive and varying in direction. For example, the PR interval is an established predictor of AF development (Cheng et al., 2009) and it was either prolonged, shortened or not affected by AF risk alleles. Similarly, there is no clear relationship

between magnitudes of variants effects on conduction and their effects on AF. For example, the variants in *NKX2-5* and *SCN10A* had extensive and strong ECG effects but small AF effects compared to the strongest common AF variant in *PITX2*. The *PITX2* variant in turn had relatively weak effects on cardiac conduction in the absence of AF and it only affected indices of atrial and AVN conduction with no observable effect on ventricular conduction. Finally, many variants including the AF variants with moderate effects in the potassium channel gene *KCNN3* and the transcription factor gene *ZFHX3* had no association with ECG measurements in the absence of AF.

4.4 Effects of novel and reported AF variants on pacemaker implantation, SSS and stroke risk (papers I and II)

To gain insight into the causal relationship between AF and related traits, AF variants were tested for association with SSS, pacemaker implantation, ischemic stroke, and cardioembolic stroke. The AF risk of each variant was then plotted against the effect on the risk of the AF related trait (Figure 10). This was done in paper I, before the discovery of coding variants in *RPL3L* and *MYZAP* associating with AF. They have, however, been added here for completion. The coding AF variant in *MYH6* that was originally discovered through its association to high risk of SSS (Holm et al., 2011) stood out with substantially greater risk of SSS and pacemaker implantation than predicted from its effect on AF risk. For all other variants, their effects on the AF related traits were consistent with being proportional to their effects on AF.

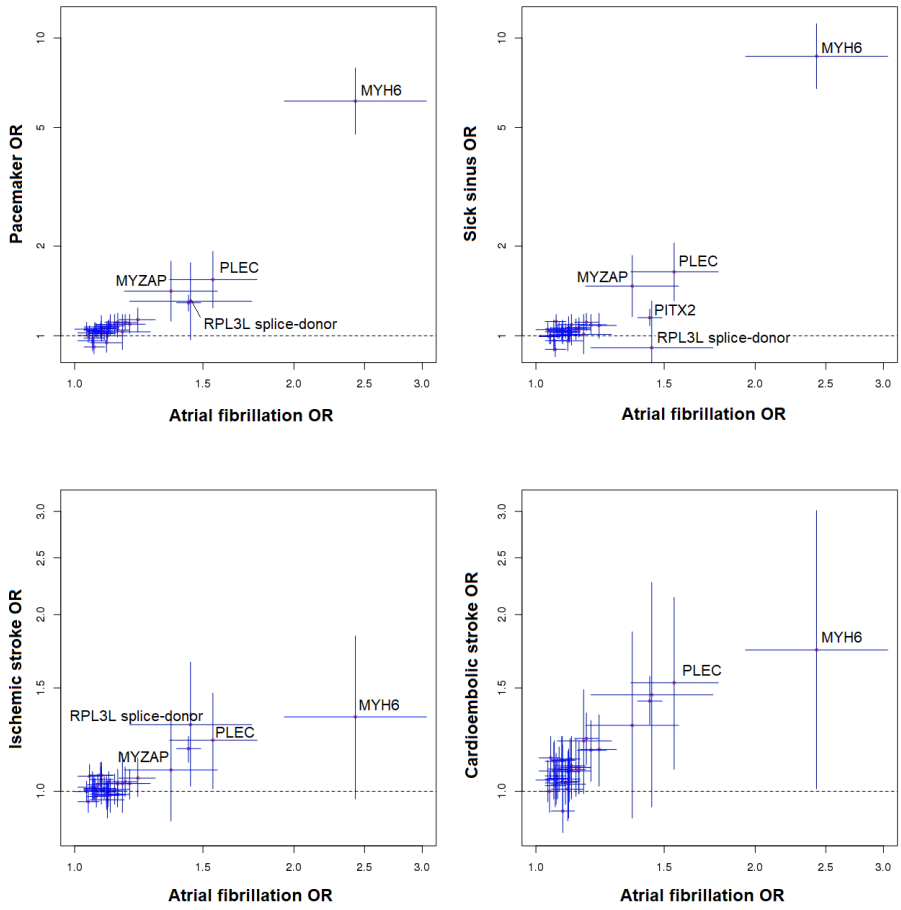


Figure 10. AF effects of AF associated variants plotted against the effect on (a) pacemaker implantation (N = 3,578), (b) sick sinus syndrome (N = 3,310), (c) ischemic stroke (N = 5,626) and (d) cardioembolic stroke (N = 1,369) respectively. A logarithmic scale is used for the axis. OR = odds ratio.

4.5 Meta-analysis of GWAS on AF including over one million participants (paper III)

4.5.1 Genome-wide significant associations

A total of 111 loci were identified ($P < 5 \times 10^{-8}$) in the meta-analysis of GWAS on AF testing 34,740,186 variants for association with AF among 60,620 AF cases and 970,216 controls (Figure 11, Appendix 9). Participants came from six contributing studies, including 13,471 cases and 358,161 controls from Iceland (see methods). Out of the 111 identified loci 80 were novel. A stepwise conditional analysis revealed additional 31 variants that

reached genome-wide significance and were not correlated (r^2 < 0.01) with any of the 111 index variants (Appendix 10), adding up to 142 independent AF variants. The 142 variants explained 4.6% of the variation in AF. Furthermore, the total genome-wide genetic variation captured in the study explained 11.2% (s.e.m. 1.4%) of the variation in AF (h^2_{SNP}heritability).

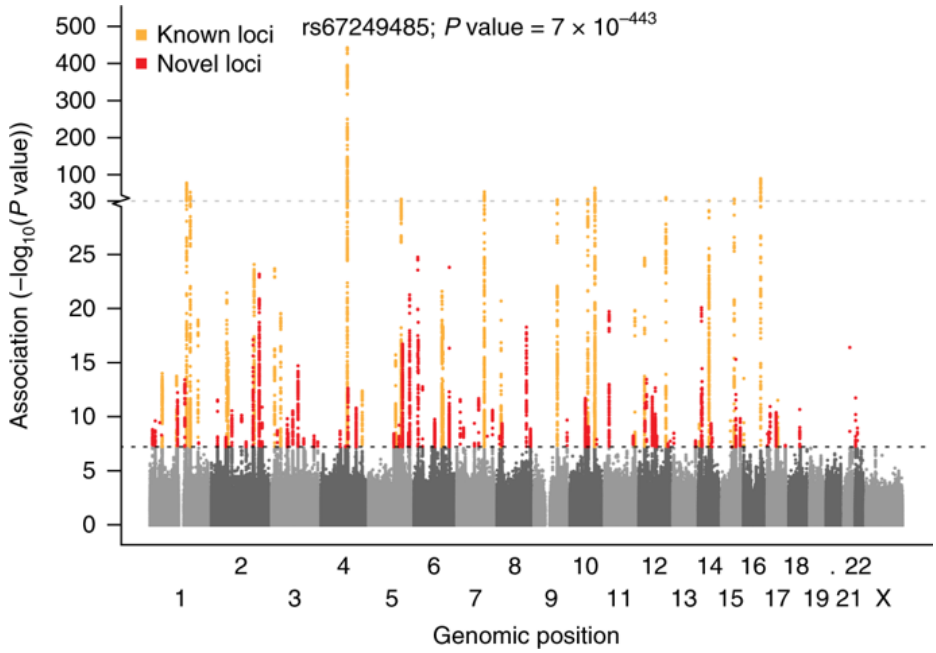


Figure 11. Manhattan plot of results of a meta-analysis of GWASs on AF among 60,620 AF cases and 970,216 controls (Paper III). Known loci are orange and novel red. A total of 34,740,186 genetic variants (each represented by a dot) were tested. The $-\log_{10}(P \text{ value})$ is plotted against the position of each SNP on each chromosome. The black horizontal dotted line represents a Bonferroni-corrected threshold of statistical significance corresponding to 1,000,000 independent tests ($P < 5 \times 10^{-8}$).

4.5.2 External validity of the results

A total of 35 variants had been reported in GWAS on AF at the time the study was published (Benjamin et al., 2009; Christophersen et al., 2017; Ellinor et al., 2012; Ellinor et al., 2010; Gudbjartsson et al., 2009; Gudbjartsson et al., 2017; Low et al., 2017; Lubitz et al., 2014; Nielsen et al., 2018a; Sinner et al., 2014; Thorolfsdottir et al., 2017; Thorolfsdottir et al., 2018; Tsai et al., 2016). Of these, 31 reached genome-wide significance in the meta-analysis after excluding the AFGen Consortium, which had published a majority of the reported variants. The other four variants were either discovered in East

Asian populations (*KCNIP1*, *NEBL*, *CUX2*) (Low et al., 2017; Tsai et al., 2016) or not included in the meta-analysis (*PLEC*) (Thorolfsdottir et al., 2017). The external validity of the results was further supported by heterogeneity analysis. Only two of the 111 variants showed significant heterogeneity of effect size across the six studies when correcting for multiple testing ($P < 0.05/111 = 4.5 \times 10^{-4}$). Both of them were index variants at established AF loci reported in multiple studies, close to *PRRX1* and *PITX2* (Appendix 9).

4.5.3 Understanding the biology of the 111 AF loci

Most of the identified AF variants were in the non-coding genome. The only two coding variants among the 111 index variants were the p.Gln254Pro in *MYZAP* and p.Ala75Val in *RPL3L*, both of which were reported in paper II. Therefore, number of approaches were applied to understand the biology behind the effects of AF risk variants, including exploring their associations with ECG measurements.

The results of pathway and functional enrichment analysis used to explore the 142 AF variants (111 index variants and 31 variants identified through conditional analysis) are described in detail in paper III but will be covered here in brief, as explained in “Declaration of contribution”. These analyses included testing for enrichment of the AF-variants with regulatory elements using the software GREGOR (Schmidt et al., 2015), revealing enrichment in regulatory features in adult and fetal heart. Furthermore, significant enrichment of genes at AF associated loci was assessed in gene sets identified by the DEPICT tool (Pers et al., 2015). The identified gene sets pointed to biological processes related to structural remodeling of the myocardium, cardiac morphology and development (Figure 12).

A systematic approach was applied to identify functional candidate genes possibly underlying the AF associations (Methods). This resulted in the prioritization of 151 candidate genes, including at least 18 genes likely to be involved in cardiac and skeletal muscle function and integrity (*AKAP6, CFL2, MYH6, MYH7, MYO18B, MYO1C, MYOCD, MYOT, MYOZ1, MYPN, PKP2, RBM20, SGCA, SSPN, SYNPO2L, TTN, TTN-AS, WIPF1*), at least 13 genes likely to be involved in mediation of developmental events (*ARNT2, EPHA3, FGF5, GATA4, GTF2I, HAND2, LRRC10, NAV2, NKX2-5, PITX2, SLIT3, SOX15, TBX5*) along with genes likely to be involved in intracellular calcium handling in the heart (*CALU, CAMK2D, CASQ2, PLN*), angiogenesis (*TNFSF12, TNFSF12-TNFSF13*), hormone signaling (*CGA, ESR2, IGF1R, NR3C1, THRB*), and function of cardiac ion channels (*HCN4, KCND3,*

KCNH2, KCNJ5, KCNN2, KCNN3, SCN10A, SCN5A, SLC9B1) (Appendix 9, column 10).

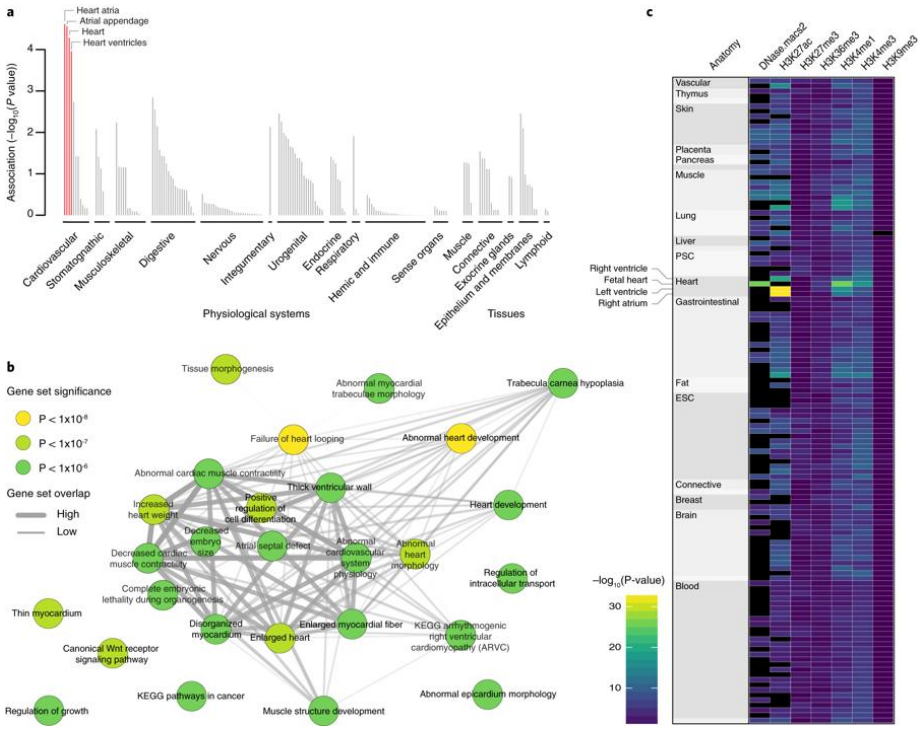


Figure 12. Tissues, reconstituted gene sets, and regulatory elements implicated in AF (paper III). (a) Based on expression patterns across 37,427 human mRNA microarrays, DEPICT predicted genes within AF associated loci to be highly expressed across various cardiac tissues. Tissues are grouped by type and significance. Red columns represent statistically significant tissues following Bonferroni correction ($P < 0.05/209 = 0.0002$). (b) Gene sets that are enriched for genes at AF loci according to DEPICT ($P < 1 \times 10^{-6}$). A total of 826 gene sets were identified in the analysis (FDR < 0.05) but only the most significantly enriched are shown here. Each node, colored according to the *P* value for enrichment, represents a gene set annotated by biological function. The gray connecting lines represent pairwise overlap of genes within the gene sets. (c) Heatmap indicating the overlap between AF associated risk variants and regulatory elements across 127 Roadmap Epigenomics tissues (each represented by a row) using GREGOR. Black indicates no data. PSC, pluripotent stem cell; ESC, embryonic stem cell.

In order to assess the effects of the 111 AF index variants on cardiac conduction in the absence of AF, they were tested for association with 123 ECG parameters from sinus rhythm ECGs of 62,974 Icelanders, after

exclusion of AF cases (Figure 13). Consistent with the results of the ECG analysis in paper I and II, the AF variants had diverse effects on ECG measurements ranging from none to extensive and varying in directions.

Sixty variants were associated with at least one ECG parameter after controlling for an FDR of 0.05 at the variant level. Out of the sixty, 39 were novel AF variants, including many with substantial ECG effects, such as the variants near *NACA, THRB, CAMK2D* and *CDKN1A*.

Although there is inconsistency in directions of effects of the AF risk variants on ECG measurements, there are a few detectable clusters of variants with similar pattern of effects. For example, a group of them have mainly lowering effects on amplitudes in both the atria (P amplitude) and ventricles (R amplitude) and shorten the duration of ventricular conduction (QRS duration). The most likely causative genes (prioritized candidate genes) linked with the variants in this group have various functions. However, two of them are intronic to genes encoding hormone receptors, *THRB* encoding the thyroid hormone receptor beta and *IGF1R* encoding insulin like growth factor 1 receptor. Both hormones have extensive effects on cardiac physiology (Ren et al., 1999; Sakurai et al., 1990). Another example of a cluster are variants on the far right of Figure 13 which shorten conduction reflected by shorter PR segment, PR interval and/or QRS duration. They have variable effects on amplitudes, although mostly decreasing them.

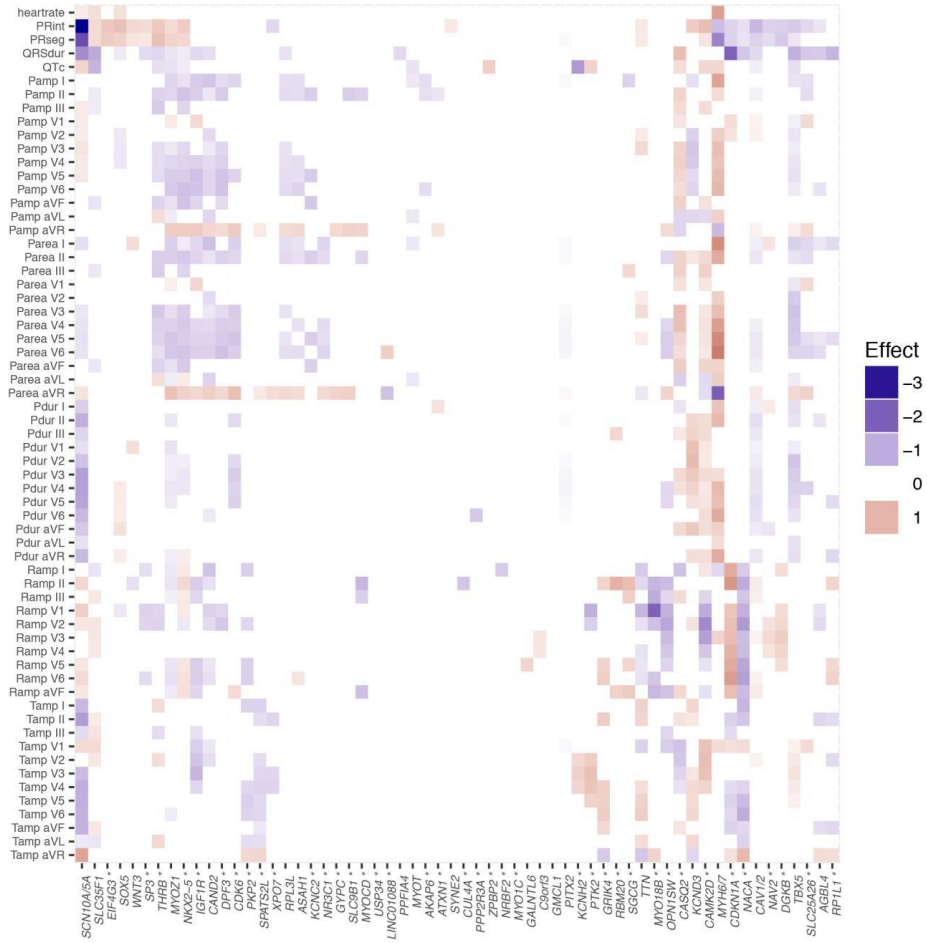


Figure 13. Heatmap showing the effects of AF variants on ECG traits in sinus rhythm ECGs, excluding AF cases (paper III). ECG measurements were available for 62,974 Icelandic individuals without diagnosis of AF. Each column shows the estimated effect of the AF risk allele on various ECG traits. The effect of each variant, annotated with the corresponding gene name, is scaled with the $\log_{10}$-AF odds ratio. Novel variants are marked with an asterisk. Red represents a positive effect of the AF risk allele on the ECG variable, and blue represents negative effect. The effect is shown only for significant associations after adjusting for multiple testing with a false discovery rate procedure for each variant. Non-significant associations are white. Sixty of 111 variants with at least one association are shown. *P* values and effect estimates were obtained using BOLT-LMM. For readability, selected highly correlated lead-specific time duration ECG variables (P interval, $r_2 > 0.51$; PR segment, $r_2 > 0.46$; QRS duration, $r_2 > 0.47$; and T duration, $r_2 > 0.16$) have been omitted from the plot. PRint = PR interval; PRseg = PR segment; QRSdur = QRS interval duration; Pamp = P wave amplitude; Parea = P wave area; Pdur = P wave duration; Ramp = R wave amplitude; Tamp = T wave amplitude.

4.5.4 Testing AF genetic risk score for association with secondary phenotypes and age of onset of AF

A genetic risk score for AF was constructed and tested for association with 1,494 International Classification of Diseases (ICD) code-defined disease groups in UK Biobank participants (Sudlow et al., 2015) (Figure 14). The genetic score was based on the 111 locus index variants and the 31 additional risk variants identified by stepwise conditional analyses (n = 142 variants). The variants were weighted by effect estimates obtained from all cohorts except the UK Biobank. The strongest association with phenotypes in the UK Biobank database was with AF ($P = 2 \times 10^{-92}$). Other associations passing Bonferroni correction ($P < 0.05/1{,}494 = 3.3 \times 10^{-5}$) were mainly with cardiovascular conditions, including heart valve disorders, heart failure, ischemic heart disease, stroke and palpitations. None of the associations remained when participants diagnosed with any type of cardiac arrhythmia (n = 24,681) were excluded from the analyses to avoid assessment bias (Figure 14). Another genetic risk score was generated using the same 142 markers but accounting for weights from all contributing studies. The score was tested for association with AF age of onset in the HUNT study, revealing that younger AF age of onset was associated with a higher genetic burden of AF.

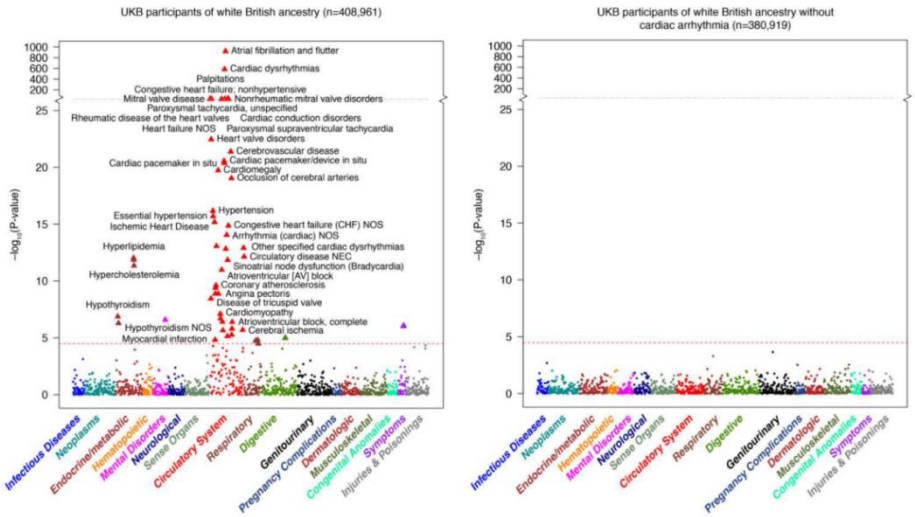


Figure 14. Association between AF genetic risk score (n = 142 variants) and 1,494 ICD-based traits among UK Biobank participants. Association tests were performed using a logistic regression adjusted for sex and birth year. The horizontal dotted red line represents a *P* value threshold of significance based on Bonferroni correction ($P < 0.05/1{,}494 = 3.3 \times 10^{-5}$). Some labels have been omitted on the left plot.

4.6 A GWAS on coarctation of the aorta (paper IV)

4.6.1 A rare missense variant in *MYH6* associates with risk of CoA

One novel genome-wide significant association, on chromosome 14q11, was identified in a GWAS testing 32.5 million sequence variants for association with CoA among 120 cases with non-syndromic CoA and 355,116 controls (Figure 15). The association was driven by a rare (0.34%) missense variant, p.Arg721Trp (c.2161C>T), in the gene *MYH6*, a previously reported arrhythmia variant. Originally, it was associated with a high risk of SSS in a study from deCODE genetics (Holm et al., 2011) and subsequently it associated genome-wide significantly with AF, as reported in paper I in this thesis (Thorolfsdottir et al., 2017). In the present study, the variant associated with CoA with an OR of 44.2 (95% CI = 20.5 – 95.5) and $P = 5.0 \times 10^{-22}$ (Figure 16). Genome-wide significance threshold for missense variants was set at 6.5×10^{-8} (Sveinbjornsson et al., 2016). It was not possible to discriminate between dominant and multiplicative models of inheritance, since none of the genotyped individuals (N = 151,677) were homozygous for the variant, consistent with what was expected under Hardy-Weinberg equilibrium (1.8 homozygotes expected). *MYH6* encodes the alpha-myosin heavy chain subunit (αMHC), a main component of the sarcomere, the basic contractile unit of cardiac muscle (Fuster et al., 2017; Sweeney & Houdusse, 2010).

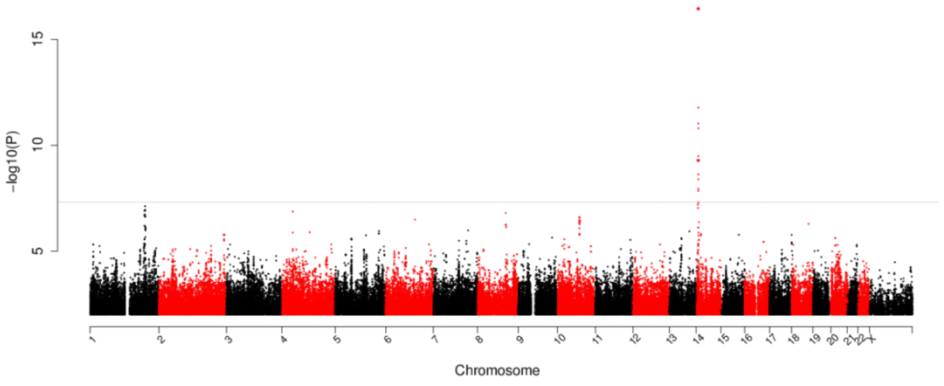


Figure 15. Manhattan plot of a GWAS on CoA in Iceland among 120 cases and 355,116 controls (paper IV). The *P* values ($-\log_{10}$) for variants association with CoA are plotted against their respective positions on each chromosome.

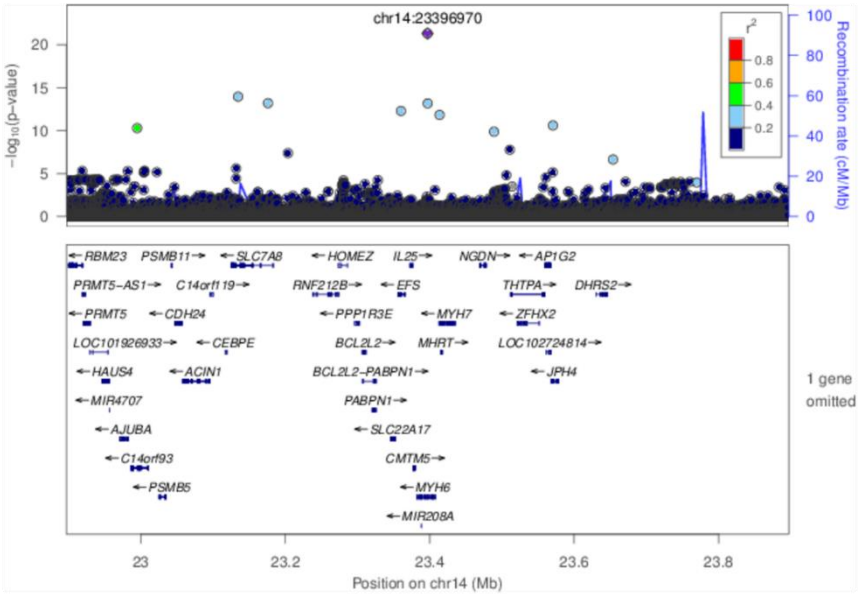


Figure 16. Region plot for the association of variants on chromosome 14q11 with CoA. A 1 Mb region on chromosome 14 is shown. The strongest association is with the missense variant p.Arg721Trp in *MYH6* located at position 23,396,970 on chromosome 14 (chr14:23396970, hg38). The nine other variants shown are weakly correlated with p.Arg721Trp, r^2 between 0.6 - 0.4 (green) and 0.4 - 0.2 (light blue).

The p.Arg721Trp variant is located in exon 18/39 in *MYH6* and leads to an arginine to tryptophan substitution at the 721st amino acid (out of 1939). This part of the gene encodes the converter domain of αMHC, a small domain which is crucial for conveying a conformational change from the active site to the lever arm upon adenosine triphosphate (ATP) hydrolysis. The variant is predicted to alter protein function by SIFT = 0 (range 0 – 1, cutoff for deleterious classification <= 0.05) (Sim et al., 2012) and PolyPhen-2 = 0.90 (range 0 – 1, where 1 is most deleterious) (Adzhubei et al., 2013), probably by altering the folding of the converter domain (Figures 17 and 18). Both SIFT and PolyPhen-2 prediction methods are based on sequence homology/conservation. The SIFT algorhithm also considers the physical properties of amino acids (Sim et al., 2012) while PolyPhen takes into account annotations of functional domains and, where available, 3D structures (Adzhubei et al., 2013).

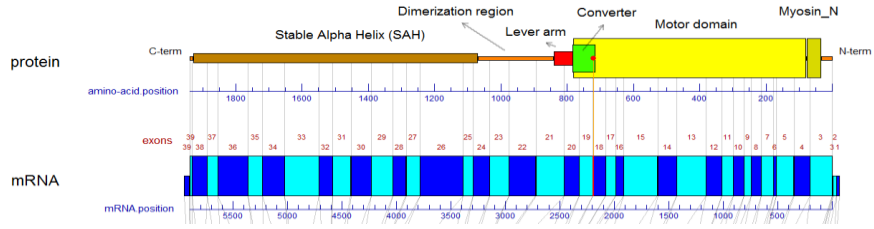


Figure 17. Transcript and protein structures of *MYH6.* Bottom panel shows transcript structure with base-pairs positions at the bottom and exon numbering at the top. Colored bars of full height depict coding exons and those with half-height depict 5' and 3' UTR's exons. Top panel shows domain composition of the protein with amino-acid positions at the bottom and domain descriptions at the top. Grey thin lines connect the panels according to exon boundaries. The position of c.2161C>T in exon 18 is highlighted by a red line in the transcript and p.Arg721Trp by a red dot in the protein. The figure is adapted from the supplementary material of Holm et al., 2011, with written permission from authors.

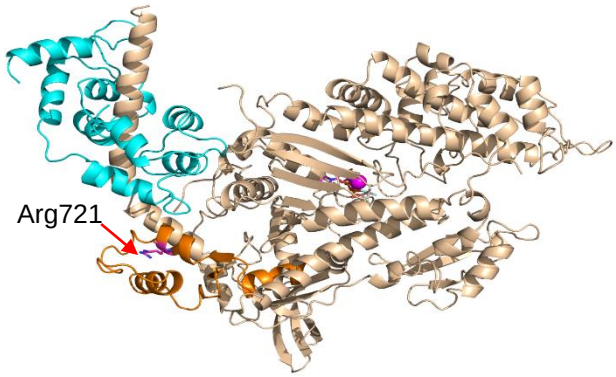


Figure 18. A 3D structural model of chicken smooth muscle myosin that is homologous to alpha-myosin heavy chain. Myosin heavy chain is shown in gold and the essential light chain shown in cyan, the active site magnesium is shown as a magenta sphere. Arginine 721 that is mutated to a tryptophan is displayed as a pink stick model. The converter domain is shown in orange. This figure was prepared in pymol (L DeLano, 2002) and is based on PDB ID 1BR1.

4.6.2 Penetrance and clinical presentation of p.Arg721Trp

Among the 151,677 chip typed Icelanders, 987 were carriers of p.Arg721Trp in *MYH6*. The variant appears to be absent or very rare outside of Iceland. It was not found in the Exome Variant Server, containing data from 6503 individuals [Exome Variant Server, NHLBI Exome Sequencing Project (ESP),

Seattle, WA, USA] (http://evs.gs.washington.edu/EVS/) (August 2016) and only one copy was found in gnomAD, holding data from 126,216 exome sequences and 15,136 WGS unrelated individuals (Karczewski et al., 2019). The higher allele frequency observed in Iceland (0.34%) is likely explained by the founder effect. A total of eight of the 987 Icelandic carriers (one per 123) were diagnosed with CoA, reflecting a low penetrance. In line with the low penetrance, carriers of the variant with CoA did not cluster in families. However, while the penetrance for CoA was low, the variant explained a substantial fraction of CoA cases in Iceland, since 20% of the 39 chip typed individuals with CoA carried p.Arg721Trp.

Through detailed review of medical records, phenotypic characteristics of CoA cases were evaluated and carriers of p.Arg721Trp compared with non-carriers (Tables 12 and 13). In total, about half of the individuals with CoA were diagnosed during the first month of life and three quarters during the first year. As expected (Engelfriet & Mulder, 2009), CoA was more common in males than in females (1.6:1). About 74% of individuals with CoA had concomitant CHDs, most commonly bicuspid aortic valve and VSD. Consistent with previous reports (Teo et al., 2011), the aortic valve was bicuspid in approximately half of individuals with CoA (Table 12). Comparison of carriers (n = 24) and non-carriers with CoA (n = 96) revealed that carriers were nominally more likely to present with mild rather than more critical and complex forms of CoA (OR = 4.2 and P = 0.023) (Table 13). Furthermore, carriers were more likely to be diagnosed with CoA as adolescents compared to non-carriers. However, only three cases in total where diagnosed as adolescents and no significant difference was observed in other age groups.

Table 12. Clinical characteristics of the CoA sample set (paper IV).

Phenotype	N[a]	Frequency[b] (%)
Gender		
Male	75	62.5
Female	45	37.5
Age at diagnosis[c]		
Neonate	56	46.7
Infant	29	24.2
Child	17	14.2
Adolescent	3	2.5
Adult	15	12.5
CoA type[d]		
Mild	13	10.8
Moderate	45	37.5
Critical +/- ventricular septal defect	41	34.2
Complex congenital heart defect	17	14.2
Interrupted aortic arch type A	2	1.7
Abdominal CoA	2	1.7
Hypoplasia		
Aortic arch	73	60.8
Ascending aorta	3	2.5
Left ventricle	3	2.5
Associated congenital heart defects		
Yes	89	74.2
Bicuspid aortic valve	57	47.5
Ventricular septal defect	38	31.7
Valvar aortic stenosis	11	9.2
Subvalvar aortic stenosis	10	8.3
Atrial septal defect	10	8.3
Mitral valve defect	10	8.3
Transposition of the great arteries	8	6.7
Tricuspid valve defect	4	3.3
Pulmonary stenosis	3	2.5
Atrioventricular septal defect	1	0.8
Single ventricle	1	0.8
No	31	25.8

[a]The number of total CoA cases with the particular phenotype. [b]The frequency of the particular phenotype among CoA cases. [c]Neonate, birth to 27 days of age; Infant, 28 days to 12 months of age; Child, 13 months to 11 years of age; Adolescent, 12 to 18 years of age; Adult, over 18 years of age. [d]Based on severity and/or anatomy of the narrowing we classified the CoA cases into five different types (Methods).

Table 13. Phenotypic differences between carriers and non-carriers of the *MYH6* p.Arg721Trp variant among CoA cases.

Phenotype	N[a]	N carriers (frequency %)[b]	N non-carriers (frequency %)[b]	OR[c]	*P* value[d]
Gender					
Male	75	14 (58.3)	61 (63.5)	0.8	0.64
Female	45	10 (41.7)	35 (36.5)	1.2	0.64
Age at diagnosis[e]					
Neonate	56	10 (41.7)	46 (47.9)	0.8	0.65
Infant	29	5 (20.8)	24 (25.0)	0.8	0.79
Child	17	4 (16.7)	13 (13.5)	1.3	0.75
Adolescent	3	3 (12.5)	0 (<4.1)	Inf	0.0072
Adult	15	2 (8.3)	13 (13.5)	0.6	0.73
CoA type[f]					
Mild	13	6 (25.0)	7 (7.3)	4.2	**0.023**
Moderate	45	9 (37.5)	36 (37.5)	1.0	1.00
Critical +/- ventricular septal defect	41	5 (20.8)	36 (37.5)	0.4	0.15
Complex congenital heart defect	17	3 (12.5)	14 (14.6)	0.8	1.00
Interrupted aortic arch	2	1 (4.2)	1 (1.0)	4.1	0.36
Abdominal CoA	2	0	2 (2.1)	0.0	1.00
Hypoplasia					
Aortic arch or isthmus	73	18 (75.0)	55 (57.3)	2.2	0.16
Ascending aorta	3	1 (4.2)	2 (20.8)	2.0	0.49
Left ventricle	3	1 (4.2)	2(20.8)	2.0	0.49
Associated CHD[g]					
Yes	89	15 (62.5)	74(77.1)	0.5	0.20
Bicuspid aortic valve	57	12 (50.0)	45 (47.0)	1.1	0.82
Ventricular septal defect	38	7 (29.2)	31 (32.3)	0.9	1.00
Valvar aortic stenosis	11	2 (18.2)	9 (9.4)	0.9	1.00
Subvalvar aortic stenosis	10	2 (8.3)	8 (8.3)	1.0	1.00
Atrial septal defect	10	2 (8.3)	8 (8.3)	1.0	1.00
Mitral valve defect	10	3 (12.5)	7 (7.3)	1.8	0.41
Transposition of the great arteries	8	0	8 (8.3)	0.0	0.36
Tricuspid valve defect	4	2 (8.3)	2 (2.1)	4.2	0.18
Pulmonary stenosis	3	0	3 (3.1)	0.0	1.00
Atrioventricular septal defect	1	0	1 (1.0)	0.0	1.00
Single ventricle	1	0	1 (1.0)	0.0	1.00
No	31	9 (29.0)	22 (22.9)	2.0	0.20
***MYH6* associated disease**					
Atrial fibrillation	6	0	6 (6.2)	0.0	0.60
Cardiomyopathy	3	1 (4.2)	2 (2.1)	2.0	0.49
Sick sinus syndrome	2	1 (4.2)	1 (1.0)	4.1	0.36

[a]The number of total CoA cases with the particular phenotype. [b]The frequency of the particular phenotype among carriers (N = 24) and non-carriers (N = 96) of p.Arg721Trp. [c]OR = odds ratio. [d]Fisher's exact test was used to evaluate the significance of the frequency difference between carriers and non-carriers of p.Arg721Trp. [e]Neonate, birth to 27 days of age; Infant, 28 days to 12 months of age; Child, 13 months to 11 years of age; Adolescent, 12 to 18 years of age; Adult, over 18 years of age. [f]Based on severity and/or anatomy of the narrowing we classified the CoA cases into five different types (Methods). [g]CHD = Congenital heart defect.

4.6.3 Association of p.Arg721Trp in *MYH6* with other cardiovascular diseases

As reported in paper I, the p.Arg721Trp variant in *MYH6* that increases risk of CoA also increases risk of arrhythmia and has extensive and strong effects on cardiac conduction in sinus rhythm (chapter 4.3, Figure 9). To further explore the effects of the variant, we tested it for association with additional cardiac phenotypes available in the deCODE phenotype database. These included other congenital heart diseases, common heart diseases, and several echocardiogram variables (Table 14); significance threshold was set at $P < 0.003$ (0.05/17 individual phenotypes tested). The previously reported associations with SSS and AF (Holm et al., 2011; Thorolfsdottir et al., 2017) became stronger with increased sample size (Table 14). However, a previously reported suggestive association with thoracic aortic aneurysm (Holm et al., 2011) was not detected. The p.Arg721Trp variant did associate with increased risk of several congenital heart diseases: bicuspid aortic valve, VSD, ASD, and patent ductus arteriosus (Table 14). As expected, the strongest association was with bicuspid aortic valve (OR = 10.5 and $P = 7.3 \times 10^{-8}$). In addition, the variant associated with late onset aortic valve stenosis. To assess if p.Arg721Trp associated with congenital heart defects in the absence of diagnosed CoA, the associations were tested again after removing individuals with CoA from the analysis. Although the effect of p.Arg721Trp was consistently weaker, the associations remained. The existence of undiagnosed CoA in these individuals cannot be excluded. The variant also associated with heart failure and ischemic stroke and with left atrial diameter but not with other variables derived from the echocardiographic data such as aortic root diameter or left ventricular end diastolic diameter (Table 14). The p.Arg721Trp variant did not associate with hypertension or coronary artery disease.

Table 14. Association of p.Arg721Trp in *MYH6* with congenital heart defects and various cardiac phenotypes.

	N affected	N controls	OR/effect (CI)[a]	*P* value
Congenital heart defects				
Coarctation of the aorta	120	355,116	44.2 (20.5 - 95.5)	5.0×10^{-22}
Bicuspid aortic valve	208	293,346	10.5 (2.6 - 38.0)	7.3×10^{-8}
Bicuspid aortic valve without CoA	178	293,293	6.0 (2.0 – 17.7)	0.0012
Ventricular septal defect	715	357,641	4.4 (1.9 - 10.0)	3.7×10^{-4}
Ventricular septal defect without CoA	688	357,511	3.2 (1.2 – 8.5)	0.016
Patent ductus arteriosus	594	357,762	4.9 (2.1 - 11.6)	2.3×10^{-4}
Patent ductus arteriosus without CoA	576	357,563	4.1 (1.6 – 10.3)	0.0029
Atrial septal defect	657	353,096	3.3 (1.5 - 7.1)	0.0026
Atrial septal defect without CoA	650	352,947	3.2 (1.5– 6.9)	0.0036
Cardiac conditions				
Sick sinus syndrome	3,310	346,082	8.7 (6.8 - 11.2)	6.2×10^{-64}
Atrial fibrillation	13,471	374,939	2.4 (1.9 - 3.0)	1.1×10^{-14}
Aortic valve stenosis	2,457	349,342	2.7 (1.8 - 4.0)	1.8×10^{-6}
Heart failure	10,48	353,508	1.8 (1.4 - 2.3)	2.3×10^{-6}
Ischemic stroke	8,948	369,624	1.5 (1.1-2.0)	0.0029
High degree atrioventricular block	1,303	361,919	2.1 (1.2 - 3.5)	0.0092
Coronary artery disease	37,782	318,845	1.2 (1.0 -1.5)	0.056
Hypertrophic cardiomyopathy	163	239,293	0.0 (0.0 - 4.5)	0.15
Thoracic aortic aneurysm	353	302,458	1.8 (0.6-5.3)	0.31
Hypertension	54,974	324,803	1.1 (0.9 - 1.3)	0.44
Echocardiogram				
Left atrial diameter	19,380		0.3 (0.1 – 0.5)	2.6×10^{-4}
Aortic root diameter	19,506		-0.1 (-0.2 - 0.1)	0.41
LVEDD[b]	5,701		0.0 (-0.3 - 0.3)	0.93

[a]Estimated odds ratio (OR) or the effect in standard deviation and the 95% confidence interval (CI) for the association with p.Arg721Trp. [b]Left ventricular end-diastolic diameter.

5 Discussion

The aim of the genetic studies described in this thesis was to further the understanding of the pathophysiology of AF and CoA. Three GWASs on AF were performed resulting in a substantial increase in the number of sequence variants associated with AF. This included novel associations with coding variants in three genes, *PLEC*, *RPL3L* and *MYZAP*, that all have a role in maintaining cardiac structure. In an attempt to gain insights into the biology underlying novel and reported AF variants their association with a variety of ECG measurements were explored. The analysis revealed that effects of AF variants on cardiac conduction indices vary substantially, suggesting diverse mechanisms behind their effects on AF risk.

Through a large meta-analysis involving over 1,000,000 participants in total from six contributing cohorts, 80 novel AF loci were identified. Most of the loci were represented by common non-coding variants. Therefore, pathway and functional enrichment analysis was undertaken to identify causative genes and pathways. This analysis further highlighted the importance of structural integrity of the heart in the development of arrhythmias as well as pathways involving fetal cardiac development, transcription factors, ion channels and calcium signaling.

The thesis also describes a GWAS on the congenital heart defect CoA. The study revealed that one of the strongest AF variants, a missense variant in the sarcomere gene *MYH6,* explains approximately 20% of CoA cases in Iceland. The *MYH6* variant was first discovered through its strong association with SSS and is in fact the only variant that has been associated with SSS risk through GWAS (Holm et al., 2011). It also increases risk of additional CVDs, including cardiac septal defects and aortic stenosis. Furthermore, in paper I we showed that it has extensive and strong effects on various ECG measurements in the absence of AF. Combined, these results reveal a pleiotropic effect of the *MYH6* missense variant on cardiac function and disease, both in the fetal and adult heart.

5.1 Novel AF associations with coding variants in structural genes

5.1.1 *PLEC*

The first of three GWASs on AF described in this thesis led to the discovery of a low frequency missense variant, p.Gly4098Ser in *PLEC,* increasing risk of AF. The *PLEC* gene encodes plectin, a large (>500 kDa) and versatile cytoskeletal linking protein which is widely distributed throughout the cytoskeleton (Wiche, 1998). Plectin directly binds to subcomponents of all major cytoskeletal filament networks; intermediate filaments, actin-based microfilaments and microtubules (Castanon et al., 2013; Eger et al., 1997; Wiche, 1998). It also anchors intermediate filaments to various cellular components, including focal adhesions, Z-disks, costameres, intercalated discs as well as the mitochondrial, nuclear and subplasma membrane skeleton (Castanon et al., 2013; Seifert et al., 1992; Wiche et al., 1983). Plectin is therefore assumed to play an important role for the many functions of the cytoskeleton, which include maintaining structural integrity of cells and tissues and controlling dynamic processes such as cell division, adhesion and intracellular trafficking (Wiche, 1998).

The importance of plectin in maintaining structural integrity has been shown through knockout studies in mice and the identification of rare human disease resulting from truncating mutations in *PLEC* (Winter & Wiche, 2013). Plectin is widely expressed in human tissue ("Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans," 2015) but evidence suggests that it has a specifically important role in skin, muscle and heart. Studies of plectin (−/−) mice revealed severe skin blistering and disrupted cytoarchitecture in skeletal muscle and heart, including disintegration of intercalated discs and sarcomere disarrangement (Andra et al., 1997). Furthermore, homozygous protein truncating mutations in *PLEC* in humans are known to cause a syndrome of skin fragility, called epidermolysis bullosa simplex (EBS). EBS is commonly accompanied by muscular dystrophy (EBS-MD) (Banwell et al., 1999; Pulkkinen et al., 1996), myasthenic syndrome (EBS-MDMyS) (Banwell et al., 1999; Selcen et al., 2011), limb-girdle muscular dystrophy type 2Q (LGMD2Q) (Gundesli et al., 2010) or pyloric atresia (EBS-PA) (Nakamura et al., 2005). Mutations in *PLEC* also cause EBS-Ogna, an autosomal dominant disease without muscular dystrophy (Koss-Harnes et al., 2002). Cardiomyopathy and arrhythmias have been described in several patients with EBS due to *PLEC* truncating mutations (Bolling et al., 2010; Celik et al.,

2005; Schroder et al., 2002; Villa et al., 2015). An EBS-like skin phenotype was not observed from reviewing deCODEs data on homozygous carriers of p.Gly4098Ser, possibly reflecting the fact that the variant is not protein truncating. In light of the cardiac phenotypes observed in plectin deficient mice and sometimes as a part of EBS in men, it is possible that p.Gly4098Ser in *PLEC* increases risk of AF by affecting the role of the protein in maintaining structural integrity of the heart.

The p.Gly4098Ser variant is predicted to be damaging and is located in a conserved region of the gene. It results in a glycine to serine substitution on the protein surface, close to a suggested interaction site for intermediate filaments (Nikolic et al., 1996). The substitution of a hydrophobic glycine by a polar serine might affect plectins interaction with other molecules. Furthermore, glycine is unique among amino acids for its conformational flexibility, attributed to its side chain of a single hydrogen atom. Its substitution by serine, a more rigid amino acid, could therefore possibly affect the protein structure and interrupt the nearby intermediate filament binding site, either by directly disrupting its structure or hindering access to it.

In the present study, p.Gly4098Ser did not associate with overt cardiomyopathy but association with subclinical or mild ventricular cardiomyopathy cannot be excluded. It is possible that p.Gly4098Ser causes subclinical or even overt atrial myopathy since this phenotype is not routinely assessed or diagnosed. Indeed, analysis of ECG measurements showed that the variant results in a widespread effect on cardiac electrical function in both atria and ventricles in the absence of AF. Furthermore, the association with lower creatine kinase levels and lower R amplitude may reflect less skeletal and cardiac muscle mass of carriers.

5.1.2 *MYZAP*

In the second paper described in this thesis, a missense variant in *MYZAP* was associated with increased risk of AF. The *MYZAP* gene encodes myocardial zonula adherens protein (myozap). It is primarily expressed in the heart in man and has been located to intercalated discs and sarcomeric Z-discs in the mouse. In fact, the *MYZAP* gene was recently discovered by Seeger et al. through a search for new components of the intercalated discs (Seeger et al., 2010), highly specialized cell-cell contact structures that are unique to the heart. Their role is to provide mechanical and electrical connection between contracting cardiomyocytes, enabling the heart to contract as a functional syncytium (Noorman et al., 2009). The location of myozap at intercalated discs implies an important role in cardiac electrical

and mechanical integrity. The fact that myozap colocalizes with actin at intercalated discs and the sarcomeric Z-discs suggests that myozap plays a role in transmission of contractile forces, maintaining mechanical integrity and possibly the versatile role of actin signaling (Seeger et al., 2010). Furthermore, in vitro functional experiments have revealed that myozap takes part in cardiac signaling transduction, as it promotes SRF-signaling to the nucleus in a Rho-dependent fashion (Seeger et al., 2010). This signaling pathway has been implicated in induction of cardiac remodeling and cardiomyopathy (Miano, 2010; Zhang et al., 2001).

Intercalated discs are composed of three functional and structural subunits; desmosomes and fascia adherens, both contributing to mechanical coupling, and gap junctions that provide electrical coupling by allowing free movement of ions between adjacent cardiomyocytes (Forbes & Sperelakis, 1985). Given the importance of intercalated discs for the function and mechanical integrity of the heart, it is not surprising that mutations in intercalated disc genes are known to cause cardiomyopathy. In particular, these mutations have been associated with arrhythmogenic right ventricular cardiomyopathy (ACM), which involves considerable risk of AF and ventricular arrhythmias and is a leading cause of sudden cardiac death among young people (Rampazzo et al., 2014). The majority of the mutations causing ACM are in genes that encode components of desmosomes, most commonly in *PKP2*, *DSP*, *DSG2* and *DSC2* (Noorman et al., 2009). Changes in the function of desmosomal proteins have been shown to interrupt the expression and distribution of gap junctions (Noorman et al., 2009). This interplay between mechanical and electrical coupling at the intercalated discs might at least partly explain the intriguing fact that in ACM arrhythmias and conduction abnormalities are frequently encountered before the appearance of structural defects (Moncayo-Arlandi & Brugada, 2017).

The *MYZAP* gene itself has not been associated with cardiomyopathy in man. However, animal studies have pointed to its importance for cardiac structure since both knockdown of the myozap ortholog in zebrafish and cardiac overexpression and knockdown of myozap in the mouse resulted in cardiomyopathy (Frank et al., 2014; Rangrez et al., 2016; Seeger et al., 2010). Furthermore, myozap$^{-/-}$ mice are born with morphologically normal hearts but show a maladaptive response to pressure overload leading to cardiac remodeling, cardiomyopathy and premature death (Rangrez et al., 2016), suggesting an important role of myozap in cardiac stress-signaling.

The desmosomes and fascia adherens functional units of intercalated discs have been shown to overlap at a mixed-type junctional structure termed the area composita (Borrmann et al., 2006). The myozap protein binds to and colocalizes with components of the area composita (Seeger et al., 2010), which is attached to both the actin cytoskeleton and the intermediate filament network of cardiomyoctes (Goossens et al., 2007). The p.Gln254Pro missense variant associating with AF is predicted to have deleterious effects on protein function (Choi & Chan, 2015). It is located next to the myozap protein region involved in both actin colocalization and activation of serum response factor-dependent transcription (amino acids 91-250). It could therefore potentially affect either one or both protein functions (Seeger et al., 2010). The induction of a conformationally constrained amino acid, like proline, can lead to perturbation in local folding hence possibly interrupting function of adjacent domains. In the current study, p.Gln254Pro did not associate with cardiomyopathies, ventricular arrhythmias or sudden cardiac death in the deCODE data. However, it is possible that the variant only affects the atria or that a ventricular effect does exist but was not identified because of lack of power.

The novel association of a coding variant in *MYZAP* strongly implicates intercalated discs in AF pathology, emphasizing what has previously been suggested by associations with non-coding variants. For example, through GWASs several AF variants have been identified close to or intronic to genes encoding components of the intercalated discs, including *GJA1* (Christophersen et al., 2017), *GJA5* (Nielsen et al., 2018b; Roselli et al., 2018) and *PKP2* (Nielsen et al., 2018b; Roselli et al., 2018). Furthermore, the strongest AF association from GWASs is with a variant close to *PITX2*, a gene that has been shown to directly regulate expression of intercalated disc genes (Tao et al., 2014).

5.1.3 *RPL3L*

In paper II two coding variants in *RPL3L* were also associated with increased risk of AF through a GWAS in Iceland and the UK biobank. *RPL3L* encodes the ribosomal protein RPL3L (ribosomal protein like 3L) (Van Raay et al., 1996). The association is particularly interesting since *RPL3L* is primarily expressed in skeletal muscle and heart unlike most ribosomal proteins that are ubiquitously expressed (Van Raay et al., 1996). The ribosome is a complex structure composed of four different ribosomal RNAs and ~80 ribosomal proteins, which main function is to translate messenger RNA into protein (McCann & Baserga, 2013). Historically, the ribosome has been

considered to function in a housekeeping capacity, performing its function without regulatory activity. Recently, this view has been challenged by studies revealing that ribosome activity may be regulated in a cell-specific manner, for example through changes in the protein composition of the ribosome by paralog substitution (Sauert et al., 2015; Xue & Barna, 2012). An example of this is when RPL3L is expressed in striated muscle, replacing its ubiquitously expressed homolog, RPL3 (Chaillou et al., 2013).

The expression of RPL3L has been shown to alter ribosome function and it has been postulated that it acts as a negative regulator of muscle growth. Evidence from in vitro experiments suggest that hypertrophic stimulus down-regulates RPL3L expression (Chaillou et al., 2013) and when induced, RPL3L expression impairs myotube growth (Chaillou et al., 2016). The two *RPL3L* variants associated with AF in the current study, might therefore increase risk of AF by affecting cardiac structure. The *RPL3L* splice-donor variant, c.1167+1G>A, causes skipping of a whole exon, as was demonstrated by RNA sequencing on cardiac samples in the current study. The skipped exon is the ninth and the second to the last exon of *RPL3L*. It encodes amino acid residues 350 to 389 which are 75% identical to the corresponding RPL3 residues. In yeast it has been shown that the amino acids corresponding to residues 382-389 in human RPL3L form a part of the contact site of the ribosomes with the signal recognition particle that targets ribosomes to the endoplasmic reticulum membrane (Ben-Shem et al., 2011). Based on functional similarities between RPL3 and RPL3L it is therefore possible that c.1167+1G>A disrupts engagement of RPL3L containing ribosomes with the endoplasmic reticulum and thus reducing ribosomal function.

The *RPL3L* missense variant associating with AF is p.Ala75Val, and Ala75 is highly conserved in both RPL3L and RPL3 over a range of species. Prediction scores for consequences of the amino acid substitution are conflicting. It is likely benign according to CADD score (Kircher et al., 2014) but is predicted to be deleterious by PROVEAN (Choi & Chan, 2015). The two amino acids, alanine and valine have similar properties, both are hydrophobic and non-charged. However, it is possible that the additional methyl group in valine has unstabilazing effects on protein structure. The chemical properties of amino acid residues are taken into account by the PROVEAN algorithm but conservation is the most important factor (Choi & Chan, 2015). The 3D structure of RPL3L is not known and no functional domains have been described, which further complicates the prediction.

Functional experiments would be necessary to determine the exact effects of the variant on protein function.

Both *RPL3L* variants increase the risk of AF suggesting they affect the function of RPL3L in the same direction. Based on the suggested effect of the splice-donor variant, this effect is more likely to be loss of, rather than gain of function. Supporting this are the results of a newly published meta-analysis of GWASs on AF by Roselli et al. (Roselli et al., 2018). One of the AF associated variants in the meta-analysis is a synonymous (not modifying amino acid sequence) variant in *RPS2* located over 9 kb away from *RPL3L.* Using the GTEx database Roselli et al. observed a significant eQTL at this locus, as the risk allele for AF associated with decreased expression of RPL3L in the heart. Furthermore, transcriptome analysis revealed association of decreased predicted RPL3L expression in right atrial appendage with increased risk of AF (Roselli et al., 2018). The study by Roselli et al. therefore strengthens the establishment of *RPL3L* as an AF risk gene. It also confirms the speculated directionality of the association, i.e. reduced function and increased AF risk, as suggested by our results.

This is the first time that a ribosomal protein is linked with risk of AF. Only a few rare inherited diseases have been specifically linked to mutations in genes encoding ribosomal proteins, including Diamond–Blackfan anemia and Shwachman–Diamond syndrome. Both syndromes are characterized by a distinct set of clinical features, including developmental abnormalities and bone marrow failure (Raiser et al., 2014). The fact that a ribosomal protein specifically expressed in skeletal muscle and the heart affects risk of AF is in line with the novel concept of ribosome specialization in muscle and underscores the importance of this specialization for normal function of the heart. Furthermore, *RPL3Ls* role in regulation of muscle growth links it to the discovery of the missense variants in *PLEC* and *MYZAP* increasing AF risk, since both associations implicate cardiac structure in arrhythmogenesis.

5.1.4 Insights into pathogenesis of AF gained from coding variants

Altogether, paper I and II of this doctoral thesis report novel AF associations with coding variants in three genes, *PLEC*, *RPL3L* and *MYZAP*. Intriguingly, all three genes encode proteins that contribute to cardiac structure. None of the variants associated significantly with clinically overt cardiomyopathy, thus the associations likely represent clinical electrophysiological consequences of an otherwise subclinical mechanical dysfunction. Prior to the current studies, two coding variants had been implicated in AF through GWASs, both also in

structural genes. First, a rare missense variant in *MYH6* strongly associated with risk of SSS, and secondarily with AF (Holm et al., 2011). In the current study (paper I) this variant associated genome-wide significantly with AF. Second, a frameshift variant in *MYL4* causes fully penetrant, early onset AF in the homozygous state (Gudbjartsson et al., 2017). Both *MYH6* and *MYL4* encode parts of myosin, a major component of the sarcomere, the building block of the cardiac contractile system (Fuster et al., 2017). Both genes are primarily expressed in the atria of the adult heart and in neither case was there evidence of clinically overt cardiomyopathy preceding arrhythmia among carriers (Gudbjartsson et al., 2017; Holm et al., 2011). The discoveries of *MYH6* and *MYL4* represented the first implications of cardiac structural or contractile units as primary players in arrhythmogenesis. The importance of cardiac structure and mechanics in maintaining normal electrical conduction has now been further emphasized by the current discoveries of coding variants in *PLEC*, *RPL3L* and *MYZAP* increasing risk of AF.

Most other variants identified in GWASs on AF hitherto are represented by non-coding variants affecting the risk of AF through unknown mechanisms and unknown genes (Bapat et al., 2018). This also applies to most other common traits that have been the subjects of GWASs, including cardiovascular diseases such as hypertension (Hoffmann et al., 2017), coronary artery disease (van der Harst & Verweij, 2018) and type 2 diabetes (Xue et al., 2018). However, it has been increasingly recognized that uncommon variants with substantial effects can contribute to the genetic architecture of complex traits. In recent years, rare or low frequency coding variants have been associated with many common diseases and traits, including for example type 2 diabetes (Steinthorsdottir et al., 2014), osteoarthritis (Styrkarsdottir et al., 2018) and height (Marouli et al., 2017). Individually, these uncommon coding variants may not explain a large proportion of disease risk at a population level but they nevertheless provide important biological insight into disease mechanisms.

Detecting rare sequence variants and testing them for association with disease has been made possible by large scale whole-genome sequencing projects, such as the one conducted by deCODE genetics (Gudbjartsson et al., 2015). Furthermore, the extended sampling of the Icelandic population allows long-range phasing and identification of the long haplotypes needed to impute rare sequence variants (Kong et al., 2008). Some of these variants, including, as an example, the missense variant in *PLEC*, have reached

higher frequencies due to the small size of the Icelandic population and the founder effect (Gudbjartsson et al., 2015).

The suitability of the genotype data at deCODE genetics for detecting rare variants and the biological insights gained from these associations was the incentive for focusing on rare and low frequency coding variants in paper II of this thesis. In contrast to previously reported non-coding AF variants, the coding variants associating with AF risk in the current studies directly implicate causal genes in AF pathology, as they affect their protein products. In addition to the common structural role of all three genes, further biological insight is provided by inquiring the likely effects of each variant on protein function. For example, p.Gly4098Ser is located close to the binding site of plectin with intermediate filament, possibly interrupting this function and the splice-donor variant in *RPL3L* causes exon skipping, making it likely to result in loss of, rather than gain of, protein function. Both examples are speculative, but might guide future functional studies on the respective proteins.

5.2 Effects of AF variants on risk of SSS, pacemaker and stroke

In paper I the AF risk of AF variants was plotted against their effects on SSS, pacemaker implantation, ischemic stroke and cardioembolic stroke and in this dissertation the novel variants from paper II were added for completion. The analysis was done with the aim of investigating the causal relationship between AF and these AF related phenotypes. It is a form of mendelian randomization, a method where sequence variants are used as proxies to investigate the causative effects of a risk factor, in this case AF, on an outcome (Smith & Ebrahim, 2003). Since the assignment of genetic predisposition is random, the method is less susceptible to confounding or reverse causation that may limit interpretations of observational studies (Lawlor et al., 2008; Smith & Ebrahim, 2003).

Stroke is a well known consequence of AF (Stewart et al., 2002; Wolf et al., 1991). The thrombogenic state in AF results from several pathophysiological mechanisms and the currently applied risk scores have limited efficacy in predicting stroke risk (Goldberger et al., 2015; Graves et al., 2018). It has been hypothesized that a genotype-based risk stratification for atrial thrombus formation might in the future become a useful tool, as was suggested by the high stroke risk of homozygotes of the AF associated variant in *MYL4* despite a low CHA_2DS_2-VASc score (Gudbjartsson et al.,

2017). Our assessment of the effects of AF associated variants on stroke risk revealed that the AF variants generally affect stroke risk in proportion to their AF risk. None of the variants stood out with a stronger effect on stroke risk than expected by the AF risk. These results indicate that the different mechanisms behind development of AF mediated by each of the loci do not affect the AF associated stroke risk. Based on this assessing the cumulative genetic risk for AF could aid in predicting stroke risk. A recent study revealed that a polygenic risk score for AF explained ~20% of the heritable component of cardioembolic stroke risk (Pulit et al., 2018). Further studies are warranted to determine whether genetic risk can serve as an additional biomarker for stroke risk stratification in patients with AF.

The causal relationship between AF and SSS is more complicated and not completely understood (Chang et al., 2013; Kezerashvili et al., 2008; Sanders et al., 2004). SSS, a common indication for pacemaker implantation, frequently coexists with AF (Ferrer, 1968; Kezerashvili et al., 2008). The two arrhythmias share risk factors and somewhat controversial evidence suggests that they can predispose to each other. AF leading to SSS is supported by studies revealing that pacing induced AF impairs SA node function in dogs (Elvan et al., 1996) and in human subjects AF causes regional substrate changes around the SA node, thereby impairing its function (Chang et al., 2013). The genetic evidence presented here supports the theory that AF can promote SSS. The AF variants likely affect AF risk through diverse mechanisms, as evident by their variable effects on normal cardiac conduction. Still, our results show that most of them affect SSS risk in proportion to their AF risk. This suggest that AF might be the main contributor to the increased SSS risk associated with these variants, rather than each of the variant acting through pathways that independently lead to AF or SSS.

One variant presents an exception to the trend of SSS effects being proportional to AF effects, the p.Arg721Trp missense variant in *MYH6* which was originally discovered through its strong association with SSS (Holm et al., 2011). The unusually high risk of SSS mediated by this variant might suggest that it acts through a single pathway that is common to the pathogenesis of AF and SSS or that it acts through more than one pathway affecting the two arrhythmias separately. It is even possible that the *MYH6* variant causes AF secondary to SSS, as it has a much stronger effect on risk of SSS than AF. This would support the notion that SSS can cause AF, which has been suggested by studies reporting diffuse atrial remodeling and increased right atrial refractoriness in patients with SSS (Luck & Engel, 1979; Sanders et al., 2004). In paper IV, p.Arg721Trp in *MYH6* was associated with

increased risk of CoA and therefore its effects on cardiac disease and function will be further discussed in chapter 5.5.

5.3 Meta-analysis substantially increasing the number of AF variants

The large GWAS meta-analysis reported in paper III resulted in a substantial increase in the number of AF variants, identifying 80 novel AF loci. Only two of the 111 index variants are coding variants, the p.Gln254Pro in *MYZAP* and p.Ala75Val in *RPL3L*, both of which were reported in paper II. The other loci were represented by variants in the non-coding genome, although some had a coding variant in high LD. As most of the 111 loci affect AF risk through unknown genes and mechanisms, a variety of approaches were undertaken to further understand their biology. In contrast to the interpretation of associations with coding variants acting through a known gene the approaches applied here analyze the loci as a group, making use of their large number to identify common features that point to potentially important pathways and mechanisms in the development of AF.

5.3.1 Insights into pathogenesis of AF gained from analyzing all established AF variants

Interestingly, a recurring theme in the analysis of all genome-wide significant AF variants in the meta-analysis was maintenance of cardiac structural integrity, reinforcing the findings of papers I and II. Furthermore, the analysis revealed that AF risk variants likely act in both the adult and fetal heart. For example, the DEPICT gene-sets enriched by genes at AF loci pointed to biological processes related to cardiac morphology and development as well as with structural remodeling of the myocardium. Furthermore, the most common mutual function of the 151 prioritized candidate genes was involvement in cardiac and skeletal muscle function and integrity (at least 18 genes). Some of these genes have been strongly implicated in cardiomyopathy, including *MYH7* (Maron & Maron, 2013), *RBM20* (Brauch et al., 2009), *TTN* (Herman et al., 2012) and *PKP2* (Gerull et al., 2004). The second most common function of the candidate genes was likely involvement in mediation of developmental events (at least 13 genes). Notably, at least one of these genes, *NKX2-5* which encodes a homeobox-containing transcription factor (Lints et al., 1993), has an important role in cardiac structure. Variants in *NKX2-5* are known to cause a diverse cardiac phenotype, including CHDs, arrhythmias (Benson et al., 1999) and adult-onset cardiomyopathy (Costa et al., 2013; Sveinbjornsson et al., 2018; Yuan

et al., 2015). In general, the results of the meta-analysis reported in paper III are consistent with the knowledge gained from associations of coding variants in structural genes associated with AF, and further emphasizes the likely importance of cardiac mechanics in the development of AF.

Several additional pathways are highlighted in the analysis of AF loci, underscoring the complexity of AF pathogenesis. These included function of cardiac ion channels, involvement in intracellular calcium handling in the heart, angiogenesis and hormone signaling. The complexity of AF development and diversity of the biology of AF variants was further implied by the diverse effects of the 111 index variants on ECG measurements, as was the case in the ECG analysis in paper I and II and will be discussed in chapter 5.4. After acceptance of paper III, another large meta-analysis, by Roselli et.al. was published (Roselli et al., 2018), identifying 97 AF loci, a majority of which overlapped with the 111 loci from the current meta-analysis. In line with the results of this meta-analysis (paper III), functional enrichment and pathway analysis performed in the Roselli study also implicated genes enriched within cardiac developmental, electrophysiological, contractile and structural pathways (Roselli et al., 2018).

5.3.2 Heritability of AF

The h^2_{SNP} heritability calculated in this meta-analysis was 11.2%. In other words, 11.2% of variation in AF is explained by all the additive genome-wide genetic variation captured in this study (Yang et al., 2017). A similar proportion, 11.4%, was recently reported in a study of a subset (152,736 subjects) of the UK Biobank dataset by Ge et al (Ge et al., 2017). However, a study by Weng and colleagues on a different subset of the UK Biobank population (120,286 subjects) reported a much higher h^2_{SNP} heritability of 22% (Weng et al., 2017). As a possible explanation for this discrepancy the authors mention that they had a broader definition of AF compared to the study by Ge et al. In fact, it is likely that all three studies present an underestimation of AF heritability for reasons such as incomplete phenotypic definitions and the fact that they only report the additive heritability (Visscher et al., 2008). Furthermore, the two studies performed in the UK Biobank only include markers with MAF>1%, missing the potentially high effects of rare variants (Ge et al., 2017; Weng et al., 2017). With regards to the relatively low heritability estimation in our study, it might at least partly be explained by the argument that heritability estimates from GWAS meta-analysis could be attenuated due to heterogeneous genetic effects across populations (de Vlaming et al., 2017).

Nevertheless, these proportions (11-22%) likely represent lower bound estimates of AF SNP heritability and provide insight into the contribution of genetic variability to AF risk. Notably, the genome-wide significant variants identified in paper III explained only 4.6% of AF variance, compared to 11.2% explained by the total genome-wide genetic variation. This is consistent with Weng et.al. reporting that only a third of AF h^2_{SNP} heritability was explained by established AF loci, putative cardiac arrhythmia, and cardiomyopathy gene regions (Weng et al., 2017). This suggests that a substantial proportion of AF risk is driven by variation in regions that have not yet been established as AF loci in GWASs.

5.3.3 Effects of AF genetic risk score on other phenotypes and age at onset of AF

Finally, a genetic risk score was constructed based on the 111 index and 31 additional risk variants identified through conditional analysis. This risk score associated with AF risk in the UK Biobank, and with several mainly cardiovascular conditions, including heart failure, ischemic heart disease and stroke. However, the associations did not remain when removing all participants diagnosed with an arrhythmia leading to the conclusion that the AF genetic risk score is specific for AF and that additional associations are possibly consequences of AF, including for example heart failure and stroke. Of note is the fact that the SSS sample size in the UK biobank is small (<700) compared to the Icelandic sample size used for assessment of secondary associations in paper I and II (>3,500). The lack of association of the AF genetic risk score with SSS in the UK biobank does therefore not constitute a discrepancy with the association of AF variants with increased SSS risk previously discussed in this thesis. Finally, a higher genetic burden of AF was associated with younger age at onset of AF, supporting epidemiological studies reporting the association of increased AF risk with lower age at onset of AF in close relatives (Lubitz et al., 2010; Oyen et al., 2012).

5.4 Effects of AF variants on normal cardiac conduction

Exploring the effects of all AF variants on ECG measurements, an endophenotype for AF, was undertaken in papers I, II and III with the aim to further understanding of their physiological effects in the absence of AF, as AF cases were excluded from the analysis and only sinus rhythm ECGs were used. The analysis revealed that their effects on cardiac electrical function are diverse. They range from none to extensive and vary in direction. This suggests that AF variants affect AF risk through very diverse mechanisms,

some of them acting in the heart before AF and some not. Furthermore, there was no obvious relationship between the effect on conduction and the AF effect size, highlighting the complex relationship between normal cardiac electrophysiology and arrhythmia development.

The ECG analysis informed on mechanisms underlying associations of both coding and non-coding AF variants, for example through comparison of individual variants. The novel coding variants in *PLEC*, *RPL3L* and *MYZAP*, diversely affect ECG measurements. Both p.Gly4098Ser in *PLEC* and p.Ala75Val in *RPL3L* significantly affect measures of both atrial and ventricular conduction. Furthermore, both the previously reported missense variant in *MYH6* (Holm et al., 2011) and p.Gly4098Ser in *PLEC* have extensive effects on cardiac conduction and interestingly, the type of ECG effects reflect their pattern of expression. In the adult heart the expression of αMHC, the product of *MYH6*, is restricted to the atria, while plectin is equally expressed in atria and ventricles (Franco et al., 1998; "Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans," 2015). Consistent with this, the *MYH6* variant has a comparatively stronger effect on the atria and the AVN than the ventricles. Neither the splice-donor variant in *RPL3L* nor the missense variant in *MYZAP* associated significantly with measurements from sinus rhythm ECGs. However, when testing for association with ECG traits using all ECGs irrespective of rhythm and history of AF, an effect was detected for the *MYZAP* variant. This could be a direct consequence of AF occurring when ECG recordings are conducted or possibly these are effects on conduction in sinus rhythm that are detected because of increased sample size.

Several of the AF variants had only mild or no association with ECG traits in this dataset, whether AF cases were included in the analysis or not. Among them is the novel non-coding variant at *LINC01142/METTL11B* and the reported variant located 1 kb away. It is unclear by which mechanism and through which gene these variants affect the risk of AF. However, it is informative that unlike the two *LINC01142/METTL11B* variants, the common variant at the close by *PRRX1* locus (400 kb away) significantly affects cardiac conduction in sinus rhythm. This might suggest that the two *LINC01142/METTL11B* variants act through a common pathway different from the one of the *PRRX1* locus.

The fact that the *LINC01142/METTL11B* variants do not associate with ECG measurements does not repudiate or contradict their AF association. In fact, 51 variants showed no association with ECG measurements in the

meta-analysis reported in paper III, including several established AF variants such as the intronic variant in *ZFHX3*. This locus has been widely replicated across ethnic groups (Gudbjartsson et al., 2009; Li et al., 2011; Roselli et al., 2018) and the variant associates genome-wide significantly with AF in the current studies (papers I and III). Possibly, these variants have subtle effects on normal cardiac conduction but we lack power to detect them or they affect other traits than those represented here. Furthermore, they might increase risk of AF through mechanisms that do not affect cardiac conduction in sinus rhythm.

In some cases, the ECG associations can serve as an indirect replication of AF risk variants identified through GWASs. For example, the *PLEC* variant is almost exclusive to Iceland and only three carriers were identified in the replication sample. However, the strong and extensive effects on cardiac conduction, even in the absence of AF, support the association. Furthermore, the extent and direction of ECG effects might help inform on underlying biology of specific AF loci in future studies.

5.5 A missense variant in *MYH6* increasing risk of arrhythmia and coarctation of the aorta

The fourth and last paper of this dissertation reports a GWAS on CoA, identifying association with the p.Arg721Trp missense variant in *MYH6.* This is one of the strongest AF variants in paper I and it was initially discovered through its strong association with SSS (Holm et al., 2011). The allele frequency of p.Arg721Trp in Iceland is 0.34%, probably explained by the founder effect. In more outbred populations the variant is extremely rare or non-existent. The variant accounts for approximately 20% of CoA cases in Iceland and is the first variant associating with non-syndromic CoA on the population level. Our results, and previous publications (Holm et al., 2011), reveal a pleiotropic effect of the variant on cardiac function and disease. In addition to genome-wide significant effects on risk of SSS, AF and CoA it is also associated with additional CHDs, adult onset aortic valve stenosis, heart failure and stroke. Furthermore, our results revealed extensive effects of p.Arg721Trp on cardiac conduction and an association with increased left atrial diameter from echocardiogram measurements. The pleiotropic effect of p.Arg721Trp is in line with implication of other variants in *MYH6* in familial cases of cardiomyopathies (Carniel et al., 2005; Niimura et al., 2002) and cardiac septal defects (Ching et al., 2005; Posch et al., 2011). Furthermore, common variants in *MYH6* have been associated with heart rate and PR interval though GWASs (den Hoed et al., 2013; Eijgelsheim et al., 2010;

Holm et al., 2011).

Myosin is a major component of the sarcomere, the building block of the cardiac contractile system. It is an ATPase cellular motor protein consisting of two heavy chain subunits and two pairs of light chain subunits. *MYH6* encodes the alpha-myosin heavy chain (αMHC) subunit of myosin. The homologous beta-myosin heavy chain (βMHC) is encoded by *MYH7*. Both αMHC and βMHC are expressed throughout the heart during embryonic development. After birth, βMHC which is a relatively slow ATPase, continues to be expressed throughout the heart. However, the expression of αMHC, a fast ATPase, becomes restricted to the atria in the adult heart (Franco et al., 1998), but the expression pattern of these two genes continues to be dynamic. In heart failure, αMHC expression is downregulated while βMHC is upregulated resulting in diminished cardiac performance and worse outcome (Hang et al., 2010; Herron & McDonald, 2002; Krenz & Robbins, 2004). Expression of αMHC has not been detected in the aorta ("Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans," 2015).

It is possible that p.Arg721Trp increases risk of CoA by reducing blood flow through the fetal aorta because of diminished contraction of the developing heart. It would thus comply with the hemodynamic theory of CoA pathogenesis which proposes that cardiac lesions resulting in decreased left ventricular outflow promote development of CoA by reducing blood flow through the fetal aorta (Rudolph et al., 1972). This theory of mechanism is supported by the fact that αMHC is expressed in both atria and ventricles during cardiac development, but has not been detected in the aorta where the coarctation occurs. It is further supported by our ECG analysis revealing extensive effects of the variant on cardiac electrical function, including in the ventricles. Finally, p.Arg721Trp has been overexpressed in neonatal rat ventricular cardiomyocytes, which resulted in impaired sarcomere structures (Ishikawa et al., 2015), further supporting that it may impair contraction and outflow from the developing ventricles.

The associations of coding variants in *MYH6* with SSS, heart rate and PR interval in GWASs represented a novel implication of the sarcomere in cardiac conduction and arrhythmogenesis in the absence of apparent cardiomyopathy. The exact role of αMHC in conduction is not known. One suggested mechanism is that p.Arg721Trp, and other *MYH6* variants affecting heart rate, modulate the expression of proteins with specific roles in conduction (Ishikawa et al., 2015). In fact, evidence has shown that a highly

conserved microRNA, miR-208a, encoded by intron 27 of *MYH6* is required for expression of cardiac transcription factors and the gap junction protein connexin 40 (Callis et al., 2009). Connexin 40 is encoded by *GJA5*, which is specifically expressed in the atria (Chaldoupi et al., 2009; "Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans," 2015) and is one of the prioritized candidate genes identified in the AF GWAS in paper III.

As discussed previously, the role of cardiac structure in arrhythmogenesis has been further established in GWASs on AF in this dissertation and other publications, but the p.Arg721Trp in *MYH6* stands out from other AF variants for its strong association with SSS. This has led to speculations that p.Arg721Trp preferentially impairs the pacemaker function of the SA node or the propagation of impulse from the node. To our knowledge, no direct link has been made between *MYH6* and SA node function specifically. Whether or not such mechanisms contribute to the high risk of SSS in carriers of p.Arg721Trp, it would not explain the pleiotropic effects of the variant on cardiac function, especially its extensive effects on ECG measurements in atria and ventricles in the absence of arrhythmia and the association with CHDs. To continue the discussion in chapter 5.2 on the AF effects of p.Arg721Trp, these pleiotropic effects therefore suggest that the effect of p.Arg721Trp on AF risk is likely not mediated by interruption of pacemaker function and SSS only.

5.6 The important role of cardiac structure in the development of arrhythmias

As discussed previously, a recurrent theme in the results of the three GWASs on AF conducted in this thesis is emphasis on the importance of myocardial structural changes in arrhythmogenesis. This assumption is drawn from associations with coding variants in genes with a structural role, *PLEC, MYZAP* and *RPL3L*, as well as interpretation of over one hundred associations with mostly common non-coding variants from the meta-analysis. Our results add to the insights gained from implication of coding variants in the sarcomere genes *MYH6* and *MYL4* in SSS and AF (Gudbjartsson et al., 2017; Holm et al., 2011). Furthermore, the importance of cardiac structure has recently been supported by other GWASs also implicating genes and pathways involving cardiac mechanics in the pathogenesis of AF (Christophersen et al., 2017; Roselli et al., 2018). It is important to note that the numerous AF loci identified through GWASs hitherto point to a very complex pathophysiology and involvement of diverse

pathways. However, the strong and relatively novel implication of cardiac structural remodeling in the development of AF warrants further consideration.

Cardiac remodeling is defined as a process resulting in myocardial cellular and interstitial changes including myocyte hypertrophy, necrosis (Tan et al., 1991), apoptosis (Olivetti et al., 1997) and fibrosis (Anderson et al., 1979; Cohn et al., 2000). The main consequence of these molecular changes is cardiac dysfunction (Azevedo et al., 2016). Cardiac remodeling can result from extrinsic or intrinsic stress such as increased hemodynamic load and neurohormonal activation (Bisping et al., 2014; Kehat & Molkentin, 2010). Several diseases or conditions have been shown to trigger cardiac remodeling, including hypertension, inflammation, fibrosis and cardiac ischemia (Burchfield et al., 2013; Pfeffer & Braunwald, 1990). On a molecular level, many signal-tansduction pathways have been implicated in cardiac hypertrophy, including for example the Rho-dependent SRF signaling pathway that can be activated by myozap overexpression, as previously discussed. Historically, cardiac remodeling has been considered a beneficial adaptive process but is now also viewed as potentially pathological, as it has been associated with activation of maladaptive pathways that eventually lead to cardiomyopathy (Frey & Olson, 2003; Selvetella et al., 2004).

The epidemiological evidence for the role of cardiac structural remodeling in the pathogenesis of AF and other arrhythmias has in fact existed for many years since conditions known to induce hypertrophy are established risk factors for AF, including hypertension, ischemia and heart failure (Huxley et al., 2011). On a molecular level, the effects of structural remodeling on the pathogenesis of AF are likely mediated through promoting both rapid ectopic activity and reentry, the two main drivers of AF development and sustainability. The effect of cardiac remodeling on reentry is considered more important. A combination of atrial dilation and fibrosis creates conduction disparity, imposes barriers and slows conduction favoring the initiation and maintenance of multiple irregular reentry circuits that sustain AF. Ectopic firing is usually required as a trigger for the reentry and evidence suggests that structural remodeling can at least indirectly induce ectopic firing through causing increased atrial expression of ion channel subunits promoting ectopic beats (Nattel et al., 2008). Through the years, the literature has focused mainly on ventricular cardiomyopathies, which have been well classified (Elliott et al., 2014) and their consequences extensively studied in human subjects and experimental models (Cook et al., 2017). For example, AF is the most frequent arrhythmia in hypertrophic cardiomyopathy thought to result

from general myocardial fibrosis and increased left atrial pressure caused by elevated left ventricular pressure and mitral valve insufficiency (Patten et al., 2018).

In recent years atrial cardiomyopathy has been defined as a clinical entity and its role in the development of AF increasingly recognized (Goette et al., 2016; Goldberger et al., 2015; Hammwohner et al., 2019). An atrial cardiomyopathy is defined as “any complex of structural, architectural, contractile or electrophysiological changes affecting the atria with the potential to produce clinically-relevant manifestations” (Goette et al., 2016). Factors such as hypertension, diabetes, heart failure and aging are known contributors to atrial cardiomyopathy. The pathological alterations vary greatly between individuals and the condition is not routinely assessed or diagnosed. The associations of variants in structural genes with AF risk have underlined the importance of atrial myopathy in AF development, especially coding variants in the atria specific myosin genes *MYH6* and *MYL4*. In our study we did not detect associations of AF risk variants in *PLEC*, *RPL3L* and *MYZAP* with clinically overt cardiomyopathy, but a frequently difficult to diagnose and likely subclinical atrial cardiomyopathy might be an underlying mechanism.

In addition to serving as a substrate for the development of AF (Boldt et al., 2004; Chimenti et al., 2010; Marrouche et al., 2014; Nattel & Harada, 2014), atrial myopathy has further implication for the diagnosis and management of the disease. Evidence suggests that assessing the degree of atrial fibrosis in patients with AF may provide a metric of disease severity and possibly predict response to therapy, specifically ablation (Oakes et al., 2009). In addition, fibrotic atrial cardiomyopathy has been independently associated with prior history of stroke in AF patients (Daccarett et al., 2011). Furthermore, adding the assessment of left atrial fibrosis by delayed-enhanced magnetic resonance imaging has been shown to increase the predictive performance of the conventional scores for assessing stroke risk in AF (Daccarett et al., 2011). Based on the temporal dissociation between timing of AF and occurrence of stroke it has even been suggested that prothrombotic atrial tissue might be an important cause of thrombus formation in patients with AF, independent of the atrial rhythm (Calenda et al., 2016). Despite these indications, a direct causal link between atrial fibrosis and stroke risk in AF has not been clearly established. However, the accumulating evidence strongly suggest that assessment of atrial fibrosis may guide in management decisions, but further research on its applicability in clinical practice is warranted.

5.7 Clinical implications of the results

A deeper understanding of the pathophysiology of AF is a critical prerequisite to developing new and more effective treatment options for this common arrhythmia. Advances in this area are urgently required as currently available treatment options have limited efficacy and may carry risk of serious adverse effects. Emphasis on the importance of impaired cardiac structure in AF pathogenesis composes a stark contrast to the previous view of AF as mainly an ion channel disorder, a theory initially supported by the earliest candidate gene studies on rare familial AF. The new perspective provided by our studies should lead to new ideas in exploring pathways involved in structural remodeling in AF as potential therapeutic targets. It is possible that future studies on the novel AF loci identified here might lead to identification of specific novel drug targets for AF treatment. The results also highlight the importance of maintaining normal cardiac structure for the prevention of AF, for example by treating hypertension and other conditions contributing to cardiac remodeling.

The numerous AF variants that have now been discovered in GWASs could be a valuable tool for risk stratification and personalization of diagnosis and treatment. Polygenic risk scores could be applied to identify individuals with high risk of developing AF and its complications, including stroke, thus encouraging a more individualized management approach. Likewise, genetic risk scores have been proposed as an addition to conventional risk prediction in coronary artery disease (Inouye et al., 2018). Future studies might identify treatment options that are particularly effective for carriers of specific AF variants, thus personalizing choice of treatment based on genotype.

The discovery of a coding variant in the cardiac myosin gene *MYH6* explaining 20% of CoA cases in Iceland has several practical implications. The association provides increased understanding of pathogenesis which might guide future studies on CoA and its complications and contribute to advances in treatment. Furthermore, the discovery might have direct meaning for genetic counselling for Icelandic individuals with CoA and their families. Although the penetrance for CoA is low (1/123) the variant explains a large proportion of CoA cases in Iceland. Our analysis revealed that, compared to non-carriers with CoA, carriers of the variant are more likely to present with mild rather than more critical forms of CoA. The symptoms of mild CoA are subtle and delayed diagnosis is common (Strafford et al., 1982). Identifying carriers among CoA cases and searching for the variant

among their close relatives might therefore promote early diagnosis and treatment of CoA among them.

5.8 Limitations and strengths

There are several limitations to the studies. For example, the p.Gly4098Ser variant in *PLEC* (paper I) is very rare outside of Iceland and we are therefore not able to replicate the AF association for this variant. The same applies to the associations of p.Arg721Trp in *MYH6* with various CVDs (papers I and IV). However, both genes were already strongly implicated in cardiac diseases, *PLEC* in cardiomyopathy in patients with EBS (Bolling et al., 2010; Celik et al., 2005; Schroder et al., 2002; Villa et al., 2015) and *MYH6* in for example cardiac septal defects (Ching et al., 2005; Posch et al., 2011) and familial cardiomyopathy (Carniel et al., 2005; Niimura et al., 2002). Furthermore, we show that both variants have extensive effects on cardiac conduction in the absence of arrhythmia and other variants in *MYH6* have also been associated with ECG measurements outside of Iceland (den Hoed et al., 2013; Eijgelsheim et al., 2010). Another limitation is the small size of the CoA sample set. A larger set might have facilitated detection of more variants associating with risk of CoA. It would also have allowed a better estimate of the fraction of CoA cases explained by the variant, as well as improved OR estimates for comparison between carriers and non-carriers of p.Arg721Trp with CoA. Performing a large meta-analysis of AF combining six cohorts (paper III) has the advantages of increased power to detect associations but a limitation is that the heterogeneous effects across studies can attenuate the predictive power of polygenic scores and heritability estimates (de Vlaming et al., 2017), as previously discussed.

The main strength of the study relates to the extensive dataset that has been collected at deCODE genetics during the past 20 years, including genotype data for a large proportion of the Icelandic nation. Several characteristics of the Icelandic population, such as the well documented genealogy and the founder effect are advantageous for genetic studies, in particular for discoveries of rare variant associations. Furthermore, the large Icelandic AF cohort likely includes the majority of people diagnosed with AF in Iceland during the period under study. Despite the small CoA sample size, the study also likely includes the majority of Icelandic cases, since this is a rare condition and only one hospital delivers specialized care for patients with CHDs. Furthermore, a major strength of the study is the ability to combine variable phenotypic data, such as disease diagnoses according to ICD codes

and ECG measurements reflecting cardiac conduction, to obtain a broad picture of the effects of sequence variants on cardiac function and disease.

6 Conclusions

Through genome-wide association studies we have identified novel AF associations with coding variants in three genes which all have a role in maintaining cardiac structure, *PLEC*, *MYZAP* and *RPL3L*. This further establishes the notion of a role for cardiac mechanics, and possibly atrial myopathy in the development of atrial arrhythmia, as implied by the previously published associations of coding variants in *MYH6* and *MYL4* with AF. We also report the results of a large meta-analysis of AF GWASs including over one million participants. Pathway and functional enrichment analysis performed on the AF associated loci further highlighted the importance of cardiac structural remodeling in the development of AF. This new understanding of AF pathogenesis supports a shift in focus towards pathways involved in cardiac mechanics for identifying novel treatment options in AF. It also supports the notion that evaluating myocardial, and specifically, atrial structure may be of value in routine clinical practice. In accordance with our initial aims we have provided increased understanding of AF pathogenesis, a crucial step towards improving the currently limited treatment options available for this common disease that poses a growing public health concern.

In addition to highlighting the role of myocardial structure, several pathways were implicated through analysis of AF associations, including cardiac development and calcium handling. This underscores the complexity of AF pathogenesis. Furthermore, the substantial increase in AF variants might translate into clinical relevance by improving calculations of genetic risk scores for AF. These may be utilized to predict risk of AF and its complications, for example stroke, a serious complication of AF that is difficult to predict with currently available methods.

Our analysis of ECG effects of all published and novel AF variants further emphasizes the complexity of AF genetics. The results suggest diverse mechanisms underlying the AF risk of these variants, reflected by their effects on normal cardiac conduction ranging from none to extensive. These diverse mechanisms also support the notion that personalized treatment based on genotype could become a future option in AF. In that regard, ECG effects of AF variants might help identify underlying biology of specific AF loci in future studies.

One of the strongest association with AF in our dataset is with p.Arg721Trp in the cardiac myosin gene *MYH6*, previously reported in SSS. A suggestive association of this variant with CoA led us to explore the variant further and perform a GWAS on CoA, revealing that the variant explains 20% of CoA cases in Iceland and is more likely to result in mild, rather than critical CoA. The pleiotropic effect of p.Arg721Trp on cardiac function and disease suggests that the variant affects risk of AF independently, not only as a result of SSS. The association with CoA increases understanding on its pathogenesis as it complies with the hemodynamic theory of CoA development. Since this missense variant in *MYH6* explains a large proportion of CoA cases in Iceland, the association might have relevance for genetic counseling for individuals with CoA and their families.

6.1 Future perspectives

Future research is expected and has the potential to increase understanding on the genetic architecture of AF, as evident by the small proportion of overall heritability explained by the variants identified here. Further advances could be facilitated by a combination of factors, including increasing sample sizes by continuing collaboration of study groups and more detailed phenotyping of participants. In addition, improved quality of genotype data by continuing efforts in large scale whole-genome sequencing and the application of long-read technologies will also promote progress (van Dijk et al., 2018). This is also the case in CoA, as our GWAS was the first to be performed. Studies in other populations might also identify rare deleterious variants explaining large proportion of CoA or other CHDs. Further research on the clinical utility of genetic risk scores for predicting risk of AF and its complications is warranted. Furthermore, the strong implication of myocardial structure in AF pathogenesis encourages research on methods to evaluate myocardial structure in AF patients and the clinical utility of such evaluation. Finally, the findings presented here can be used as a foundation for functional studies directed at further elucidating mechanisms underlying AF and CoA. The studies performed here have increased understanding of the pathogenesis of both AF and CoA and will hopefully provide a step towards improvements in the evaluation and treatment of both diseases.

References

Adalsteinsdottir, B., Teekakirikul, P., Maron, B. J., Burke, M. A., Gudbjartsson, D. F., Holm, H., Stefansson, K., DePalma, S. R., Mazaika, E., McDonough, B., Danielsen, R., Seidman, J. G., Seidman, C. E., & Gunnarsson, G. T. (2014). Nationwide study on hypertrophic cardiomyopathy in Iceland: evidence of a *MYBPC3* founder mutation. *Circulation, 130*(14), 1158-1167.

Adams, H. P., Jr., Bendixen, B. H., Kappelle, L. J., Biller, J., Love, B. B., Gordon, D. L., & Marsh, E. E., 3rd. (1993). Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke, 24*(1), 35-41.

Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics, Chapter 7*, Unit7.20-Unit27.20.

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S., & Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nat Methods, 7*(4), 248-249.

Akar, J. G., Jeske, W., & Wilber, D. J. (2008). Acute onset human atrial fibrillation is associated with local cardiac platelet activation and endothelial dysfunction. *J Am Coll Cardiol, 51*(18), 1790-1793.

Allen, H. D., Shaddy, R. E., Penny, D. J., Cetta, F., & Feltes, T. F. (2016). *Moss and Adams' heart disease in infants, children, and adolescents : including the fetus and young adult* (9th ed.). Philadelphia: Wolters Kluwer.

Altshuler, D., Daly, M. J., & Lander, E. S. (2008). Genetic mapping in human disease. *Science, 322*(5903), 881-888.

Anderson, K. R., Sutton, M. G., & Lie, J. T. (1979). Histopathological types of cardiac fibrosis in myocardial disease. *J Pathol, 128*(2), 79-85.

Anderson, R. H., Lenox, C. C., & Zuberbuhler, J. R. (1983). Morphology of ventricular septal defect associated with coarctation of aorta. *Br Heart J, 50*(2), 176-181.

Andra, K., Lassmann, H., Bittner, R., Shorny, S., Fassler, R., Propst, F., & Wiche, G. (1997). Targeted inactivation of plectin reveals essential

function in maintaining the integrity of skin, muscle, and heart cytoarchitecture. *Genes Dev, 11*(23), 3143-3156.

Andrade, J., Khairy, P., Dobrev, D., & Nattel, S. (2014). The clinical profile and pathophysiology of atrial fibrillation: relationships among clinical features, epidemiology, and mechanisms. *Circ Res, 114*(9), 1453-1468.

Arnar, D. O., Thorvaldsson, S., Manolio, T. A., Thorgeirsson, G., Kristjansson, K., Hakonarson, H., & Stefansson, K. (2006). Familial aggregation of atrial fibrillation in Iceland. *Eur Heart J, 27*(6), 708-712.

Arrington, C. B., Bleyl, S. B., Matsunami, N., Bonnell, G. D., Otterud, B. E., Nielsen, D. C., Stevens, J., Levy, S., Leppert, M. F., & Bowles, N. E. (2012). Exome analysis of a family with pleiotropic congenital heart disease. *Circ Cardiovasc Genet, 5*(2), 175-182.

Asad, Z., Abbas, M., Javed, I., Korantzopoulos, P., & Stavrakis, S. (2018). Obesity is associated with incident atrial fibrillation independent of gender: A meta-analysis. *J Cardiovasc Electrophysiol, 29*(5), 725-732.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics, 25*(1), 25-29.

Aune, D., Feng, T., Schlesinger, S., Janszky, I., Norat, T., & Riboli, E. (2018). Diabetes mellitus, blood glucose and the risk of atrial fibrillation: A systematic review and meta-analysis of cohort studies. *J Diabetes Complications, 32*(5), 501-511.

Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., & Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature, 526*(7571), 68-74.

Azevedo, P. S., Polegato, B. F., Minicucci, M. F., Paiva, S. A. R., & Zornoff, L. A. M. (2016). Cardiac Remodeling: Concepts, Clinical Impact, Pathophysiological Mechanisms and Pharmacologic Treatment. *Arquivos brasileiros de cardiologia, 106*(1), 62-69.

Banwell, B. L., Russel, J., Fukudome, T., Shen, X. M., Stilling, G., & Engel, A. G. (1999). Myopathy, myasthenic syndrome, and epidermolysis bullosa simplex due to plectin deficiency. *J Neuropathol Exp Neurol, 58*(8), 832-846.

Bapat, A., Anderson, C. D., Ellinor, P. T., & Lubitz, S. A. (2018). Genomic basis of atrial fibrillation. *Heart, 104*(3), 201-206.

Ben-Shem, A., Garreau de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G., & Yusupov, M. (2011). The structure of the eukaryotic ribosome at 3.0 A resolution. *Science, 334*(6062), 1524-1529.

Benjamin, E. J., Muntner, P., Alonso, A., Bittencourt, M. S., Callaway, C. W., Carson, A. P., Chamberlain, A. M., Chang, A. R., Cheng, S., Das, S. R., Delling, F. N., Djousse, L., Elkind, M. S. V., Ferguson, J. F., Fornage, M., Jordan, L. C., Khan, S. S., Kissela, B. M., Knutson, K. L., Kwan, T. W., Lackland, D. T., Lewis, T. T., Lichtman, J. H., Longenecker, C. T., Loop, M. S., Lutsey, P. L., Martin, S. S., Matsushita, K., Moran, A. E., Mussolino, M. E., O'Flaherty, M., Pandey, A., Perak, A. M., Rosamond, W. D., Roth, G. A., Sampson, U. K. A., Satou, G. M., Schroeder, E. B., Shah, S. H., Spartano, N. L., Stokes, A., Tirschwell, D. L., Tsao, C. W., Turakhia, M. P., VanWagner, L. B., Wilkins, J. T., Wong, S. S., & Virani, S. S. (2019). Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association. *Circulation, 139*(10), e56-e528.

Benjamin, E. J., Rice, K. M., Arking, D. E., Pfeufer, A., van Noord, C., Smith, A. V., Schnabel, R. B., Bis, J. C., Boerwinkle, E., Sinner, M. F., Dehghan, A., Lubitz, S. A., D'Agostino, R. B., Sr., Lumley, T., Ehret, G. B., Heeringa, J., Aspelund, T., Newton-Cheh, C., Larson, M. G., Marciante, K. D., Soliman, E. Z., Rivadeneira, F., Wang, T. J., Eiriksdottir, G., Levy, D., Psaty, B. M., Li, M., Chamberlain, A. M., Hofman, A., Vasan, R. S., Harris, T. B., Rotter, J. I., Kao, W. H., Agarwal, S. K., Stricker, B. H., Wang, K., Launer, L. J., Smith, N. L., Chakravarti, A., Uitterlinden, A. G., Wolf, P. A., Sotoodehnia, N., Kottgen, A., van Duijn, C. M., Meitinger, T., Mueller, M., Perz, S., Steinbeck, G., Wichmann, H. E., Lunetta, K. L., Heckbert, S. R., Gudnason, V., Alonso, A., Kaab, S., Ellinor, P. T., & Witteman, J. C. (2009). Variants in *ZFHX3* are associated with atrial fibrillation in individuals of European ancestry. *Nat Genet, 41*(8), 879-881.

Benjamin, E. J., Wolf, P. A., D'Agostino, R. B., Silbershatz, H., Kannel, W. B., & Levy, D. (1998). Impact of atrial fibrillation on the risk of death: the Framingham Heart Study. *Circulation, 98*(10), 946-952.

Benonisdottir, S., Oddsson, A., Helgason, A., Kristjansson, R. P., Sveinbjornsson, G., Oskarsdottir, A., Thorleifsson, G., Davidsson, O. B., Arnadottir, G. A., Sulem, G., Jensson, B. O., Holm, H., Alexandersson, K. F., Tryggvadottir, L., Walters, G. B., Gudjonsson, S. A., Ward, L. D., Sigurdsson, J. K., Iordache, P. D., Frigge, M. L.,

Rafnar, T., Kong, A., Masson, G., Helgason, H., Thorsteinsdottir, U., Gudbjartsson, D. F., Sulem, P., & Stefansson, K. (2016). Epigenetic and genetic components of height regulation. *Nat Commun, 7*, 13490.

Benson, D. W., Silberbach, G. M., Kavanaugh-McHugh, A., Cottrill, C., Zhang, Y., Riggs, S., Smalls, O., Johnson, M. C., Watson, M. S., Seidman, J. G., Seidman, C. E., Plowden, J., & Kugler, J. D. (1999). Mutations in the cardiac transcription factor *NKX2.5* affect diverse cardiac developmental pathways. *J Clin Invest, 104*(11), 1567-1573.

Berry, F. B., Miura, Y., Mihara, K., Kaspar, P., Sakata, N., Hashimoto-Tamaoki, T., & Tamaoki, T. (2001). Positive and negative regulation of myogenic differentiation of C2C12 cells by isoforms of the multiple homeodomain zinc finger transcription factor ATBF1. *J Biol Chem, 276*(27), 25057-25065.

Bisping, E., Wakula, P., Poteser, M., & Heinzel, F. R. (2014). Targeting cardiac hypertrophy: toward a causal heart failure therapy. *J Cardiovasc Pharmacol, 64*(4), 293-305.

Blake, J. A., Bult, C. J., Eppig, J. T., Kadin, J. A., & Richardson, J. E. (2014). The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse. *Nucleic Acids Res, 42*(Database issue), D810-817.

Boldt, A., Wetzel, U., Lauschke, J., Weigl, J., Gummert, J., Hindricks, G., Kottkamp, H., & Dhein, S. (2004). Fibrosis in left atrial tissue of patients with atrial fibrillation with and without underlying mitral valve disease. *Heart, 90*(4), 400-405.

Bolling, M. C., Pas, H. H., de Visser, M., Aronica, E., Pfendner, E. G., van den Berg, M. P., Diercks, G. F., Suurmeijer, A. J., & Jonkman, M. F. (2010). *PLEC1* mutations underlie adult-onset dilated cardiomyopathy in epidermolysis bullosa simplex with muscular dystrophy. *J Invest Dermatol, 130*(4), 1178-1181.

Bonachea, E. M., Chang, S. W., Zender, G., LaHaye, S., Fitzgerald-Butt, S., McBride, K. L., & Garg, V. (2014). Rare *GATA5* sequence variants identified in individuals with bicuspid aortic valve. *Pediatr Res, 76*(2), 211-216.

Borrmann, C. M., Grund, C., Kuhn, C., Hofmann, I., Pieperhoff, S., & Franke, W. W. (2006). The area composita of adhering junctions connecting heart muscle cells of vertebrates. II. Colocalizations of desmosomal and fascia adhaerens molecules in the intercalated disk. *Eur J Cell Biol, 85*(6), 469-485.

Brauch, K. M., Karst, M. L., Herron, K. J., de Andrade, M., Pellikka, P. A., Rodeheffer, R. J., Michels, V. V., & Olson, T. M. (2009). Mutations in ribonucleic acid binding protein gene cause familial dilated cardiomyopathy. *J Am Coll Cardiol, 54*(10), 930-941.

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol, 34*(5), 525-527.

Brugada, R., Tapscott, T., Czernuszewicz, G. Z., Marian, A. J., Iglesias, A., Mont, L., Brugada, J., Girona, J., Domingo, A., Bachinski, L. L., & Roberts, R. (1997). Identification of a genetic locus for familial atrial fibrillation. *N Engl J Med, 336*(13), 905-911.

Bulik-Sullivan, B. K., Loh, P. R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., Daly, M. J., Price, A. L., & Neale, B. M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet, 47*(3), 291-295.

Burchfield, J. S., Xie, M., & Hill, J. A. (2013). Pathological ventricular remodeling: mechanisms: part 1 of 2. *Circulation, 128*(4), 388-400.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., McVean, G., Leslie, S., Donnelly, P., & Marchini, J. (2017). Genome-wide genetic data on ~500,000 UK Biobank participants. doi: https://doi.org/10.1101/166298.

Cai, H., Li, Z., Goette, A., Mera, F., Honeycutt, C., Feterik, K., Wilcox, J. N., Dudley, S. C., Jr., Harrison, D. G., & Langberg, J. J. (2002). Downregulation of endocardial nitric oxide synthase expression and nitric oxide production in atrial fibrillation: potential mechanisms for atrial thrombosis and stroke. *Circulation, 106*(22), 2854-2858.

Calenda, B. W., Fuster, V., Halperin, J. L., & Granger, C. B. (2016). Stroke risk assessment in atrial fibrillation: risk factors and markers of atrial myopathy. *Nat Rev Cardiol, 13*(9), 549-559.

Callis, T. E., Pandya, K., Seok, H. Y., Tang, R. H., Tatsuguchi, M., Huang, Z. P., Chen, J. F., Deng, Z., Gunn, B., Shumate, J., Willis, M. S., Selzman, C. H., & Wang, D. Z. (2009). MicroRNA-208a is a regulator of cardiac hypertrophy and conduction in mice. *J Clin Invest, 119*(9), 2772-2786.

Carey, D. J., Fetterolf, S. N., Davis, F. D., Faucett, W. A., Kirchner, H. L., Mirshahi, U., Murray, M. F., Smelser, D. T., Gerhard, G. S., & Ledbetter, D. H. (2016). The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research. *Genet Med, 18*(9), 906-913.

Carniel, E., Taylor, M. R., Sinagra, G., Di Lenarda, A., Ku, L., Fain, P. R., Boucek, M. M., Cavanaugh, J., Miocic, S., Slavov, D., Graw, S. L., Feiger, J., Zhu, X. Z., Dao, D., Ferguson, D. A., Bristow, M. R., & Mestroni, L. (2005). Alpha-myosin heavy chain: a sarcomeric gene associated with dilated and hypertrophic phenotypes of cardiomyopathy. *Circulation, 112*(1), 54-59.

Castanon, M. J., Walko, G., Winter, L., & Wiche, G. (2013). Plectin-intermediate filament partnership in skin, skeletal muscle, and peripheral nerve. *Histochem Cell Biol, 140*(1), 33-53.

Celik, C., Uysal, H., Heper, A. O., & Karaoglan, B. (2005). Epidermolysis bullosa simplex associated with muscular dystrophy and cardiac involvement. *J Clin Neuromuscul Dis, 6*(4), 157-161.

Chaillou, T., Lee, J. D., England, J. H., Esser, K. A., & McCarthy, J. J. (2013). Time course of gene expression during mouse skeletal muscle hypertrophy. *J Appl Physiol (1985), 115*(7), 1065-1074.

Chaillou, T., Zhang, X., & McCarthy, J. J. (2016). Expression of Muscle-Specific Ribosomal Protein L3-Like Impairs Myotube Growth. *J Cell Physiol, 231*(9), 1894-1902.

Chaldoupi, S. M., Loh, P., Hauer, R. N., de Bakker, J. M., & van Rijen, H. V. (2009). The role of connexin40 in atrial fibrillation. *Cardiovasc Res, 84*(1), 15-23.

Chang, H. Y., Lin, Y. J., Lo, L. W., Chang, S. L., Hu, Y. F., Li, C. H., Chao, T. F., Yin, W. H., & Chen, S. A. (2013). Sinus node dysfunction in atrial fibrillation patients: the evidence of regional atrial substrate remodelling. *Europace, 15*(2), 205-211.

Chang, R. K., Gurvitz, M., & Rodriguez, S. (2008). Missed diagnosis of critical congenital heart disease. *Arch Pediatr Adolesc Med, 162*(10), 969-974.

Chen, Y. C., Chen, S. A., Chen, Y. J., Chang, M. S., Chan, P., & Lin, C. I. (2002). Effects of thyroid hormone on the arrhythmogenic activity of pulmonary vein cardiomyocytes. *J Am Coll Cardiol, 39*(2), 366-372.

Chen, Y. H., Xu, S. J., Bendahhou, S., Wang, X. L., Wang, Y., Xu, W. Y., Jin, H. W., Sun, H., Su, X. Y., Zhuang, Q. N., Yang, Y. Q., Li, Y. B., Liu, Y., Xu, H. J., Li, X. F., Ma, N., Mou, C. P., Chen, Z., Barhanin, J., & Huang, W. (2003). *KCNQ1* gain-of-function mutation in familial atrial fibrillation. *Science, 299*(5604), 251-254.

Cheng, S., Keyes, M. J., Larson, M. G., McCabe, E. L., Newton-Cheh, C., Levy, D., Benjamin, E. J., Vasan, R. S., & Wang, T. J. (2009). Long-term

outcomes in individuals with prolonged PR interval or first-degree atrioventricular block. *Jama, 301*(24), 2571-2577.

Cheung, V. G., & Spielman, R. S. (2009). Genetics of human gene expression: mapping DNA variants that influence gene expression. *Nat Rev Genet, 10*(9), 595-604.

Chimenti, C., Russo, M. A., Carpi, A., & Frustaci, A. (2010). Histological substrate of human atrial fibrillation. *Biomed Pharmacother, 64*(3), 177-183.

Ching, Y. H., Ghosh, T. K., Cross, S. J., Packham, E. A., Honeyman, L., Loughna, S., Robinson, T. E., Dearlove, A. M., Ribas, G., Bonser, A. J., Thomas, N. R., Scotter, A. J., Caves, L. S., Tyrrell, G. P., Newbury-Ecob, R. A., Munnich, A., Bonnet, D., & Brook, J. D. (2005). Mutation in myosin heavy chain 6 causes atrial septal defect. *Nat Genet, 37*(4), 423-428.

Choi, H. J., Park-Snyder, S., Pascoe, L. T., Green, K. J., & Weis, W. I. (2002). Structures of two intermediate filament-binding fragments of desmoplakin reveal a unique repeat motif structure. *Nat Struct Biol, 9*(8), 612-620.

Choi, Y., & Chan, A. P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics, 31*(16), 2745-2747.

Christophersen, I. E., Budtz-Jorgensen, E., Olesen, M. S., Haunso, S., Christensen, K., & Svendsen, J. H. (2013). Familial atrial fibrillation predicts increased risk of mortality: a study in Danish twins. *Circ Arrhythm Electrophysiol, 6*(1), 10-15.

Christophersen, I. E., & Ellinor, P. T. (2016). Genetics of atrial fibrillation: from families to genomes. *J Hum Genet, 61*(1), 61-70.

Christophersen, I. E., Ravn, L. S., Budtz-Joergensen, E., Skytthe, A., Haunsoe, S., Svendsen, J. H., & Christensen, K. (2009). Familial aggregation of atrial fibrillation: a study in Danish twins. *Circ Arrhythm Electrophysiol, 2*(4), 378-383.

Christophersen, I. E., Rienstra, M., Roselli, C., Yin, X., Geelhoed, B., Barnard, J., Lin, H., Arking, D. E., Smith, A. V., Albert, C. M., Chaffin, M., Tucker, N. R., Li, M., Klarin, D., Bihlmeyer, N. A., Low, S. K., Weeke, P. E., Muller-Nurasyid, M., Smith, J. G., Brody, J. A., Niemeijer, M. N., Dorr, M., Trompet, S., Huffman, J., Gustafsson, S., Schurmann, C., Kleber, M. E., Lyytikainen, L. P., Seppala, I., Malik, R., Horimoto, A., Perez, M., Sinisalo, J., Aeschbacher, S., Theriault, S., Yao, J., Radmanesh, F., Weiss, S., Teumer, A., Choi, S. H., Weng, L. C., Clauss,

S., Deo, R., Rader, D. J., Shah, S. H., Sun, A., Hopewell, J. C., Debette, S., Chauhan, G., Yang, Q., Worrall, B. B., Pare, G., Kamatani, Y., Hagemeijer, Y. P., Verweij, N., Siland, J. E., Kubo, M., Smith, J. D., Van Wagoner, D. R., Bis, J. C., Perz, S., Psaty, B. M., Ridker, P. M., Magnani, J. W., Harris, T. B., Launer, L. J., Shoemaker, M. B., Padmanabhan, S., Haessler, J., Bartz, T. M., Waldenberger, M., Lichtner, P., Arendt, M., Krieger, J. E., Kahonen, M., Risch, L., Mansur, A. J., Peters, A., Smith, B. H., Lind, L., Scott, S. A., Lu, Y., Bottinger, E. B., Hernesniemi, J., Lindgren, C. M., Wong, J. A., Huang, J., Eskola, M., Morris, A. P., Ford, I., Reiner, A. P., Delgado, G., Chen, L. Y., Chen, Y. I., Sandhu, R. K., Li, M., Boerwinkle, E., Eisele, L., Lannfelt, L., Rost, N., Anderson, C. D., Taylor, K. D., Campbell, A., Magnusson, P. K., Porteous, D., Hocking, L. J., Vlachopoulou, E., Pedersen, N. L., Nikus, K., Orho-Melander, M., Hamsten, A., Heeringa, J., Denny, J. C., Kriebel, J., Darbar, D., Newton-Cheh, C., Shaffer, C., Macfarlane, P. W., Heilmann-Heimbach, S., Almgren, P., Huang, P. L., Sotoodehnia, N., Soliman, E. Z., Uitterlinden, A. G., Hofman, A., Franco, O. H., Volker, U., Jockel, K. H., Sinner, M. F., Lin, H. J., Guo, X., Dichgans, M., Ingelsson, E., Kooperberg, C., Melander, O., Loos, R. J. F., Laurikka, J., Conen, D., Rosand, J., van der Harst, P., Lokki, M. L., Kathiresan, S., Pereira, A., Jukema, J. W., Hayward, C., Rotter, J. I., Marz, W., Lehtimaki, T., Stricker, B. H., Chung, M. K., Felix, S. B., Gudnason, V., Alonso, A., Roden, D. M., Kaab, S., Chasman, D. I., Heckbert, S. R., Benjamin, E. J., Tanaka, T., Lunetta, K. L., Lubitz, S. A., & Ellinor, P. T. (2017). Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nat Genet, 49*(6), 946-952.

Chugh, S. S., Havmoeller, R., Narayanan, K., Singh, D., Rienstra, M., Benjamin, E. J., Gillum, R. F., Kim, Y. H., McAnulty, J. H., Jr., Zheng, Z. J., Forouzanfar, M. H., Naghavi, M., Mensah, G. A., Ezzati, M., & Murray, C. J. (2014). Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. *Circulation, 129*(8), 837-847.

Cohen, M., Fuster, V., Steele, P. M., Driscoll, D., & McGoon, D. C. (1989). Coarctation of the aorta. Long-term follow-up and prediction of outcome after surgical correction. *Circulation, 80*(4), 840-845.

Cohn, J. N., Ferrari, R., & Sharpe, N. (2000). Cardiac remodeling--concepts and clinical implications: a consensus paper from an international forum on cardiac remodeling. Behalf of an International Forum on Cardiac Remodeling. *J Am Coll Cardiol, 35*(3), 569-582.

Cook, J. P., Mahajan, A., & Morris, A. P. (2017). Guidance for the utility of linear models in meta-analysis of genetic association studies of binary phenotypes. *European journal of human genetics : EJHG, 25*(2), 240-245.

Cooper, G. M., Stone, E. A., Asimenos, G., Green, E. D., Batzoglou, S., & Sidow, A. (2005). Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res, 15*(7), 901-913.

Costa, M. W., Guo, G., Wolstein, O., Vale, M., Castro, M. L., Wang, L., Otway, R., Riek, P., Cochrane, N., Furtado, M., Semsarian, C., Weintraub, R. G., Yeoh, T., Hayward, C., Keogh, A., Macdonald, P., Feneley, M., Graham, R. M., Seidman, J. G., Seidman, C. E., Rosenthal, N., Fatkin, D., & Harvey, R. P. (2013). Functional characterization of a novel mutation in *NKX2-5* associated with congenital heart disease and adult-onset cardiomyopathy. *Circ Cardiovasc Genet, 6*(3), 238-247.

Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., Jupe, S., Kalatskaya, I., Mahajan, S., May, B., Ndegwa, N., Schmidt, E., Shamovsky, V., Yung, C., Birney, E., Hermjakob, H., D'Eustachio, P., & Stein, L. (2011). Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res, 39*(Database issue), D691-697.

Daccarett, M., Badger, T. J., Akoum, N., Burgon, N. S., Mahnkopf, C., Vergara, G., Kholmovski, E., McGann, C. J., Parker, D., Brachmann, J., Macleod, R. S., & Marrouche, N. F. (2011). Association of left atrial fibrosis detected by delayed-enhancement magnetic resonance imaging and the risk of stroke in patients with atrial fibrillation. *J Am Coll Cardiol, 57*(7), 831-838.

Darbar, D., Motsinger, A. A., Ritchie, M. D., Gainer, J. V., & Roden, D. M. (2007). Polymorphism modulates symptomatic response to antiarrhythmic drug therapy in patients with lone atrial fibrillation. *Heart Rhythm, 4*(6), 743-749.

Darbar, D., & Roden, D. M. (2013). Genetic mechanisms of atrial fibrillation: impact on response to treatment. *Nat Rev Cardiol, 10*(6), 317-329.

de Vlaming, R., Okbay, A., Rietveld, C. A., Johannesson, M., Magnusson, P. K., Uitterlinden, A. G., van Rooij, F. J., Hofman, A., Groenen, P. J., Thurik, A. R., & Koellinger, P. D. (2017). Meta-GWAS Accuracy and Power (MetaGAP) Calculator Shows that Hiding Heritability Is Partially Due to Imperfect Genetic Correlations across Studies. *PLoS Genet, 13*(1), e1006495.

de Vos, C. B., Pisters, R., Nieuwlaat, R., Prins, M. H., Tieleman, R. G., Coelen, R. J., van den Heijkant, A. C., Allessie, M. A., & Crijns, H. J. (2010). Progression from paroxysmal to persistent atrial fibrillation clinical correlates and prognosis. *J Am Coll Cardiol, 55*(8), 725-731.

Delaney, J. T., Jeff, J. M., Brown, N. J., Pretorius, M., Okafor, H. E., Darbar, D., Roden, D. M., & Crawford, D. C. (2012). Characterization of genome-wide association-identified variants for atrial fibrillation in African Americans. *PLoS One, 7*(2), e32338.

DeLuca, D. S., Levin, J. Z., Sivachenko, A., Fennell, T., Nazaire, M.-D., Williams, C., Reich, M., Winckler, W., & Getz, G. (2012). RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics, 28*(11), 1530-1532.

den Hoed, M., Eijgelsheim, M., Esko, T., Brundel, B. J., Peal, D. S., Evans, D. M., Nolte, I. M., Segre, A. V., Holm, H., Handsaker, R. E., Westra, H. J., Johnson, T., Isaacs, A., Yang, J., Lundby, A., Zhao, J. H., Kim, Y. J., Go, M. J., Almgren, P., Bochud, M., Boucher, G., Cornelis, M. C., Gudbjartsson, D., Hadley, D., van der Harst, P., Hayward, C., den Heijer, M., Igl, W., Jackson, A. U., Kutalik, Z., Luan, J., Kemp, J. P., Kristiansson, K., Ladenvall, C., Lorentzon, M., Montasser, M. E., Njajou, O. T., O'Reilly, P. F., Padmanabhan, S., St Pourcain, B., Rankinen, T., Salo, P., Tanaka, T., Timpson, N. J., Vitart, V., Waite, L., Wheeler, W., Zhang, W., Draisma, H. H., Feitosa, M. F., Kerr, K. F., Lind, P. A., Mihailov, E., Onland-Moret, N. C., Song, C., Weedon, M. N., Xie, W., Yengo, L., Absher, D., Albert, C. M., Alonso, A., Arking, D. E., de Bakker, P. I., Balkau, B., Barlassina, C., Benaglio, P., Bis, J. C., Bouatia-Naji, N., Brage, S., Chanock, S. J., Chines, P. S., Chung, M., Darbar, D., Dina, C., Dorr, M., Elliott, P., Felix, S. B., Fischer, K., Fuchsberger, C., de Geus, E. J., Goyette, P., Gudnason, V., Harris, T. B., Hartikainen, A. L., Havulinna, A. S., Heckbert, S. R., Hicks, A. A., Hofman, A., Holewijn, S., Hoogstra-Berends, F., Hottenga, J. J., Jensen, M. K., Johansson, A., Junttila, J., Kaab, S., Kanon, B., Ketkar, S., Khaw, K. T., Knowles, J. W., Kooner, A. S., Kors, J. A., Kumari, M., Milani, L., Laiho, P., Lakatta, E. G., Langenberg, C., Leusink, M., Liu, Y., Luben, R. N., Lunetta, K. L., Lynch, S. N., Markus, M. R., Marques-Vidal, P., Mateo Leach, I., McArdle, W. L., McCarroll, S. A., Medland, S. E., Miller, K. A., Montgomery, G. W., Morrison, A. C., Muller-Nurasyid, M., Navarro, P., Nelis, M., O'Connell, J. R., O'Donnell, C. J., Ong, K. K., Newman, A. B., Peters, A., Polasek, O., Pouta, A., Pramstaller, P. P., Psaty, B. M., Rao, D. C., Ring, S. M., Rossin, E. J., Rudan, D., Sanna, S., Scott, R. A., Sehmi, J. S., Sharp, S., Shin, J. T.,

Singleton, A. B., Smith, A. V., Soranzo, N., Spector, T. D., Stewart, C., Stringham, H. M., Tarasov, K. V., Uitterlinden, A. G., Vandenput, L., Hwang, S. J., Whitfield, J. B., Wijmenga, C., Wild, S. H., Willemsen, G., Wilson, J. F., Witteman, J. C., Wong, A., Wong, Q., Jamshidi, Y., Zitting, P., Boer, J. M., Boomsma, D. I., Borecki, I. B., van Duijn, C. M., Ekelund, U., Forouhi, N. G., Froguel, P., Hingorani, A., Ingelsson, E., Kivimaki, M., Kronmal, R. A., Kuh, D., Lind, L., Martin, N. G., Oostra, B. A., Pedersen, N. L., Quertermous, T., Rotter, J. I., van der Schouw, Y. T., Verschuren, W. M., Walker, M., Albanes, D., Arnar, D. O., Assimes, T. L., Bandinelli, S., Boehnke, M., de Boer, R. A., Bouchard, C., Caulfield, W. L., Chambers, J. C., Curhan, G., Cusi, D., Eriksson, J., Ferrucci, L., van Gilst, W. H., Glorioso, N., de Graaf, J., Groop, L., Gyllensten, U., Hsueh, W. C., Hu, F. B., Huikuri, H. V., Hunter, D. J., Iribarren, C., Isomaa, B., Jarvelin, M. R., Jula, A., Kahonen, M., Kiemeney, L. A., van der Klauw, M. M., Kooner, J. S., Kraft, P., Iacoviello, L., Lehtimaki, T., Lokki, M. L., Mitchell, B. D., Navis, G., Nieminen, M. S., Ohlsson, C., Poulter, N. R., Qi, L., Raitakari, O. T., Rimm, E. B., Rioux, J. D., Rizzi, F., Rudan, I., Salomaa, V., Sever, P. S., Shields, D. C., Shuldiner, A. R., Sinisalo, J., Stanton, A. V., Stolk, R. P., Strachan, D. P., Tardif, J. C., Thorsteinsdottir, U., Tuomilehto, J., van Veldhuisen, D. J., Virtamo, J., Viikari, J., Vollenweider, P., Waeber, G., Widen, E., Cho, Y. S., Olsen, J. V., Visscher, P. M., Willer, C., Franke, L., Erdmann, J., Thompson, J. R., Pfeufer, A., Sotoodehnia, N., Newton-Cheh, C., Ellinor, P. T., Stricker, B. H., Metspalu, A., Perola, M., Beckmann, J. S., Smith, G. D., Stefansson, K., Wareham, N. J., Munroe, P. B., Sibon, O. C., Milan, D. J., Snieder, H., Samani, N. J., & Loos, R. J. (2013). Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. *Nat Genet, 45*(6), 621-631.

Denny, J. C., Bastarache, L., Ritchie, M. D., Carroll, R. J., Zink, R., Mosley, J. D., Field, J. R., Pulley, J. M., Ramirez, A. H., Bowton, E., Basford, M. A., Carrell, D. S., Peissig, P. L., Kho, A. N., Pacheco, J. A., Rasmussen, L. V., Crosslin, D. R., Crane, P. K., Pathak, J., Bielinski, S. J., Pendergrass, S. A., Xu, H., Hindorff, L. A., Li, R., Manolio, T. A., Chute, C. G., Chisholm, R. L., Larson, E. B., Jarvik, G. P., Brilliant, M. H., McCarty, C. A., Kullo, I. J., Haines, J. L., Crawford, D. C., Masys, D. R., & Roden, D. M. (2013). Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol, 31*(12), 1102-1110.

Desai, A. D., Yaw, T. S., Yamazaki, T., Kaykha, A., Chun, S., & Froelicher, V. F. (2006). Prognostic Significance of Quantitative QRS Duration. *Am J Med, 119*(7), 600-606.

Digilio, M. C., Gnazzo, M., Lepri, F., Dentici, M. L., Pisaneschi, E., Baban, A., Passarelli, C., Capolino, R., Angioni, A., Novelli, A., Marino, B., & Dallapiccola, B. (2017). Congenital heart defects in molecularly proven Kabuki syndrome patients. *Am J Med Genet A, 173*(11), 2912-2922.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics, 29*(1), 15-21.

Dobrev, D., & Nattel, S. (2010). New antiarrhythmic drugs for treatment of atrial fibrillation. *Lancet, 375*(9721), 1212-1223.

Eger, A., Stockinger, A., Wiche, G., & Foisner, R. (1997). Polarisation-dependent association of plectin with desmoplakin and the lateral submembrane skeleton in MDCK cells. *J Cell Sci, 110 (Pt 11)*, 1307-1316.

Eijgelsheim, M., Newton-Cheh, C., Sotoodehnia, N., de Bakker, P. I., Muller, M., Morrison, A. C., Smith, A. V., Isaacs, A., Sanna, S., Dorr, M., Navarro, P., Fuchsberger, C., Nolte, I. M., de Geus, E. J., Estrada, K., Hwang, S. J., Bis, J. C., Ruckert, I. M., Alonso, A., Launer, L. J., Hottenga, J. J., Rivadeneira, F., Noseworthy, P. A., Rice, K. M., Perz, S., Arking, D. E., Spector, T. D., Kors, J. A., Aulchenko, Y. S., Tarasov, K. V., Homuth, G., Wild, S. H., Marroni, F., Gieger, C., Licht, C. M., Prineas, R. J., Hofman, A., Rotter, J. I., Hicks, A. A., Ernst, F., Najjar, S. S., Wright, A. F., Peters, A., Fox, E. R., Oostra, B. A., Kroemer, H. K., Couper, D., Volzke, H., Campbell, H., Meitinger, T., Uda, M., Witteman, J. C., Psaty, B. M., Wichmann, H. E., Harris, T. B., Kaab, S., Siscovick, D. S., Jamshidi, Y., Uitterlinden, A. G., Folsom, A. R., Larson, M. G., Wilson, J. F., Penninx, B. W., Snieder, H., Pramstaller, P. P., van Duijn, C. M., Lakatta, E. G., Felix, S. B., Gudnason, V., Pfeufer, A., Heckbert, S. R., Stricker, B. H., Boerwinkle, E., & O'Donnell, C. J. (2010). Genome-wide association analysis identifies multiple loci related to resting heart rate. *Hum Mol Genet, 19*(19), 3885-3894.

El Harchi, A., Zhang, H., & Hancox, J. C. (2010). The S140G *KCNQ1* atrial fibrillation mutation affects 'I(KS)' profile during both atrial and ventricular action potentials. *J Physiol Pharmacol, 61*(6), 759-764.

Ellinor, P. T., Lunetta, K. L., Albert, C. M., Glazer, N. L., Ritchie, M. D., Smith, A. V., Arking, D. E., Muller-Nurasyid, M., Krijthe, B. P., Lubitz, S. A., Bis, J. C., Chung, M. K., Dorr, M., Ozaki, K., Roberts, J. D., Smith, J. G., Pfeufer, A., Sinner, M. F., Lohman, K., Ding, J., Smith, N. L., Smith, J. D., Rienstra, M., Rice, K. M., Van Wagoner, D. R., Magnani, J. W., Wakili, R., Clauss, S., Rotter, J. I., Steinbeck, G., Launer, L. J., Davies, R. W., Borkovich, M., Harris, T. B., Lin, H., Volker, U., Volzke, H., Milan, D. J., Hofman, A., Boerwinkle, E., Chen, L. Y., Soliman, E. Z., Voight, B. F., Li, G., Chakravarti, A., Kubo, M., Tedrow, U. B., Rose, L. M., Ridker, P. M., Conen, D., Tsunoda, T., Furukawa, T., Sotoodehnia, N., Xu, S., Kamatani, N., Levy, D., Nakamura, Y., Parvez, B., Mahida, S., Furie, K. L., Rosand, J., Muhammad, R., Psaty, B. M., Meitinger, T., Perz, S., Wichmann, H. E., Witteman, J. C., Kao, W. H., Kathiresan, S., Roden, D. M., Uitterlinden, A. G., Rivadeneira, F., McKnight, B., Sjogren, M., Newman, A. B., Liu, Y., Gollob, M. H., Melander, O., Tanaka, T., Stricker, B. H., Felix, S. B., Alonso, A., Darbar, D., Barnard, J., Chasman, D. I., Heckbert, S. R., Benjamin, E. J., Gudnason, V., & Kaab, S. (2012). Meta-analysis identifies six new susceptibility loci for atrial fibrillation. *Nat Genet, 44*(6), 670-675.

Ellinor, P. T., Lunetta, K. L., Glazer, N. L., Pfeufer, A., Alonso, A., Chung, M. K., Sinner, M. F., de Bakker, P. I., Mueller, M., Lubitz, S. A., Fox, E., Darbar, D., Smith, N. L., Smith, J. D., Schnabel, R. B., Soliman, E. Z., Rice, K. M., Van Wagoner, D. R., Beckmann, B. M., van Noord, C., Wang, K., Ehret, G. B., Rotter, J. I., Hazen, S. L., Steinbeck, G., Smith, A. V., Launer, L. J., Harris, T. B., Makino, S., Nelis, M., Milan, D. J., Perz, S., Esko, T., Kottgen, A., Moebus, S., Newton-Cheh, C., Li, M., Mohlenkamp, S., Wang, T. J., Kao, W. H., Vasan, R. S., Nothen, M. M., MacRae, C. A., Stricker, B. H., Hofman, A., Uitterlinden, A. G., Levy, D., Boerwinkle, E., Metspalu, A., Topol, E. J., Chakravarti, A., Gudnason, V., Psaty, B. M., Roden, D. M., Meitinger, T., Wichmann, H. E., Witteman, J. C., Barnard, J., Arking, D. E., Benjamin, E. J., Heckbert, S. R., & Kaab, S. (2010). Common variants in *KCNN3* are associated with lone atrial fibrillation. *Nat Genet, 42*(3), 240-244.

Ellinor, P. T., Shin, J. T., Moore, R. K., Yoerger, D. M., & MacRae, C. A. (2003). Locus for atrial fibrillation maps to chromosome 6q14-16. *Circulation, 107*(23), 2880-2883.

Ellinor, P. T., Yoerger, D. M., Ruskin, J. N., & MacRae, C. A. (2005). Familial aggregation in lone atrial fibrillation. *Hum Genet, 118*(2), 179-184.

Elliott, P., & Peakman, T. C. (2008). The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int J Epidemiol, 37*(2), 234-244.

Elliott, P. M., Anastasakis, A., Borger, M. A., Borggrefe, M., Cecchi, F., Charron, P., Hagege, A. A., Lafont, A., Limongelli, G., Mahrholdt, H., McKenna, W. J., Mogensen, J., Nihoyannopoulos, P., Nistri, S., Pieper, P. G., Pieske, B., Rapezzi, C., Rutten, F. H., Tillmanns, C., & Watkins, H. (2014). 2014 ESC Guidelines on diagnosis and management of hypertrophic cardiomyopathy: the Task Force for the Diagnosis and Management of Hypertrophic Cardiomyopathy of the European Society of Cardiology (ESC). *Eur Heart J, 35*(39), 2733-2779.

Elseed, A. M., Shinebourne, E. A., & Paneth, M. (1974). Manifestation of juxtaductal coarctation after surgical ligation of persistent ductus arteriosus in infancy. *Br Heart J, 36*(7), 687-692.

Elvan, A., Wylie, K., & Zipes, D. P. (1996). Pacing-induced chronic atrial fibrillation impairs sinus node function in dogs. Electrophysiological remodeling. *Circulation, 94*(11), 2953-2960.

Engelfriet, P., & Mulder, B. J. (2009). Gender differences in adult congenital heart disease. *Neth Heart J, 17*(11), 414-417.

Fehrmann, R. S., Karjalainen, J. M., Krajewska, M., Westra, H. J., Maloney, D., Simeonov, A., Pers, T. H., Hirschhorn, J. N., Jansen, R. C., Schultes, E. A., van Haagen, H. H., de Vries, E. G., te Meerman, G. J., Wijmenga, C., van Vugt, M. A., & Franke, L. (2015). Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat Genet, 47*(2), 115-125.

Ferrer, M. I. (1968). The sick sinus syndrome in atrial disease. *Jama, 206*(3), 645-646.

Forbes, M. S., & Sperelakis, N. (1985). Intercalated discs of mammalian heart: a review of structure and function. *Tissue Cell, 17*(5), 605-648.

Fox, C. S., Parise, H., D'Agostino, R. B., Sr., Lloyd-Jones, D. M., Vasan, R. S., Wang, T. J., Levy, D., Wolf, P. A., & Benjamin, E. J. (2004). Parental atrial fibrillation as a risk factor for atrial fibrillation in offspring. *Jama, 291*(23), 2851-2855.

Franco, D., Lamers, W. H., & Moorman, A. F. (1998). Patterns of expression in the developing myocardium: towards a morphologically integrated transcriptional model. *Cardiovasc Res, 38*(1), 25-53.

Frank, D., Rangrez, A. Y., Poyanmehr, R., Seeger, T. S., Kuhn, C., Eden, M., Stiebeling, K., Bernt, A., Grund, C., Franke, W. W., & Frey, N. (2014). Mice with cardiac-restricted overexpression of Myozap are sensitized to biomechanical stress and develop a protein-aggregate-associated cardiomyopathy. *J Mol Cell Cardiol, 72*, 196-207.

Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M., Pasternak, S., Wheeler, D. A., Willis, T. D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhao, H., Zhou, J., Gabriel, S. B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R. C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Waye, M. M., Tsui, S. K., Xue, H., Wong, J. T., Galver, L. M., Fan, J. B., Gunderson, K., Murray, S. S., Oliphant, A. R., Chee, M. S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J. F., Phillips, M. S., Roumy, S., Sallee, C., Verner, A., Hudson, T. J., Kwok, P. Y., Cai, D., Koboldt, D. C., Miller, R. D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L. C., Mak, W., Song, Y. Q., Tam, P. K., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C. P., Delgado, M., Dermitzakis, E. T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B. E., Whittaker, P., Bentley, D. R., Daly, M. J., de Bakker, P. I., Barrett, J., Chretien, Y. R., Maller, J., McCarroll, S., Patterson, N., Pe'er, I., Price, A., Purcell, S., Richter, D. J., Sabeti, P., Saxena, R., Schaffner, S. F., Sham, P. C., Varilly, P., Altshuler, D., Stein, L. D., Krishnan, L., Smith, A. V., Tello-Ruiz, M. K., Thorisson, G. A., Chakravarti, A., Chen, P. E., Cutler, D. J., Kashuk, C. S., Lin, S., Abecasis, G. R., Guan, W., Li, Y., Munro, H. M., Qin, Z. S., Thomas, D. J., McVean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L. R., Clarke, G., Evans, D. M., Morris, A. P., Weir, B. S., Tsunoda, T., Mullikin, J. C., Sherry, S. T., Feolo, M., Skol, A., Zhang, H., Zeng, C., Zhao, H., Matsuda, I., Fukushima, Y., Macer, D. R., Suda, E., Rotimi, C. N., Adebamowo, C. A., Ajayi, I., Aniagwu, T., Marshall, P. A., Nkwodimmah, C., Royal, C. D., Leppert, M. F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I. F., Knoppers, B. M., Foster, M. W., Clayton, E. W., Watkin, J., Gibbs, R. A., Belmont, J. W., Muzny, D., Nazareth, L.,

Sodergren, E., Weinstock, G. M., Wheeler, D. A., Yakub, I., Gabriel, S. B., Onofrio, R. C., Richter, D. J., Ziaugra, L., Birren, B. W., Daly, M. J., Altshuler, D., Wilson, R. K., Fulton, L. L., Rogers, J., Burton, J., Carter, N. P., Clee, C. M., Griffiths, M., Jones, M. C., McLay, K., Plumb, R. W., Ross, M. T., Sims, S. K., Willey, D. L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J. C., L'Archeveque, P., Bellemare, G., Saeki, K., Wang, H., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A. L., Brooks, L. D., McEwen, J. E., Guyer, M. S., Wang, V. O., Peterson, J. L., Shi, M., Spiegel, J., Sung, L. M., Zacharia, L. F., Collins, F. S., Kennedy, K., Jamieson, R., & Stewart, J. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature, 449*(7164), 851-861.

Freedman, B., Potpara, T. S., & Lip, G. Y. (2016). Stroke prevention in atrial fibrillation. *Lancet, 388*(10046), 806-817.

Freeman, J. L., Perry, G. H., Feuk, L., Redon, R., McCarroll, S. A., Altshuler, D. M., Aburatani, H., Jones, K. W., Tyler-Smith, C., Hurles, M. E., Carter, N. P., Scherer, S. W., & Lee, C. (2006). Copy number variation: new insights in genome diversity. *Genome Res, 16*(8), 949-961.

Frey, N., Frank, D., Lippl, S., Kuhn, C., Kogler, H., Barrientos, T., Rohr, C., Will, R., Muller, O. J., Weiler, H., Bassel-Duby, R., Katus, H. A., & Olson, E. N. (2008). Calsarcin-2 deficiency increases exercise capacity in mice through calcineurin/NFAT activation. *J Clin Invest, 118*(11), 3598-3608.

Frey, N., & Olson, E. N. (2003). Cardiac hypertrophy: the good, the bad, and the ugly. *Annu Rev Physiol, 65*, 45-79.

Freylikhman, O., Tatarinova, T., Smolina, N., Zhuk, S., Klyushina, A., Kiselev, A., Moiseeva, O., Sjoberg, G., Malashicheva, A., & Kostareva, A. (2014). Variants in the *NOTCH1* gene in patients with aortic coarctation. *Congenit Heart Dis, 9*(5), 391-396.

Fuster, V., Harrington, R. A., Narula, J., & Eapen, Z. J. (2017). *Hurst's the Heart* (14th ed.). New York City: McGraw-Hill Education.

Fuster, V., Ryden, L. E., Asinger, R. W., Cannom, D. S., Crijns, H. J., Frye, R. L., Halperin, J. L., Kay, G. N., Klein, W. W., Levy, S., McNamara, R. L., Prystowsky, E. N., Wann, L. S., Wyse, D. G., Gibbons, R. J., Antman, E. M., Alpert, J. S., Faxon, D. P., Fuster, V., Gregoratos, G., Hiratzka, L. F., Jacobs, A. K., Russell, R. O., Smith, S. C., Jr., Klein, W. W., Alonso-Garcia, A., Blomstrom-Lundqvist, C., de Backer, G., Flather, M., Hradec, J., Oto, A., Parkhomenko, A., Silber, S., & Torbicki, A. (2001). ACC/AHA/ESC Guidelines for the Management of Patients With

Atrial Fibrillation: Executive Summary A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice Guidelines and Policy Conferences (Committee to Develop Guidelines for the Management of Patients With Atrial Fibrillation) Developed in Collaboration With the North American Society of Pacing and Electrophysiology. *Circulation, 104*(17), 2118-2150.

Ge, T., Chen, C. Y., Neale, B. M., Sabuncu, M. R., & Smoller, J. W. (2017). Phenome-wide heritability analysis of the UK Biobank. *PLoS Genet, 13*(4), e1006711.

Gerull, B., Heuser, A., Wichter, T., Paul, M., Basson, C. T., McDermott, D. A., Lerman, B. B., Markowitz, S. M., Ellinor, P. T., MacRae, C. A., Peters, S., Grossmann, K. S., Drenckhahn, J., Michely, B., Sasse-Klaassen, S., Birchmeier, W., Dietz, R., Breithardt, G., Schulze-Bahr, E., & Thierfelder, L. (2004). Mutations in the desmosomal protein plakophilin-2 are common in arrhythmogenic right ventricular cardiomyopathy. *Nat Genet, 36*(11), 1162-1164.

Go, A. S., Hylek, E. M., Phillips, K. A., Chang, Y., Henault, L. E., Selby, J. V., & Singer, D. E. (2001). Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study. *Jama, 285*(18), 2370-2375.

Goette, A., Kalman, J. M., Aguinaga, L., Akar, J., Cabrera, J. A., Chen, S. A., Chugh, S. S., Corradi, D., D'Avila, A., Dobrev, D., Fenelon, G., Gonzalez, M., Hatem, S. N., Helm, R., Hindricks, G., Ho, S. Y., Hoit, B., Jalife, J., Kim, Y. H., Lip, G. Y., Ma, C. S., Marcus, G. M., Murray, K., Nogami, A., Sanders, P., Uribe, W., Van Wagoner, D. R., & Nattel, S. (2016). EHRA/HRS/APHRS/SOLAECE expert consensus on atrial cardiomyopathies: definition, characterization, and clinical implication. *Europace, 18*(10), 1455-1490.

Goldberger, J. J., Arora, R., Green, D., Greenland, P., Lee, D. C., Lloyd-Jones, D. M., Markl, M., Ng, J., & Shah, S. J. (2015). Evaluating the Atrial Myopathy Underlying Atrial Fibrillation: Identifying the Arrhythmogenic and Thrombogenic Substrate. *Circulation, 132*(4), 278-291.

Goossens, S., Janssens, B., Bonne, S., De Rycke, R., Braet, F., van Hengel, J., & van Roy, F. (2007). A unique and specific interaction between alphaT-catenin and plakophilin-2 in the area composita, the mixed-

type junctional structure of cardiac intercalated discs. *J Cell Sci, 120*(Pt 12), 2126-2136.

Gottesman, II, & Shields, J. (1967). A polygenic theory of schizophrenia. *Proc Natl Acad Sci U S A, 58*(1), 199-205.

Graves, K. G., May, H. T., Knowlton, K. U., Muhlestein, J. B., Jacobs, V., Lappe, D. L., Anderson, J. L., Horne, B. D., & Bunch, T. J. (2018). Improving CHA2DS2-VASc stratification of non-fatal stroke and mortality risk using the Intermountain Mortality Risk Score among patients with atrial fibrillation. *Open Heart, 5*(2), e000907.

Gregg, R., Helfenbein, E., Shao, M., & Zhou, S. (2005). *Global QT measurements in the Philips 12-lead Algorithm* (Vol. 38).

Gretarsdottir, S., Thorleifsson, G., Manolescu, A., Styrkarsdottir, U., Helgadottir, A., Gschwendtner, A., Kostulas, K., Kuhlenbaumer, G., Bevan, S., Jonsdottir, T., Bjarnason, H., Saemundsdottir, J., Palsson, S., Arnar, D. O., Holm, H., Thorgeirsson, G., Valdimarsson, E. M., Sveinbjornsdottir, S., Gieger, C., Berger, K., Wichmann, H. E., Hillert, J., Markus, H., Gulcher, J. R., Ringelstein, E. B., Kong, A., Dichgans, M., Gudbjartsson, D. F., Thorsteinsdottir, U., & Stefansson, K. (2008). Risk variants for atrial fibrillation on chromosome 4q25 associate with ischemic stroke. *Ann Neurol, 64*(4), 402-409.

Gudbjartsson, D. F., Arnar, D. O., Helgadottir, A., Gretarsdottir, S., Holm, H., Sigurdsson, A., Jonasdottir, A., Baker, A., Thorleifsson, G., Kristjansson, K., Palsson, A., Blondal, T., Sulem, P., Backman, V. M., Hardarson, G. A., Palsdottir, E., Helgason, A., Sigurjonsdottir, R., Sverrisson, J. T., Kostulas, K., Ng, M. C., Baum, L., So, W. Y., Wong, K. S., Chan, J. C., Furie, K. L., Greenberg, S. M., Sale, M., Kelly, P., MacRae, C. A., Smith, E. E., Rosand, J., Hillert, J., Ma, R. C., Ellinor, P. T., Thorgeirsson, G., Gulcher, J. R., Kong, A., Thorsteinsdottir, U., & Stefansson, K. (2007). Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature, 448*(7151), 353-357.

Gudbjartsson, D. F., Helgason, H., Gudjonsson, S. A., Zink, F., Oddson, A., Gylfason, A., Besenbacher, S., Magnusson, G., Halldorsson, B. V., Hjartarson, E., Sigurdsson, G. T., Stacey, S. N., Frigge, M. L., Holm, H., Saemundsdottir, J., Helgadottir, H. T., Johannsdottir, H., Sigfusson, G., Thorgeirsson, G., Sverrisson, J. T., Gretarsdottir, S., Walters, G. B., Rafnar, T., Thjodleifsson, B., Bjornsson, E. S., Olafsson, S., Thorarinsdottir, H., Steingrimsdottir, T., Gudmundsdottir, T. S., Theodors, A., Jonasson, J. G., Sigurdsson, A., Bjornsdottir, G., Jonsson, J. J., Thorarensen, O., Ludvigsson, P., Gudbjartsson, H., Eyjolfsson, G. I., Sigurdardottir, O., Olafsson, I., Arnar, D. O.,

Magnusson, O. T., Kong, A., Masson, G., Thorsteinsdottir, U., Helgason, A., Sulem, P., & Stefansson, K. (2015). Large-scale whole-genome sequencing of the Icelandic population. *Nat Genet, 47*(5), 435-444.

Gudbjartsson, D. F., Holm, H., Gretarsdottir, S., Thorleifsson, G., Walters, G. B., Thorgeirsson, G., Gulcher, J., Mathiesen, E. B., Njolstad, I., Nyrnes, A., Wilsgaard, T., Hald, E. M., Hveem, K., Stoltenberg, C., Kucera, G., Stubblefield, T., Carter, S., Roden, D., Ng, M. C., Baum, L., So, W. Y., Wong, K. S., Chan, J. C., Gieger, C., Wichmann, H. E., Gschwendtner, A., Dichgans, M., Kuhlenbaumer, G., Berger, K., Ringelstein, E. B., Bevan, S., Markus, H. S., Kostulas, K., Hillert, J., Sveinbjornsdottir, S., Valdimarsson, E. M., Lochen, M. L., Ma, R. C., Darbar, D., Kong, A., Arnar, D. O., Thorsteinsdottir, U., & Stefansson, K. (2009). A sequence variant in *ZFHX3* on 16q22 associates with atrial fibrillation and ischemic stroke. *Nat Genet, 41*(8), 876-878.

Gudbjartsson, D. F., Holm, H., Sulem, P., Masson, G., Oddsson, A., Magnusson, O. T., Saemundsdottir, J., Helgadottir, H. T., Helgason, H., Johannsdottir, H., Gretarsdottir, S., Gudjonsson, S. A., Njolstad, I., Lochen, M. L., Baum, L., Ma, R. C., Sigfusson, G., Kong, A., Thorgeirsson, G., Sverrisson, J. T., Thorsteinsdottir, U., Stefansson, K., & Arnar, D. O. (2017). A frameshift deletion in the sarcomere gene *MYL4* causes early-onset familial atrial fibrillation. *Eur Heart J, 38*(1), 27-34.

Gundesli, H., Talim, B., Korkusuz, P., Balci-Hayta, B., Cirak, S., Akarsu, N. A., Topaloglu, H., & Dincer, P. (2010). Mutation in exon 1f of *PLEC*, leading to disruption of plectin isoform 1f, causes autosomal-recessive limb-girdle muscular dystrophy. *Am J Hum Genet, 87*(6), 834-841.

Gussow, A. B., Petrovski, S., Wang, Q., Allen, A. S., & Goldstein, D. B. (2016). The intolerance to functional genetic variation of protein domains predicts the localization of pathogenic mutations within genes. *Genome Biol, 17*, 9.

Haissaguerre, M., Jais, P., Shah, D. C., Takahashi, A., Hocini, M., Quiniou, G., Garrigue, S., Le Mouroux, A., Le Metayer, P., & Clementy, J. (1998). Spontaneous initiation of atrial fibrillation by ectopic beats originating in the pulmonary veins. *N Engl J Med, 339*(10), 659-666.

Hall, M. H., & Smoller, J. W. (2010). A new role for endophenotypes in the GWAS era: functional characterization of risk variants. *Harv Rev Psychiatry, 18*(1), 67-74.

Hamdan, M. A., Maheshwari, S., Fahey, J. T., & Hellenbrand, W. E. (2001). Endovascular stents for coarctation of the aorta: initial results and intermediate-term follow-up. *J Am Coll Cardiol, 38*(5), 1518-1523.

Hammwohner, M., Bukowska, A., Mahardika, W., & Goette, A. (2019). Clinical importance of atrial cardiomyopathy. *Int J Cardiol, 287*, 174-180.

Hang, C. T., Yang, J., Han, P., Cheng, H. L., Shang, C., Ashley, E., Zhou, B., & Chang, C. P. (2010). Chromatin regulation by Brg1 underlies heart muscle development and disease. *Nature, 466*(7302), 62-67.

Harada, M., Van Wagoner, D. R., & Nattel, S. (2015). Role of inflammation in atrial fibrillation pathophysiology and management. *Circ J, 79*(3), 495-502.

Heikkinen, L. O., Ala-Kulju, K. V., & Salo, J. A. (1991). Dilatation of ascending aorta in patients with repaired coarctation. *Scand J Thorac Cardiovasc Surg, 25*(1), 25-28.

Helgadottir, A., Gretarsdottir, S., Thorleifsson, G., Hjartarson, E., Sigurdsson, A., Magnusdottir, A., Jonasdottir, A., Kristjansson, H., Sulem, P., Oddsson, A., Sveinbjornsson, G., Steinthorsdottir, V., Rafnar, T., Masson, G., Jonsdottir, I., Olafsson, I., Eyjolfsson, G. I., Sigurdardottir, O., Daneshpour, M. S., Khalili, D., Azizi, F., Swinkels, D. W., Kiemeney, L., Quyyumi, A. A., Levey, A. I., Patel, R. S., Hayek, S. S., Gudmundsdottir, I. J., Thorgeirsson, G., Thorsteinsdottir, U., Gudbjartsson, D. F., Holm, H., & Stefansson, K. (2016). Variants with large effects on blood lipids and the role of cholesterol and triglycerides in coronary disease. *Nat Genet, 48*(6), 634-639.

Herman, D. S., Lam, L., Taylor, M. R., Wang, L., Teekakirikul, P., Christodoulou, D., Conner, L., DePalma, S. R., McDonough, B., Sparks, E., Teodorescu, D. L., Cirino, A. L., Banner, N. R., Pennell, D. J., Graw, S., Merlo, M., Di Lenarda, A., Sinagra, G., Bos, J. M., Ackerman, M. J., Mitchell, R. N., Murry, C. E., Lakdawala, N. K., Ho, C. Y., Barton, P. J., Cook, S. A., Mestroni, L., Seidman, J. G., & Seidman, C. E. (2012). Truncations of titin causing dilated cardiomyopathy. *N Engl J Med, 366*(7), 619-628.

Herron, T. J., & McDonald, K. S. (2002). Small amounts of alpha-myosin heavy chain isoform expression significantly increase power output of rat cardiac myocyte fragments. *Circ Res, 90*(11), 1150-1152.

Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., & Manolio, T. A. (2009). Potential etiologic and

functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A, 106*(23), 9362-9367.

Hirschhorn, J. N., & Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet, 6*(2), 95-108.

Hirschhorn, J. N., & Gajdos, Z. K. (2011). Genome-wide association studies: results from the first few years and potential implications for clinical medicine. *Annu Rev Med, 62*, 11-24.

Ho, S. Y., & Anderson, R. H. (1979). Coarctation, tubular hypoplasia, and the ductus arteriosus. Histological study of 35 specimens. *Br Heart J, 41*(3), 268-274.

Ho, V. B., Bakalov, V. K., Cooley, M., Van, P. L., Hood, M. N., Burklow, T. R., & Bondy, C. A. (2004). Major vascular anomalies in Turner syndrome: prevalence and magnetic resonance angiographic features. *Circulation, 110*(12), 1694-1700.

Hoffman, J. I. (2018). The challenge in diagnosing coarctation of the aorta. *Cardiovasc J Afr, 29*(4), 252-255.

Hoffman, J. I., & Kaplan, S. (2002). The incidence of congenital heart disease. *J Am Coll Cardiol, 39*(12), 1890-1900.

Hoffmann, T. J., Ehret, G. B., Nandakumar, P., Ranatunga, D., Schaefer, C., Kwok, P. Y., Iribarren, C., Chakravarti, A., & Risch, N. (2017). Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat Genet, 49*(1), 54-64.

Holm, H., Gudbjartsson, D. F., Arnar, D. O., Thorleifsson, G., Thorgeirsson, G., Stefansdottir, H., Gudjonsson, S. A., Jonasdottir, A., Mathiesen, E. B., Njolstad, I., Nyrnes, A., Wilsgaard, T., Hald, E. M., Hveem, K., Stoltenberg, C., Lochen, M. L., Kong, A., Thorsteinsdottir, U., & Stefansson, K. (2010). Several common variants modulate heart rate, PR interval and QRS duration. *Nat Genet, 42*(2), 117-122.

Holm, H., Gudbjartsson, D. F., Sulem, P., Masson, G., Helgadottir, H. T., Zanon, C., Magnusson, O. T., Helgason, A., Saemundsdottir, J., Gylfason, A., Stefansdottir, H., Gretarsdottir, S., Matthiasson, S. E., Thorgeirsson, G. M., Jonasdottir, A., Sigurdsson, A., Stefansson, H., Werge, T., Rafnar, T., Kiemeney, L. A., Parvez, B., Muhammad, R., Roden, D. M., Darbar, D., Thorleifsson, G., Walters, G. B., Kong, A., Thorsteinsdottir, U., Arnar, D. O., & Stefansson, K. (2011). A rare variant in *MYH6* is associated with high risk of sick sinus syndrome. *Nat Genet, 43*(4), 316-320.

Houmsse, M., Tyler, J., & Kalbfleisch, S. (2011). Supraventricular tachycardia causing heart failure. *Curr Opin Cardiol, 26*(3), 261-269.

Huber, O. (2003). Structure and function of desmosomal proteins and their role in development and disease. *Cell Mol Life Sci, 60*(9), 1872-1890.

Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. (2015). *Science, 348*(6235), 648-660.

Huxley, R. R., Lopez, F. L., Folsom, A. R., Agarwal, S. K., Loehr, L. R., Soliman, E. Z., Maclehose, R., Konety, S., & Alonso, A. (2011). Absolute and attributable risks of atrial fibrillation in relation to optimal and borderline risk factors: the Atherosclerosis Risk in Communities (ARIC) study. *Circulation, 123*(14), 1501-1508.

Iafrate, A. J., Feuk, L., Rivera, M. N., Listewnik, M. L., Donahoe, P. K., Qi, Y., Scherer, S. W., & Lee, C. (2004). Detection of large-scale variation in the human genome. *Nat Genet, 36*(9), 949-951.

Ihida-Stansbury, K., McKean, D. M., Gebb, S. A., Martin, J. F., Stevens, T., Nemenoff, R., Akeson, A., Vaughn, J., & Jones, P. L. (2004). Paired-related homeobox gene Prx1 is required for pulmonary vascular development. *Circ Res, 94*(11), 1507-1514.

Inouye, M., Abraham, G., Nelson, C. P., Wood, A. M., Sweeting, M. J., Dudbridge, F., Lai, F. Y., Kaptoge, S., Brozynska, M., Wang, T., Ye, S., Webb, T. R., Rutter, M. K., Tzoulaki, I., Patel, R. S., Loos, R. J. F., Keavney, B., Hemingway, H., Thompson, J., Watkins, H., Deloukas, P., Di Angelantonio, E., Butterworth, A. S., Danesh, J., & Samani, N. J. (2018). Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. *J Am Coll Cardiol, 72*(16), 1883-1893.

Ishikawa, T., Jou, C. J., Nogami, A., Kowase, S., Arrington, C. B., Barnett, S. M., Harrell, D. T., Arimura, T., Tsuji, Y., Kimura, A., & Makita, N. (2015). Novel mutation in the alpha-myosin heavy chain gene is associated with sick sinus syndrome. *Circ Arrhythm Electrophysiol, 8*(2), 400-408.

Iwasaki, Y. K., Nishida, K., Kato, T., & Nattel, S. (2011). Atrial fibrillation pathophysiology: implications for management. *Circulation, 124*(20), 2264-2274.

Jacobsen, B. K., Eggen, A. E., Mathiesen, E. B., Wilsgaard, T., & Njolstad, I. (2012). Cohort profile: the Tromso Study. *Int J Epidemiol, 41*(4), 961-967.

January, C. T., Wann, L. S., Calkins, H., Chen, L. Y., Cigarroa, J. E., Cleveland, J. C., Jr., Ellinor, P. T., Ezekowitz, M. D., Field, M. E., Furie, K. L., Heidenreich, P. A., Murray, K. T., Shea, J. B., Tracy, C. M., & Yancy, C. W. (2019). 2019 AHA/ACC/HRS Focused Update of the 2014 AHA/ACC/HRS Guideline for the Management of Patients With Atrial Fibrillation: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society. *J Am Coll Cardiol*.

Jimenez-Sanchez, G., Childs, B., & Valle, D. (2001). Human disease genes. *Nature, 409*(6822), 853-855.

Jung, C. G., Kim, H. J., Kawaguchi, M., Khanna, K. K., Hida, H., Asai, K., Nishino, H., & Miura, Y. (2005). Homeotic factor ATBF1 induces the cell cycle arrest associated with neuronal differentiation. *Development, 132*(23), 5137-5145.

Kadish, A. H., Buxton, A. E., Kennedy, H. L., Knight, B. P., Mason, J. W., Schuger, C. D., Tracy, C. M., Winters, W. L., Jr., Boone, A. W., Elnicki, M., Hirshfeld, J. W., Jr., Lorell, B. H., Rodgers, G. P., Tracy, C. M., & Weitz, H. H. (2001). ACC/AHA clinical competence statement on electrocardiography and ambulatory electrocardiography: A report of the ACC/AHA/ACP-ASIM task force on clinical competence (ACC/AHA Committee to develop a clinical competence statement on electrocardiography and ambulatory electrocardiography) endorsed by the International Society for Holter and noninvasive electrocardiology. *Circulation, 104*(25), 3169-3178.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., & Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res, 40*(Database issue), D109-114.

Karczewski, K., C Francioli, L., Tiao, G., B Cummings, B., Alföldi, J., Wang, Q., Collins, R., M Laricchia, K., Ganna, A., P Birnbaum, D., D Gauthier, L., Brand, H., Solomonson, M., A Watts, N., Rhodes, D., Singer-Berk, M., Seaby, E., A Kosmicki, J., K Walters, R., & Macarthur, D. (2019). *Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes: Supplementary Information*.

Kathiresan, S., & Srivastava, D. (2012). Genetics of human cardiovascular disease. *Cell, 148*(6), 1242-1257.

Kaushal, S., Backer, C. L., Patel, J. N., Patel, S. K., Walker, B. L., Weigel, T. J., Randolph, G., Wax, D., & Mavroudis, C. (2009). Coarctation of the

aorta: midterm outcomes of resection with extended end-to-end anastomosis. *Ann Thorac Surg, 88*(6), 1932-1938.

Kehat, I., & Molkentin, J. D. (2010). Molecular pathways underlying cardiac remodeling during pathophysiological stimulation. *Circulation, 122*(25), 2727-2735.

Kenny, D., Polson, J. W., Martin, R. P., Paton, J. F., & Wolf, A. R. (2011). Hypertension and coarctation of the aorta: an inevitable consequence of developmental pathophysiology. *Hypertens Res, 34*(5), 543-547.

Kerstjens-Frederikse, W. S., van de Laar, I. M., Vos, Y. J., Verhagen, J. M., Berger, R. M., Lichtenbelt, K. D., Klein Wassink-Ruiter, J. S., van der Zwaag, P. A., du Marchie Sarvaas, G. J., Bergman, K. A., Bilardo, C. M., Roos-Hesselink, J. W., Janssen, J. H., Frohn-Mulder, I. M., van Spaendonck-Zwarts, K. Y., van Melle, J. P., Hofstra, R. M., & Wessels, M. W. (2016). Cardiovascular malformations caused by *NOTCH1* mutations do not keep left: data on 428 probands with left-sided CHD and their families. *Genet Med, 18*(9), 914-923.

Kezerashvili, A., Krumerman, A. K., & Fisher, J. D. (2008). Sinus Node Dysfunction in Atrial Fibrillation: Cause or Effect? *J Atr Fibrillation, 1*(3), 30.

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics, 46*(3), 310-315.

Kirchhof, P. (2017). The future of atrial fibrillation management: integrated care and stratified therapy. *Lancet, 390*(10105), 1873-1887.

Kirchhof, P., Benussi, S., Kotecha, D., Ahlsson, A., Atar, D., Casadei, B., Castella, M., Diener, H. C., Heidbuchel, H., Hendriks, J., Hindricks, G., Manolis, A. S., Oldgren, J., Popescu, B. A., Schotten, U., Van Putte, B., Vardas, P., Agewall, S., Camm, J., Baron Esquivias, G., Budts, W., Carerj, S., Casselman, F., Coca, A., De Caterina, R., Deftereos, S., Dobrev, D., Ferro, J. M., Filippatos, G., Fitzsimons, D., Gorenek, B., Guenoun, M., Hohnloser, S. H., Kolh, P., Lip, G. Y., Manolis, A., McMurray, J., Ponikowski, P., Rosenhek, R., Ruschitzka, F., Savelieva, I., Sharma, S., Suwalski, P., Tamargo, J. L., Taylor, C. J., Van Gelder, I. C., Voors, A. A., Windecker, S., Zamorano, J. L., & Zeppenfeld, K. (2016). 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *Europace, 18*(11), 1609-1678.

Koizumi, R., Funamoto, K., Hayase, T., Kanke, Y., Shibata, M., Shiraishi, Y., & Yambe, T. (2015). Numerical analysis of hemodynamic changes in the left atrium due to atrial fibrillation. *J Biomech, 48*(3), 472-478.

Kong, A., Masson, G., Frigge, M. L., Gylfason, A., Zusmanovich, P., Thorleifsson, G., Olason, P. I., Ingason, A., Steinberg, S., Rafnar, T., Sulem, P., Mouy, M., Jonsson, F., Thorsteinsdottir, U., Gudbjartsson, D. F., Stefansson, H., & Stefansson, K. (2008). Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat Genet, 40*(9), 1068-1075.

Kopan, R. (2012). Notch signaling. *Cold Spring Harb Perspect Biol, 4*(10).

Koss-Harnes, D., Hoyheim, B., Anton-Lamprecht, I., Gjesti, A., Jorgensen, R. S., Jahnsen, F. L., Olaisen, B., Wiche, G., & Gedde-Dahl, T., Jr. (2002). A site-specific plectin mutation causes dominant epidermolysis bullosa simplex Ogna: two identical de novo mutations. *J Invest Dermatol, 118*(1), 87-93.

Krenz, M., & Robbins, J. (2004). Impact of beta-myosin heavy chain expression on cardiac function during stress. *J Am Coll Cardiol, 44*(12), 2390-2397.

Krokstad, S., Langhammer, A., Hveem, K., Holmen, T. L., Midthjell, K., Stene, T. R., Bratberg, G., Heggland, J., & Holmen, J. (2013). Cohort Profile: the HUNT Study, Norway. *Int J Epidemiol, 42*(4), 968-977.

Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc, 4*(7), 1073-1081.

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M. J., Amin, V., Whitaker, J. W., Schultz, M. D., Ward, L. D., Sarkar, A., Quon, G., Sandstrom, R. S., Eaton, M. L., Wu, Y. C., Pfenning, A. R., Wang, X., Claussnitzer, M., Liu, Y., Coarfa, C., Harris, R. A., Shoresh, N., Epstein, C. B., Gjoneska, E., Leung, D., Xie, W., Hawkins, R. D., Lister, R., Hong, C., Gascard, P., Mungall, A. J., Moore, R., Chuah, E., Tam, A., Canfield, T. K., Hansen, R. S., Kaul, R., Sabo, P. J., Bansal, M. S., Carles, A., Dixon, J. R., Farh, K. H., Feizi, S., Karlic, R., Kim, A. R., Kulkarni, A., Li, D., Lowdon, R., Elliott, G., Mercer, T. R., Neph, S. J., Onuchic, V., Polak, P., Rajagopal, N., Ray, P., Sallari, R. C., Siebenthall, K. T., Sinnott-Armstrong, N. A., Stevens, M., Thurman, R. E., Wu, J., Zhang, B., Zhou, X., Beaudet, A. E., Boyer, L. A., De Jager, P. L., Farnham, P. J., Fisher, S. J., Haussler, D., Jones, S. J., Li, W., Marra, M. A., McManus, M. T., Sunyaev, S., Thomson, J. A., Tlsty, T.

D., Tsai, L. H., Wang, W., Waterland, R. A., Zhang, M. Q., Chadwick, L. H., Bernstein, B. E., Costello, J. F., Ecker, J. R., Hirst, M., Meissner, A., Milosavljevic, A., Ren, B., Stamatoyannopoulos, J. A., Wang, T., & Kellis, M. (2015). Integrative analysis of 111 reference human epigenomes. *Nature, 518*(7539), 317-330.

L DeLano, W. (2002). *The PyMOL Molecular Graphics System (2002) DeLano Scientific, Palo Alto, CA, USA. http://www.pymol.org*.

Lage, K., Karlberg, E. O., Storling, Z. M., Olason, P. I., Pedersen, A. G., Rigina, O., Hinsby, A. M., Tumer, Z., Pociot, F., Tommerup, N., Moreau, Y., & Brunak, S. (2007). A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat Biotechnol, 25*(3), 309-316.

Lannering, K., Bartos, M., & Mellander, M. (2015). Late Diagnosis of Coarctation Despite Prenatal Ultrasound and Postnatal Pulse Oximetry. *Pediatrics, 136*(2), e406-412.

Larsson, S. C., Drca, N., & Wolk, A. (2014). Alcohol consumption and risk of atrial fibrillation: a prospective study and dose-response meta-analysis. *J Am Coll Cardiol, 64*(3), 281-289.

Lawlor, D. A., Harbord, R. M., Sterne, J. A., Timpson, N., & Davey Smith, G. (2008). Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med, 27*(8), 1133-1163.

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., O'Donnell-Luria, A. H., Ware, J. S., Hill, A. J., Cummings, B. B., Tukiainen, T., Birnbaum, D. P., Kosmicki, J. A., Duncan, L. E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D. N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J., Gupta, N., Howrigan, D., Kiezun, A., Kurki, M. I., Moonshine, A. L., Natarajan, P., Orozco, L., Peloso, G. M., Poplin, R., Rivas, M. A., Ruano-Rubio, V., Rose, S. A., Ruderfer, D. M., Shakir, K., Stenson, P. D., Stevens, C., Thomas, B. P., Tiao, G., Tusie-Luna, M. T., Weisburd, B., Won, H. H., Yu, D., Altshuler, D. M., Ardissino, D., Boehnke, M., Danesh, J., Donnelly, S., Elosua, R., Florez, J. C., Gabriel, S. B., Getz, G., Glatt, S. J., Hultman, C. M., Kathiresan, S., Laakso, M., McCarroll, S., McCarthy, M. I., McGovern, D., McPherson, R., Neale, B. M., Palotie, A., Purcell, S. M., Saleheen, D., Scharf, J. M., Sklar, P., Sullivan, P. F., Tuomilehto, J., Tsuang, M. T., Watkins, H. C., Wilson, J. G., Daly, M. J., & MacArthur, D. G. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature, 536*(7616), 285-291.

Lewis, C. M. (2002). Genetic association studies: design, analysis and interpretation. *Brief Bioinform, 3*(2), 146-153.

Li, C., Wang, F., Yang, Y., Fu, F., Xu, C., Shi, L., Li, S., Xia, Y., Wu, G., Cheng, X., Liu, H., Wang, C., Wang, P., Hao, J., Ke, Y., Zhao, Y., Liu, M., Zhang, R., Gao, L., Yu, B., Zeng, Q., Liao, Y., Yang, B., Tu, X., & Wang, Q. K. (2011). Significant association of SNP rs2106261 in the *ZFHX3* gene with atrial fibrillation in a Chinese Han GeneID population. *Hum Genet, 129*(3), 239-246.

Li, N., Timofeyev, V., Tuteja, D., Xu, D., Lu, L., Zhang, Q., Zhang, Z., Singapuri, A., Albert, T. R., Rajagopal, A. V., Bond, C. T., Periasamy, M., Adelman, J., & Chiamvimonvat, N. (2009). Ablation of a Ca2+-activated K+ channel (SK2 channel) results in action potential prolongation in atrial myocytes and atrial fibrillation. *J Physiol, 587*(Pt 5), 1087-1100.

Liberman, R. F., Getz, K. D., Lin, A. E., Higgins, C. A., Sekhavat, S., Markenson, G. R., & Anderka, M. (2014). Delayed diagnosis of critical congenital heart defects: trends and associated factors. *Pediatrics, 134*(2), e373-381.

Lints, T. J., Parsons, L. M., Hartley, L., Lyons, I., & Harvey, R. P. (1993). Nkx-2.5: a novel murine homeobox gene expressed in early heart progenitor cells and their myogenic descendants. *Development, 119*(2), 419-431.

Loh, P. R., Tucker, G., Bulik-Sullivan, B. K., Vilhjalmsson, B. J., Finucane, H. K., Salem, R. M., Chasman, D. I., Ridker, P. M., Neale, B. M., Berger, B., Patterson, N., & Price, A. L. (2015). Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet, 47*(3), 284-290.

Low, S. K., Takahashi, A., Ebana, Y., Ozaki, K., Christophersen, I. E., Ellinor, P. T., Ogishima, S., Yamamoto, M., Satoh, M., Sasaki, M., Yamaji, T., Iwasaki, M., Tsugane, S., Tanaka, K., Naito, M., Wakai, K., Tanaka, H., Furukawa, T., Kubo, M., Ito, K., Kamatani, Y., & Tanaka, T. (2017). Identification of six new genetic loci associated with atrial fibrillation in the Japanese population. *Nat Genet, 49*(6), 953-958.

Lubitz, S. A., Lunetta, K. L., Lin, H., Arking, D. E., Trompet, S., Li, G., Krijthe, B. P., Chasman, D. I., Barnard, J., Kleber, M. E., Dorr, M., Ozaki, K., Smith, A. V., Muller-Nurasyid, M., Walter, S., Agarwal, S. K., Bis, J. C., Brody, J. A., Chen, L. Y., Everett, B. M., Ford, I., Franco, O. H., Harris, T. B., Hofman, A., Kaab, S., Mahida, S., Kathiresan, S., Kubo, M., Launer, L. J., MacFarlane, P. W., Magnani, J. W., McKnight, B.,

McManus, D. D., Peters, A., Psaty, B. M., Rose, L. M., Rotter, J. I., Silbernagel, G., Smith, J. D., Sotoodehnia, N., Stott, D. J., Taylor, K. D., Tomaschitz, A., Tsunoda, T., Uitterlinden, A. G., Van Wagoner, D. R., Volker, U., Volzke, H., Murabito, J. M., Sinner, M. F., Gudnason, V., Felix, S. B., Marz, W., Chung, M., Albert, C. M., Stricker, B. H., Tanaka, T., Heckbert, S. R., Jukema, J. W., Alonso, A., Benjamin, E. J., & Ellinor, P. T. (2014). Novel genetic markers associate with atrial fibrillation risk in Europeans and Japanese. *J Am Coll Cardiol, 63*(12), 1200-1210.

Lubitz, S. A., Yin, X., Fontes, J. D., Magnani, J. W., Rienstra, M., Pai, M., Villalon, M. L., Vasan, R. S., Pencina, M. J., Levy, D., Larson, M. G., Ellinor, P. T., & Benjamin, E. J. (2010). Association between familial atrial fibrillation and risk of new-onset atrial fibrillation. *Jama, 304*(20), 2263-2269.

Luck, J. C., & Engel, T. R. (1979). Dispersion of atrial refractoriness in patients with sinus node dysfunction. *Circulation, 60*(2), 404-412.

Macri, V., Mahida, S. N., Zhang, M. L., Sinner, M. F., Dolmatova, E. V., Tucker, N. R., McLellan, M., Shea, M. A., Milan, D. J., Lunetta, K. L., Benjamin, E. J., & Ellinor, P. T. (2014). A novel trafficking-defective *HCN4* mutation is associated with early-onset atrial fibrillation. *Heart Rhythm, 11*(6), 1055-1062.

Magnani, J. W., Williamson, M. A., Ellinor, P. T., Monahan, K. M., & Benjamin, E. J. (2009). P wave indices: current status and future directions in epidemiology, clinical, and research applications. *Circ Arrhythm Electrophysiol, 2*(1), 72-79.

Magnussen, C., Niiranen, T. J., Ojeda, F. M., Gianfagna, F., Blankenberg, S., Njolstad, I., Vartiainen, E., Sans, S., Pasterkamp, G., Hughes, M., Costanzo, S., Donati, M. B., Jousilahti, P., Linneberg, A., Palosaari, T., de Gaetano, G., Bobak, M., den Ruijter, H. M., Mathiesen, E., Jorgensen, T., Soderberg, S., Kuulasmaa, K., Zeller, T., Iacoviello, L., Salomaa, V., & Schnabel, R. B. (2017). Sex Differences and Similarities in Atrial Fibrillation Epidemiology, Risk Factors, and Mortality in Community Cohorts: Results From the BiomarCaRE Consortium (Biomarker for Cardiovascular Risk Assessment in Europe). *Circulation, 136*(17), 1588-1597.

Maisel, W. H., & Stevenson, L. W. (2003). Atrial fibrillation in heart failure: epidemiology, pathophysiology, and rationale for therapy. *Am J Cardiol, 91*(6a), 2d-8d.

Mandyam, M. C., Vedantham, V., Scheinman, M. M., Tseng, Z. H., Badhwar, N., Lee, B. K., Lee, R. J., Gerstenfeld, E. P., Olgin, J. E., & Marcus, G. M. (2012). Alcohol and vagal tone as triggers for paroxysmal atrial fibrillation. *Am J Cardiol, 110*(3), 364-368.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., Boehnke, M., Clark, A. G., Eichler, E. E., Gibson, G., Haines, J. L., Mackay, T. F., McCarroll, S. A., & Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature, 461*(7265), 747-753.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst, 22*(4), 719-748.

Marchini, J., Howie, B., Myers, S., McVean, G., & Donnelly, P. (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet, 39*(7), 906-913.

Markides, V., & Schilling, R. J. (2003). Atrial fibrillation: classification, pathophysiology, mechanisms and drug treatment. *Heart, 89*(8), 939-943.

Maron, B. J., & Maron, M. S. (2013). Hypertrophic cardiomyopathy. *Lancet, 381*(9862), 242-255.

Marouli, E., Graff, M., Medina-Gomez, C., Lo, K. S., Wood, A. R., Kjaer, T. R., Fine, R. S., Lu, Y., Schurmann, C., Highland, H. M., Rueger, S., Thorleifsson, G., Justice, A. E., Lamparter, D., Stirrups, K. E., Turcot, V., Young, K. L., Winkler, T. W., Esko, T., Karaderi, T., Locke, A. E., Masca, N. G., Ng, M. C., Mudgal, P., Rivas, M. A., Vedantam, S., Mahajan, A., Guo, X., Abecasis, G., Aben, K. K., Adair, L. S., Alam, D. S., Albrecht, E., Allin, K. H., Allison, M., Amouyel, P., Appel, E. V., Arveiler, D., Asselbergs, F. W., Auer, P. L., Balkau, B., Banas, B., Bang, L. E., Benn, M., Bergmann, S., Bielak, L. F., Bluher, M., Boeing, H., Boerwinkle, E., Boger, C. A., Bonnycastle, L. L., Bork-Jensen, J., Bots, M. L., Bottinger, E. P., Bowden, D. W., Brandslund, I., Breen, G., Brilliant, M. H., Broer, L., Burt, A. A., Butterworth, A. S., Carey, D. J., Caulfield, M. J., Chambers, J. C., Chasman, D. I., Chen, Y. I., Chowdhury, R., Christensen, C., Chu, A. Y., Cocca, M., Collins, F. S., Cook, J. P., Corley, J., Galbany, J. C., Cox, A. J., Cuellar-Partida, G., Danesh, J., Davies, G., de Bakker, P. I., de Borst, G. J., de Denus, S., de Groot, M. C., de Mutsert, R., Deary, I. J., Dedoussis, G., Demerath,

E. W., den Hollander, A. I., Dennis, J. G., Di Angelantonio, E., Drenos, F., Du, M., Dunning, A. M., Easton, D. F., Ebeling, T., Edwards, T. L., Ellinor, P. T., Elliott, P., Evangelou, E., Farmaki, A. E., Faul, J. D., Feitosa, M. F., Feng, S., Ferrannini, E., Ferrario, M. M., Ferrieres, J., Florez, J. C., Ford, I., Fornage, M., Franks, P. W., Frikke-Schmidt, R., Galesloot, T. E., Gan, W., Gandin, I., Gasparini, P., Giedraitis, V., Giri, A., Girotto, G., Gordon, S. D., Gordon-Larsen, P., Gorski, M., Grarup, N., Grove, M. L., Gudnason, V., Gustafsson, S., Hansen, T., Harris, K. M., Harris, T. B., Hattersley, A. T., Hayward, C., He, L., Heid, I. M., Heikkila, K., Helgeland, O., Hernesniemi, J., Hewitt, A. W., Hocking, L. J., Hollensted, M., Holmen, O. L., Hovingh, G. K., Howson, J. M., Hoyng, C. B., Huang, P. L., Hveem, K., Ikram, M. A., Ingelsson, E., Jackson, A. U., Jansson, J. H., Jarvik, G. P., Jensen, G. B., Jhun, M. A., Jia, Y., Jiang, X., Johansson, S., Jorgensen, M. E., Jorgensen, T., Jousilahti, P., Jukema, J. W., Kahali, B., Kahn, R. S., Kahonen, M., Kamstrup, P. R., Kanoni, S., Kaprio, J., Karaleftheri, M., Kardia, S. L., Karpe, F., Kee, F., Keeman, R., Kiemeney, L. A., Kitajima, H., Kluivers, K. B., Kocher, T., Komulainen, P., Kontto, J., Kooner, J. S., Kooperberg, C., Kovacs, P., Kriebel, J., Kuivaniemi, H., Kury, S., Kuusisto, J., La Bianca, M., Laakso, M., Lakka, T. A., Lange, E. M., Lange, L. A., Langefeld, C. D., Langenberg, C., Larson, E. B., Lee, I. T., Lehtimaki, T., Lewis, C. E., Li, H., Li, J., Li-Gao, R., Lin, H., Lin, L. A., Lin, X., Lind, L., Lindstrom, J., Linneberg, A., Liu, Y., Liu, Y., Lophatananon, A., Luan, J., Lubitz, S. A., Lyytikainen, L. P., Mackey, D. A., Madden, P. A., Manning, A. K., Mannisto, S., Marenne, G., Marten, J., Martin, N. G., Mazul, A. L., Meidtner, K., Metspalu, A., Mitchell, P., Mohlke, K. L., Mook-Kanamori, D. O., Morgan, A., Morris, A. D., Morris, A. P., Muller-Nurasyid, M., Munroe, P. B., Nalls, M. A., Nauck, M., Nelson, C. P., Neville, M., Nielsen, S. F., Nikus, K., Njolstad, P. R., Nordestgaard, B. G., Ntalla, I., O'Connel, J. R., Oksa, H., Loohuis, L. M., Ophoff, R. A., Owen, K. R., Packard, C. J., Padmanabhan, S., Palmer, C. N., Pasterkamp, G., Patel, A. P., Pattie, A., Pedersen, O., Peissig, P. L., Peloso, G. M., Pennell, C. E., Perola, M., Perry, J. A., Perry, J. R., Person, T. N., Pirie, A., Polasek, O., Posthuma, D., Raitakari, O. T., Rasheed, A., Rauramaa, R., Reilly, D. F., Reiner, A. P., Renstrom, F., Ridker, P. M., Rioux, J. D., Robertson, N., Robino, A., Rolandsson, O., Rudan, I., Ruth, K. S., Saleheen, D., Salomaa, V., Samani, N. J., Sandow, K., Sapkota, Y., Sattar, N., Schmidt, M. K., Schreiner, P. J., Schulze, M. B., Scott, R. A., Segura-Lepe, M. P., Shah, S., Sim, X., Sivapalaratnam, S., Small, K. S., Smith, A. V., Smith, J. A., Southam, L., Spector, T. D., Speliotes, E. K., Starr, J. M.,

Steinthorsdottir, V., Stringham, H. M., Stumvoll, M., Surendran, P., t Hart, L. M., Tansey, K. E., Tardif, J. C., Taylor, K. D., Teumer, A., Thompson, D. J., Thorsteinsdottir, U., Thuesen, B. H., Tonjes, A., Tromp, G., Trompet, S., Tsafantakis, E., Tuomilehto, J., Tybjaerg-Hansen, A., Tyrer, J. P., Uher, R., Uitterlinden, A. G., Ulivi, S., van der Laan, S. W., Van Der Leij, A. R., van Duijn, C. M., van Schoor, N. M., van Setten, J., Varbo, A., Varga, T. V., Varma, R., Edwards, D. R., Vermeulen, S. H., Vestergaard, H., Vitart, V., Vogt, T. F., Vozzi, D., Walker, M., Wang, F., Wang, C. A., Wang, S., Wang, Y., Wareham, N. J., Warren, H. R., Wessel, J., Willems, S. M., Wilson, J. G., Witte, D. R., Woods, M. O., Wu, Y., Yaghootkar, H., Yao, J., Yao, P., Yerges-Armstrong, L. M., Young, R., Zeggini, E., Zhan, X., Zhang, W., Zhao, J. H., Zhao, W., Zhao, W., Zheng, H., Zhou, W., Rotter, J. I., Boehnke, M., Kathiresan, S., McCarthy, M. I., Willer, C. J., Stefansson, K., Borecki, I. B., Liu, D. J., North, K. E., Heard-Costa, N. L., Pers, T. H., Lindgren, C. M., Oxvig, C., Kutalik, Z., Rivadeneira, F., Loos, R. J., Frayling, T. M., Hirschhorn, J. N., Deloukas, P., & Lettre, G. (2017). Rare and low-frequency coding variants alter human adult height. *Nature, 542*(7640), 186-190.

Marrouche, N. F., Wilber, D., Hindricks, G., Jais, P., Akoum, N., Marchlinski, F., Kholmovski, E., Burgon, N., Hu, N., Mont, L., Deneke, T., Duytschaever, M., Neumann, T., Mansour, M., Mahnkopf, C., Herweg, B., Daoud, E., Wissner, E., Bansmann, P., & Brachmann, J. (2014). Association of atrial tissue fibrosis identified by delayed enhancement MRI and atrial fibrillation catheter ablation: the DECAAF study. *Jama, 311*(5), 498-506.

McBride, K. L., Pignatelli, R., Lewin, M., Ho, T., Fernbach, S., Menesses, A., Lam, W., Leal, S. M., Kaplan, N., Schliekelman, P., Towbin, J. A., & Belmont, J. W. (2005). Inheritance analysis of congenital left ventricular outflow tract obstruction malformations: Segregation, multiplex relative risk, and heritability. *Am J Med Genet A, 134a*(2), 180-186.

McBride, K. L., Riley, M. F., Zender, G. A., Fitzgerald-Butt, S. M., Towbin, J. A., Belmont, J. W., & Cole, S. E. (2008). *NOTCH1* mutations in individuals with left ventricular outflow tract malformations reduce ligand-induced signaling. *Hum Mol Genet, 17*(18), 2886-2893.

McCann, K. L., & Baserga, S. J. (2013). Genetics. Mysterious ribosomopathies. *Science, 341*(6148), 849-850.

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., Kang, H. M., Fuchsberger, C., Danecek, P., Sharp, K., Luo, Y.,

Sidore, C., Kwong, A., Timpson, N., Koskinen, S., Vrieze, S., Scott, L. J., Zhang, H., Mahajan, A., Veldink, J., Peters, U., Pato, C., van Duijn, C. M., Gillies, C. E., Gandin, I., Mezzavilla, M., Gilly, A., Cocca, M., Traglia, M., Angius, A., Barrett, J. C., Boomsma, D., Branham, K., Breen, G., Brummett, C. M., Busonero, F., Campbell, H., Chan, A., Chen, S., Chew, E., Collins, F. S., Corbin, L. J., Smith, G. D., Dedoussis, G., Dorr, M., Farmaki, A. E., Ferrucci, L., Forer, L., Fraser, R. M., Gabriel, S., Levy, S., Groop, L., Harrison, T., Hattersley, A., Holmen, O. L., Hveem, K., Kretzler, M., Lee, J. C., McGue, M., Meitinger, T., Melzer, D., Min, J. L., Mohlke, K. L., Vincent, J. B., Nauck, M., Nickerson, D., Palotie, A., Pato, M., Pirastu, N., McInnis, M., Richards, J. B., Sala, C., Salomaa, V., Schlessinger, D., Schoenherr, S., Slagboom, P. E., Small, K., Spector, T., Stambolian, D., Tuke, M., Tuomilehto, J., Van den Berg, L. H., Van Rheenen, W., Volker, U., Wijmenga, C., Toniolo, D., Zeggini, E., Gasparini, P., Sampson, M. G., Wilson, J. F., Frayling, T., de Bakker, P. I., Swertz, M. A., McCarroll, S., Kooperberg, C., Dekker, A., Altshuler, D., Willer, C., Iacono, W., Ripatti, S., Soranzo, N., Walter, K., Swaroop, A., Cucca, F., Anderson, C. A., Myers, R. M., Boehnke, M., McCarthy, M. I., & Durbin, R. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet, 48*(10), 1279-1283.

McElhinney, D. B., Geiger, E., Blinder, J., Benson, D. W., & Goldmuntz, E. (2003). *NKX2.5* mutations in patients with congenital heart disease. *J Am Coll Cardiol, 42*(9), 1650-1655.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res, 20*(9), 1297-1303.

McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P., & Cunningham, F. (2010). Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics, 26*(16), 2069-2070.

Miano, J. M. (2010). Role of serum response factor in the pathogenesis of disease. *Lab Invest, 90*(9), 1274-1284.

Miyasaka, Y., Barnes, M. E., Gersh, B. J., Cha, S. S., Bailey, K. R., Abhayaratna, W. P., Seward, J. B., & Tsang, T. S. (2006). Secular trends in incidence of atrial fibrillation in Olmsted County, Minnesota, 1980 to 2000, and implications on the projections for future prevalence. *Circulation, 114*(2), 119-125.

Mommersteeg, M. T., Brown, N. A., Prall, O. W., de Gier-de Vries, C., Harvey, R. P., Moorman, A. F., & Christoffels, V. M. (2007). Pitx2c and Nkx2-5 are required for the formation and identity of the pulmonary myocardium. *Circ Res, 101*(9), 902-909.

Moncayo-Arlandi, J., & Brugada, R. (2017). Unmasking the molecular link between arrhythmogenic cardiomyopathy and Brugada syndrome. *Nat Rev Cardiol, 14*(12), 744-756.

Morgan, G. J., Lee, K. J., Chaturvedi, R., Bradley, T. J., Mertens, L., & Benson, L. (2013). Systemic blood pressure after stent management for arch coarctation implications for clinical care. *JACC Cardiovasc Interv, 6*(2), 192-201.

Nakamura, H., Sawamura, D., Goto, M., Nakamura, H., McMillan, J. R., Park, S., Kono, S., Hasegawa, S., Paku, S., Nakamura, T., Ogiso, Y., & Shimizu, H. (2005). Epidermolysis bullosa simplex associated with pyloric atresia is a novel clinical subtype caused by mutations in the plectin gene (*PLEC1*). *J Mol Diagn, 7*(1), 28-35.

Nattel, S. (2002). New ideas about atrial fibrillation 50 years on. *Nature, 415*(6868), 219-226.

Nattel, S., Burstein, B., & Dobrev, D. (2008). Atrial remodeling and atrial fibrillation: mechanisms and implications. *Circ Arrhythm Electrophysiol, 1*(1), 62-73.

Nattel, S., & Harada, M. (2014). Atrial remodeling and atrial fibrillation: recent advances and translational perspectives. *J Am Coll Cardiol, 63*(22), 2335-2345.

Nielsen, J. B., Fritsche, L. G., Zhou, W., Teslovich, T. M., Holmen, O. L., Gustafsson, S., Gabrielsen, M. E., Schmidt, E. M., Beaumont, R., Wolford, B. N., Lin, M., Brummett, C. M., Preuss, M. H., Refsgaard, L., Bottinger, E. P., Graham, S. E., Surakka, I., Chu, Y., Skogholt, A. H., Dalen, H., Boyle, A. P., Oral, H., Herron, T. J., Kitzman, J., Jalife, J., Svendsen, J. H., Olesen, M. S., Njolstad, I., Lochen, M. L., Baras, A., Gottesman, O., Marcketta, A., O'Dushlaine, C., Ritchie, M. D., Wilsgaard, T., Loos, R. J. F., Frayling, T. M., Boehnke, M., Ingelsson, E., Carey, D. J., Dewey, F. E., Kang, H. M., Abecasis, G. R., Hveem, K., & Willer, C. J. (2018a). Genome-wide Study of Atrial Fibrillation Identifies Seven Risk Loci and Highlights Biological Pathways and Regulatory Elements Involved in Cardiac Development. *Am J Hum Genet, 102*(1), 103-115.

Nielsen, J. B., Kuhl, J. T., Pietersen, A., Graff, C., Lind, B., Struijk, J. J., Olesen, M. S., Sinner, M. F., Bachmann, T. N., Haunso, S., Nordestgaard, B.

G., Ellinor, P. T., Svendsen, J. H., Kofoed, K. F., Kober, L., & Holst, A. G. (2015). P-wave duration and the risk of atrial fibrillation: Results from the Copenhagen ECG Study. *Heart Rhythm, 12*(9), 1887-1895.

Nielsen, J. B., Pietersen, A., Graff, C., Lind, B., Struijk, J. J., Olesen, M. S., Haunso, S., Gerds, T. A., Ellinor, P. T., Kober, L., Svendsen, J. H., & Holst, A. G. (2013). Risk of atrial fibrillation as a function of the electrocardiographic PR interval: results from the Copenhagen ECG Study. *Heart Rhythm, 10*(9), 1249-1256.

Nielsen, J. B., Thorolfsdottir, R. B., Fritsche, L. G., Zhou, W., Skov, M. W., Graham, S. E., Herron, T. J., McCarthy, S., Schmidt, E. M., Sveinbjornsson, G., Surakka, I., Mathis, M. R., Yamazaki, M., Crawford, R. D., Gabrielsen, M. E., Skogholt, A. H., Holmen, O. L., Lin, M., Wolford, B. N., Dey, R., Dalen, H., Sulem, P., Chung, J. H., Backman, J. D., Arnar, D. O., Thorsteinsdottir, U., Baras, A., O'Dushlaine, C., Holst, A. G., Wen, X., Hornsby, W., Dewey, F. E., Boehnke, M., Kheterpal, S., Mukherjee, B., Lee, S., Kang, H. M., Holm, H., Kitzman, J., Shavit, J. A., Jalife, J., Brummett, C. M., Teslovich, T. M., Carey, D. J., Gudbjartsson, D. F., Stefansson, K., Abecasis, G. R., Hveem, K., & Willer, C. J. (2018b). Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat Genet, 50*(9), 1234-1239.

Niimura, H., Patton, K. K., McKenna, W. J., Soults, J., Maron, B. J., Seidman, J. G., & Seidman, C. E. (2002). Sarcomere protein gene mutations in hypertrophic cardiomyopathy of the elderly. *Circulation, 105*(4), 446-451.

Nikolic, B., Mac Nulty, E., Mir, B., & Wiche, G. (1996). Basic amino acid residue cluster within nuclear targeting sequence motif is essential for cytoplasmic plectin-vimentin network junctions. *J Cell Biol, 134*(6), 1455-1467.

Noorman, M., van der Heyden, M. A., van Veen, T. A., Cox, M. G., Hauer, R. N., de Bakker, J. M., & van Rijen, H. V. (2009). Cardiac cell-cell junctions in health and disease: Electrical versus mechanical coupling. *J Mol Cell Cardiol, 47*(1), 23-31.

Oakes, R. S., Badger, T. J., Kholmovski, E. G., Akoum, N., Burgon, N. S., Fish, E. N., Blauer, J. J., Rao, S. N., DiBella, E. V., Segerson, N. M., Daccarett, M., Windfelder, J., McGann, C. J., Parker, D., MacLeod, R. S., & Marrouche, N. F. (2009). Detection and quantification of left atrial structural remodeling with delayed-enhancement magnetic resonance imaging in patients with atrial fibrillation. *Circulation, 119*(13), 1758-1767.

Olesen, J. B., Sorensen, R., Hansen, M. L., Lamberts, M., Weeke, P., Mikkelsen, A. P., Kober, L., Gislason, G. H., Torp-Pedersen, C., & Fosbol, E. L. (2015). Non-vitamin K antagonist oral anticoagulation agents in anticoagulant naive atrial fibrillation patients: Danish nationwide descriptive data 2011-2013. *Europace, 17*(2), 187-193.

Olivetti, G., Abbi, R., Quaini, F., Kajstura, J., Cheng, W., Nitahara, J. A., Quaini, E., Di Loreto, C., Beltrami, C. A., Krajewski, S., Reed, J. C., & Anversa, P. (1997). Apoptosis in the failing human heart. *N Engl J Med, 336*(16), 1131-1141.

Oyen, N., Ranthe, M. F., Carstensen, L., Boyd, H. A., Olesen, M. S., Olesen, S. P., Wohlfahrt, J., & Melbye, M. (2012). Familial aggregation of lone atrial fibrillation in young persons. *J Am Coll Cardiol, 60*(10), 917-921.

Pagani, F., & Baralle, F. E. (2004). Genomic variants in exons and introns: identifying the splicing spoilers. *Nat Rev Genet, 5*(5), 389-396.

Patten, M., Pecha, S., & Aydin, A. (2018). Atrial Fibrillation in Hypertrophic Cardiomyopathy: Diagnosis and Considerations for Management. *J Atr Fibrillation, 10*(5), 1556.

Pers, T. H., Karjalainen, J. M., Chan, Y., Westra, H. J., Wood, A. R., Yang, J., Lui, J. C., Vedantam, S., Gustafsson, S., Esko, T., Frayling, T., Speliotes, E. K., Boehnke, M., Raychaudhuri, S., Fehrmann, R. S., Hirschhorn, J. N., & Franke, L. (2015). Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun, 6*, 5890.

Pfeffer, M. A., & Braunwald, E. (1990). Ventricular remodeling after myocardial infarction. Experimental observations and clinical implications. *Circulation, 81*(4), 1161-1172.

Pfeufer, A., van Noord, C., Marciante, K. D., Arking, D. E., Larson, M. G., Smith, A. V., Tarasov, K. V., Muller, M., Sotoodehnia, N., Sinner, M. F., Verwoert, G. C., Li, M., Kao, W. H., Kottgen, A., Coresh, J., Bis, J. C., Psaty, B. M., Rice, K., Rotter, J. I., Rivadeneira, F., Hofman, A., Kors, J. A., Stricker, B. H., Uitterlinden, A. G., van Duijn, C. M., Beckmann, B. M., Sauter, W., Gieger, C., Lubitz, S. A., Newton-Cheh, C., Wang, T. J., Magnani, J. W., Schnabel, R. B., Chung, M. K., Barnard, J., Smith, J. D., Van Wagoner, D. R., Vasan, R. S., Aspelund, T., Eiriksdottir, G., Harris, T. B., Launer, L. J., Najjar, S. S., Lakatta, E., Schlessinger, D., Uda, M., Abecasis, G. R., Muller-Myhsok, B., Ehret, G. B., Boerwinkle, E., Chakravarti, A., Soliman, E. Z., Lunetta, K. L., Perz, S., Wichmann, H. E., Meitinger, T., Levy, D., Gudnason, V.,

Ellinor, P. T., Sanna, S., Kaab, S., Witteman, J. C., Alonso, A., Benjamin, E. J., & Heckbert, S. R. (2010). Genome-wide association study of PR interval. *Nat Genet, 42*(2), 153-159.

Piccini, J. P., & Fauchier, L. (2016). Rhythm control in atrial fibrillation. *Lancet, 388*(10046), 829-840.

Piedra, M. E., Icardo, J. M., Albajar, M., Rodriguez-Rey, J. C., & Ros, M. A. (1998). Pitx2 participates in the late phase of the pathway controlling left-right asymmetry. *Cell, 94*(3), 319-324.

Posch, M. G., Waldmuller, S., Muller, M., Scheffold, T., Fournier, D., Andrade-Navarro, M. A., De Geeter, B., Guillaumont, S., Dauphin, C., Yousseff, D., Schmitt, K. R., Perrot, A., Berger, F., Hetzer, R., Bouvagnet, P., & Ozcelik, C. (2011). Cardiac alpha-myosin (*MYH6*) is the predominant sarcomeric disease gene for familial atrial septal defects. *PLoS One, 6*(12), e28872.

Presbitero, P., Demarie, D., Villani, M., Perinetto, E. A., Riva, G., Orzan, F., Bobbio, M., Morea, M., & Brusca, A. (1987). Long term results (15-30 years) of surgical repair of aortic coarctation. *Br Heart J, 57*(5), 462-467.

Pulit, S. L., Weng, L. C., McArdle, P. F., Trinquart, L., Choi, S. H., Mitchell, B. D., Rosand, J., de Bakker, P. I. W., Benjamin, E. J., Ellinor, P. T., Kittner, S. J., Lubitz, S. A., & Anderson, C. D. (2018). Atrial fibrillation genetic risk differentiates cardioembolic stroke from other stroke subtypes. *Neurol Genet, 4*(6), e293.

Pulkkinen, L., Smith, F. J., Shimizu, H., Murata, S., Yaoita, H., Hachisuka, H., Nishikawa, T., McLean, W. H., & Uitto, J. (1996). Homozygous deletion mutations in the plectin gene (*PLEC1*) in patients with epidermolysis bullosa simplex associated with late-onset muscular dystrophy. *Hum Mol Genet, 5*(10), 1539-1546.

Qi, Y., Ranish, J. A., Zhu, X., Krones, A., Zhang, J., Aebersold, R., Rose, D. W., Rosenfeld, M. G., & Carriere, C. (2008). Atbf1 is required for the Pit1 gene early activation. *Proc Natl Acad Sci U S A, 105*(7), 2481-2486.

Raiser, D. M., Narla, A., & Ebert, B. L. (2014). The emerging importance of ribosomal dysfunction in the pathogenesis of hematologic disorders. *Leuk Lymphoma, 55*(3), 491-500.

Rampazzo, A., Calore, M., van Hengel, J., & van Roy, F. (2014). Intercalated discs and arrhythmogenic cardiomyopathy. *Circ Cardiovasc Genet, 7*(6), 930-940.

Rangrez, A. Y., Eden, M., Poyanmehr, R., Kuhn, C., Stiebeling, K., Dierck, F., Bernt, A., Lullmann-Rauch, R., Weiler, H., Kirchof, P., Frank, D., & Frey, N. (2016). Myozap Deficiency Promotes Adverse Cardiac Remodeling via Differential Regulation of Mitogen-activated Protein Kinase/Serum-response Factor and beta-Catenin/GSK-3beta Protein Signaling. *J Biol Chem, 291*(8), 4128-4143.

Reich, D. E., & Lander, E. S. (2001). On the allelic spectrum of human disease. *Trends Genet, 17*(9), 502-510.

Ren, J., Samson, W. K., & Sowers, J. R. (1999). Insulin-like growth factor I as a cardiac hormone: physiological and pathophysiological implications in heart disease. *J Mol Cell Cardiol, 31*(11), 2049-2061.

Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science, 273*(5281), 1516-1517.

Ritchie, G. R., Dunham, I., Zeggini, E., & Flicek, P. (2014). Functional annotation of noncoding sequence variants. *Nat Methods, 11*(3), 294-296.

Roselli, C., Chaffin, M. D., Weng, L. C., Aeschbacher, S., Ahlberg, G., Albert, C. M., Almgren, P., Alonso, A., Anderson, C. D., Aragam, K. G., Arking, D. E., Barnard, J., Bartz, T. M., Benjamin, E. J., Bihlmeyer, N. A., Bis, J. C., Bloom, H. L., Boerwinkle, E., Bottinger, E. B., Brody, J. A., Calkins, H., Campbell, A., Cappola, T. P., Carlquist, J., Chasman, D. I., Chen, L. Y., Chen, Y. I., Choi, E. K., Choi, S. H., Christophersen, I. E., Chung, M. K., Cole, J. W., Conen, D., Cook, J., Crijns, H. J., Cutler, M. J., Damrauer, S. M., Daniels, B. R., Darbar, D., Delgado, G., Denny, J. C., Dichgans, M., Dorr, M., Dudink, E. A., Dudley, S. C., Esa, N., Esko, T., Eskola, M., Fatkin, D., Felix, S. B., Ford, I., Franco, O. H., Geelhoed, B., Grewal, R. P., Gudnason, V., Guo, X., Gupta, N., Gustafsson, S., Gutmann, R., Hamsten, A., Harris, T. B., Hayward, C., Heckbert, S. R., Hernesniemi, J., Hocking, L. J., Hofman, A., Horimoto, A., Huang, J., Huang, P. L., Huffman, J., Ingelsson, E., Ipek, E. G., Ito, K., Jimenez-Conde, J., Johnson, R., Jukema, J. W., Kaab, S., Kahonen, M., Kamatani, Y., Kane, J. P., Kastrati, A., Kathiresan, S., Katschnig-Winter, P., Kavousi, M., Kessler, T., Kietselaer, B. L., Kirchhof, P., Kleber, M. E., Knight, S., Krieger, J. E., Kubo, M., Launer, L. J., Laurikka, J., Lehtimaki, T., Leineweber, K., Lemaitre, R. N., Li, M., Lim, H. E., Lin, H. J., Lin, H., Lind, L., Lindgren, C. M., Lokki, M. L., London, B., Loos, R. J. F., Low, S. K., Lu, Y., Lyytikainen, L. P., Macfarlane, P. W., Magnusson, P. K., Mahajan, A., Malik, R., Mansur, A. J., Marcus, G. M., Margolin, L., Margulies, K. B., Marz, W., McManus, D. D., Melander, O., Mohanty, S., Montgomery, J. A.,

Morley, M. P., Morris, A. P., Muller-Nurasyid, M., Natale, A., Nazarian, S., Neumann, B., Newton-Cheh, C., Niemeijer, M. N., Nikus, K., Nilsson, P., Noordam, R., Oellers, H., Olesen, M. S., Orho-Melander, M., Padmanabhan, S., Pak, H. N., Pare, G., Pedersen, N. L., Pera, J., Pereira, A., Porteous, D., Psaty, B. M., Pulit, S. L., Pullinger, C. R., Rader, D. J., Refsgaard, L., Ribases, M., Ridker, P. M., Rienstra, M., Risch, L., Roden, D. M., Rosand, J., Rosenberg, M. A., Rost, N., Rotter, J. I., Saba, S., Sandhu, R. K., Schnabel, R. B., Schramm, K., Schunkert, H., Schurman, C., Scott, S. A., Seppala, I., Shaffer, C., Shah, S., Shalaby, A. A., Shim, J., Shoemaker, M. B., Siland, J. E., Sinisalo, J., Sinner, M. F., Slowik, A., Smith, A. V., Smith, B. H., Smith, J. G., Smith, J. D., Smith, N. L., Soliman, E. Z., Sotoodehnia, N., Stricker, B. H., Sun, A., Sun, H., Svendsen, J. H., Tanaka, T., Tanriverdi, K., Taylor, K. D., Teder-Laving, M., Teumer, A., Theriault, S., Trompet, S., Tucker, N. R., Tveit, A., Uitterlinden, A. G., Van Der Harst, P., Van Gelder, I. C., Van Wagoner, D. R., Verweij, N., Vlachopoulou, E., Volker, U., Wang, B., Weeke, P. E., Weijs, B., Weiss, R., Weiss, S., Wells, Q. S., Wiggins, K. L., Wong, J. A., Woo, D., Worrall, B. B., Yang, P. S., Yao, J., Yoneda, Z. T., Zeller, T., Zeng, L., Lubitz, S. A., Lunetta, K. L., & Ellinor, P. T. (2018). Multi-ethnic genome-wide association study for atrial fibrillation. *Nat Genet, 50*(9), 1225-1233.

Rudolph, A. M., Heymann, M. A., & Spitznas, U. (1972). Hemodynamic considerations in the development of narrowing of the aorta. *Am J Cardiol, 30*(5), 514-525.

Ruffier, M., Kahari, A., Komorowska, M., Keenan, S., Laird, M., Longden, I., Proctor, G., Searle, S., Staines, D., Taylor, K., Vullo, A., Yates, A., Zerbino, D., & Flicek, P. (2017). Ensembl core software resources: storage and programmatic access for DNA sequence and genome annotation. *Database (Oxford), 2017*(1).

Russell, G. A., Berry, P. J., Watterson, K., Dhasmana, J. P., & Wisheart, J. D. (1991). Patterns of ductal tissue in coarctation of the aorta in the first three months of life. *J Thorac Cardiovasc Surg, 102*(4), 596-601.

Sabatine, M. S., Giugliano, R. P., Keech, A., Honarpour, N., Wang, H., Liu, T., Wasserman, S. M., Scott, R., Sever, P. S., & Pedersen, T. R. (2016). Rationale and design of the Further cardiovascular OUtcomes Research with PCSK9 Inhibition in subjects with Elevated Risk trial. *Am Heart J, 173*, 94-101.

Saksena, S., & Camm, A. J. (2012). *Electrophysiological disorders of the heart* (2nd ed.). Philadelphia: Elsevier/Saunders.

Sakurai, A., Nakai, A., & DeGroot, L. J. (1990). Structural analysis of human thyroid hormone receptor beta gene. *Mol Cell Endocrinol, 71*(2), 83-91.

Sánchez-Quintana, D., López-Mínguez, J. R., Pizarro, G., Murillo, M., & Cabrera, J. A. (2012). Triggers and anatomical substrates in the genesis and perpetuation of atrial fibrillation. *Current cardiology reviews, 8*(4), 310-326.

Sanders, P., Morton, J. B., Kistler, P. M., Spence, S. J., Davidson, N. C., Hussin, A., Vohra, J. K., Sparks, P. B., & Kalman, J. M. (2004). Electrophysiological and electroanatomic characterization of the atria in sinus node disease: evidence of diffuse atrial remodeling. *Circulation, 109*(12), 1514-1522.

Sauert, M., Temmel, H., & Moll, I. (2015). Heterogeneity of the translational machinery: Variations on a common theme. *Biochimie, 114*, 39-47.

Savio-Galimberti, E., Weeke, P., Muhammad, R., Blair, M., Ansari, S., Short, L., Atack, T. C., Kor, K., Vanoye, C. G., Olesen, M. S., LuCamp, Yang, T., George, A. L., Jr., Roden, D. M., & Darbar, D. (2014). *SCN10A*/Nav1.8 modulation of peak and late sodium currents in patients with early onset atrial fibrillation. *Cardiovasc Res, 104*(2), 355-363.

Schmidt, E. M., Zhang, J., Zhou, W., Chen, J., Mohlke, K. L., Chen, Y. E., & Willer, C. J. (2015). GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. *Bioinformatics, 31*(16), 2601-2606.

Schotten, U., Verheule, S., Kirchhof, P., & Goette, A. (2011). Pathophysiological mechanisms of atrial fibrillation: a translational appraisal. *Physiol Rev, 91*(1), 265-325.

Schroder, R., Kunz, W. S., Rouan, F., Pfendner, E., Tolksdorf, K., Kappes-Horn, K., Altenschmidt-Mehring, M., Knoblich, R., van der Ven, P. F., Reimann, J., Furst, D. O., Blumcke, I., Vielhaber, S., Zillikens, D., Eming, S., Klockgether, T., Uitto, J., Wiche, G., & Rolfs, A. (2002). Disorganization of the desmin cytoskeleton and mitochondrial dysfunction in plectin-related epidermolysis bullosa simplex with muscular dystrophy. *J Neuropathol Exp Neurol, 61*(6), 520-530.

Seeger, T. S., Frank, D., Rohr, C., Will, R., Just, S., Grund, C., Lyon, R., Luedde, M., Koegl, M., Sheikh, F., Rottbauer, W., Franke, W. W., Katus, H. A., Olson, E. N., & Frey, N. (2010). Myozap, a novel intercalated disc protein, activates serum response factor-dependent signaling and is

required to maintain cardiac function in vivo. *Circ Res, 106*(5), 880-890.

Seifert, G. J., Lawson, D., & Wiche, G. (1992). Immunolocalization of the intermediate filament-associated protein plectin at focal contacts and actin stress fibers. *Eur J Cell Biol, 59*(1), 138-147.

Selcen, D., Juel, V. C., Hobson-Webb, L. D., Smith, E. C., Stickler, D. E., Bite, A. V., Ohno, K., & Engel, A. G. (2011). Myasthenic syndrome caused by plectinopathy. *Neurology, 76*(4), 327-336.

Selvetella, G., Hirsch, E., Notte, A., Tarone, G., & Lembo, G. (2004). Adaptive and maladaptive hypertrophic pathways: points of convergence and divergence. *Cardiovasc Res, 63*(3), 373-380.

Shameer, K., Tripathi, L. P., Kalari, K. R., Dudley, J. T., & Sowdhamini, R. (2016). Interpreting functional effects of coding variants: challenges in proteome-scale prediction, annotation and assessment. *Brief Bioinform, 17*(5), 841-862.

Shi, L., Li, C., Wang, C., Xia, Y., Wu, G., Wang, F., Xu, C., Wang, P., Li, X., Wang, D., Xiong, X., Bai, Y., Liu, M., Liu, J., Ren, X., Gao, L., Wang, B., Zeng, Q., Yang, B., Ma, X., Yang, Y., Tu, X., & Wang, Q. K. (2009). Assessment of association of rs2200733 on chromosome 4q25 with atrial fibrillation and ischemic stroke in a Chinese Han population. *Hum Genet, 126*(6), 843-849.

Shinebourne, E. A., Tam, A. S., Elseed, A. M., Paneth, M., Lennox, S. C., & Cleland, W. P. (1976). Coarctation of the aorta in infancy and childhood. *Br Heart J, 38*(4), 375-380.

Sim, N. L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., & Ng, P. C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res, 40*(Web Server issue), W452-457.

Singh, P. K., Marzo, A., Staicu, C., William, M. G., Wilkinson, I., Lawford, P. V., Rufenacht, D. A., Bijlenga, P., Frangi, A. F., Hose, R., Patel, U. J., & Coley, S. C. (2010). The effects of aortic coarctation on cerebral hemodynamics and its importance in the etiopathogenesis of intracranial aneurysms. *J Vasc Interv Neurol, 3*(1), 17-30.

Sinner, M. F., Tucker, N. R., Lunetta, K. L., Ozaki, K., Smith, J. G., Trompet, S., Bis, J. C., Lin, H., Chung, M. K., Nielsen, J. B., Lubitz, S. A., Krijthe, B. P., Magnani, J. W., Ye, J., Gollob, M. H., Tsunoda, T., Muller-Nurasyid, M., Lichtner, P., Peters, A., Dolmatova, E., Kubo, M., Smith, J. D., Psaty, B. M., Smith, N. L., Jukema, J. W., Chasman, D. I., Albert, C. M., Ebana, Y., Furukawa, T., Macfarlane, P. W., Harris, T. B., Darbar, D., Dorr, M., Holst, A. G., Svendsen, J. H., Hofman, A.,

Uitterlinden, A. G., Gudnason, V., Isobe, M., Malik, R., Dichgans, M., Rosand, J., Van Wagoner, D. R., Benjamin, E. J., Milan, D. J., Melander, O., Heckbert, S. R., Ford, I., Liu, Y., Barnard, J., Olesen, M. S., Stricker, B. H., Tanaka, T., Kaab, S., & Ellinor, P. T. (2014). Integrating genetic, transcriptional, and functional analyses to identify 5 novel genes for atrial fibrillation. *Circulation, 130*(15), 1225-1235.

Smith, G. D., & Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol, 32*(1), 1-22.

So, H. C., Gui, A. H., Cherny, S. S., & Sham, P. C. (2011). Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genet Epidemiol, 35*(5), 310-317.

Steffens, J. C., Bourne, M. W., Sakuma, H., O'Sullivan, M., & Higgins, C. B. (1994). Quantification of collateral blood flow in coarctation of the aorta by velocity encoded cine magnetic resonance imaging. *Circulation, 90*(2), 937-943.

Steinthorsdottir, V., Thorleifsson, G., Sulem, P., Helgason, H., Grarup, N., Sigurdsson, A., Helgadottir, H. T., Johannsdottir, H., Magnusson, O. T., Gudjonsson, S. A., Justesen, J. M., Harder, M. N., Jorgensen, M. E., Christensen, C., Brandslund, I., Sandbaek, A., Lauritzen, T., Vestergaard, H., Linneberg, A., Jorgensen, T., Hansen, T., Daneshpour, M. S., Fallah, M. S., Hreidarsson, A. B., Sigurdsson, G., Azizi, F., Benediktsson, R., Masson, G., Helgason, A., Kong, A., Gudbjartsson, D. F., Pedersen, O., Thorsteinsdottir, U., & Stefansson, K. (2014). Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. *Nat Genet, 46*(3), 294-298.

Stewart-Hutchinson, P. J., Hale, C. M., Wirtz, D., & Hodzic, D. (2008). Structural requirements for the assembly of LINC complexes and their function in cellular mechanical stiffness. *Exp Cell Res, 314*(8), 1892-1905.

Stewart, S., Hart, C. L., Hole, D. J., & McMurray, J. J. (2002). A population-based study of the long-term risks associated with atrial fibrillation: 20-year follow-up of the Renfrew/Paisley study. *Am J Med, 113*(5), 359-364.

Stewart, S., Murphy, N. F., Walker, A., McGuire, A., & McMurray, J. J. (2004). Cost of an emerging epidemic: an economic analysis of atrial fibrillation in the UK. *Heart, 90*(3), 286-292.

Stout, K. K., Daniels, C. J., Aboulhosn, J. A., Bozkurt, B., Broberg, C. S., Colman, J. M., Crumb, S. R., Dearani, J. A., Fuller, S., Gurvitz, M., Khairy, P., Landzberg, M. J., Saidi, A., Valente, A. M., & Van Hare, G. F. (2019). 2018 AHA/ACC Guideline for the Management of Adults With Congenital Heart Disease: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *J Am Coll Cardiol, 73*(12), 1494-1563.

Strachan, T., & Read, A. P. (2011). *Human molecular genetics* (4th ed.). New York: Garland Science/Taylor & Francis Group.

Strafford, M. A., Griffiths, S. P., & Gersony, W. M. (1982). Coarctation of the aorta: a study in delayed detection. *Pediatrics, 69*(2), 159-163.

Styrkarsdottir, U., Lund, S. H., Thorleifsson, G., Zink, F., Stefansson, O. A., Sigurdsson, J. K., Juliusson, K., Bjarnadottir, K., Sigurbjornsdottir, S., Jonsson, S., Norland, K., Stefansdottir, L., Sigurdsson, A., Sveinbjornsson, G., Oddsson, A., Bjornsdottir, G., Gudmundsson, R. L., Halldorsson, G. H., Rafnar, T., Jonsdottir, I., Steingrimsson, E., Norddahl, G. L., Masson, G., Sulem, P., Jonsson, H., Ingvarsson, T., Gudbjartsson, D. F., Thorsteinsdottir, U., & Stefansson, K. (2018). Meta-analysis of Icelandic and UK data sets identifies missense variants in *SMO*, *IL11*, *COL11A1* and 13 more new loci associated with osteoarthritis. *Nat Genet, 50*(12), 1681-1687.

Styrkarsdottir, U., Thorleifsson, G., Sulem, P., Gudbjartsson, D. F., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., Oddsson, A., Helgason, A., Magnusson, O. T., Walters, G. B., Frigge, M. L., Helgadottir, H. T., Johannsdottir, H., Bergsteinsdottir, K., Ogmundsdottir, M. H., Center, J. R., Nguyen, T. V., Eisman, J. A., Christiansen, C., Steingrimsson, E., Jonasson, J. G., Tryggvadottir, L., Eyjolfsson, G. I., Theodors, A., Jonsson, T., Ingvarsson, T., Olafsson, I., Rafnar, T., Kong, A., Sigurdsson, G., Masson, G., Thorsteinsdottir, U., & Stefansson, K. (2013). Nonsense mutation in the *LGR4* gene is associated with several human diseases and other traits. *Nature, 497*(7450), 517-520.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med, 12*(3), e1001779.

Sveinbjornsson, G., Albrechtsen, A., Zink, F., Gudjonsson, S. A., Oddson, A., Masson, G., Holm, H., Kong, A., Thorsteinsdottir, U., Sulem, P., Gudbjartsson, D. F., & Stefansson, K. (2016). Weighting sequence variants based on their annotation increases power of whole-genome association studies. *Nat Genet, 48*(3), 314-317.

Sveinbjornsson, G., Olafsdottir, E. F., Thorolfsdottir, R. B., Davidsson, O. B., Helgadottir, A., Jonasdottir, A., Jonasdottir, A., Bjornsson, E., Jensson, B. O., Arnadottir, G. A., Kristinsdottir, H., Stephensen, S. S., Oskarsson, G., Gudbjartsson, T., Sigurdsson, E. L., Andersen, K., Danielsen, R., Arnar, D. O., Jonsdottir, I., Thorsteinsdottir, U., Sulem, P., Thorgeirsson, G., Gudbjartsson, D. F., Holm, H., & Stefansson, K. (2018). Variants in *NKX2-5* and *FLNC* Cause Dilated Cardiomyopathy and Sudden Cardiac Death. *Circ Genom Precis Med, 11*(8), e002151.

Sweeney, H. L., & Houdusse, A. (2010). Myosin VI rewrites the rules for myosin motors. *Cell, 141*(4), 573-582.

Tabor, H. K., Risch, N. J., & Myers, R. M. (2002). Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet, 3*(5), 391-397.

Tan, H. L., Glen, E., Topf, A., Hall, D., O'Sullivan, J. J., Sneddon, L., Wren, C., Avery, P., Lewis, R. J., ten Dijke, P., Arthur, H. M., Goodship, J. A., & Keavney, B. D. (2012). Nonsynonymous variants in the *SMAD6* gene predispose to congenital cardiovascular malformation. *Hum Mutat, 33*(4), 720-727.

Tan, L. B., Jalil, J. E., Pick, R., Janicki, J. S., & Weber, K. T. (1991). Cardiac myocyte necrosis induced by angiotensin II. *Circ Res, 69*(5), 1185-1195.

Tao, Y., Zhang, M., Li, L., Bai, Y., Zhou, Y., Moon, A. M., Kaminski, H. J., & Martin, J. F. (2014). Pitx2, an atrial fibrillation predisposition gene, directly regulates ion transport and intercalated disc genes. *Circ Cardiovasc Genet, 7*(1), 23-32.

Teo, L. L., Cannell, T., Babu-Narayan, S. V., Hughes, M., & Mohiaddin, R. H. (2011). Prevalence of associated cardiovascular abnormalities in 500 patients with aortic coarctation referred for cardiovascular magnetic resonance imaging to a tertiary center. *Pediatr Cardiol, 32*(8), 1120-1127.

Thorolfsdottir, R. B., Sveinbjornsson, G., Sulem, P., Helgadottir, A., Gretarsdottir, S., Benonisdottir, S., Magnusdottir, A., Davidsson, O. B., Rajamani, S., Roden, D. M., Darbar, D., Pedersen, T. R., Sabatine, M. S., Jonsdottir, I., Arnar, D. O., Thorsteinsdottir, U., Gudbjartsson,

D. F., Holm, H., & Stefansson, K. (2017). A Missense Variant in *PLEC* Increases Risk of Atrial Fibrillation. *J Am Coll Cardiol, 70*(17), 2157-2168.

Thorolfsdottir, R. B., Sveinbjornsson, G., Sulem, P., Nielsen, J. B., Jonsson, S., Halldorsson, G. H., Melsted, P., Ivarsdottir, E. V., Davidsson, O. B., Kristjansson, R. P., Thorleifsson, G., Helgadottir, A., Gretarsdottir, S., Norddahl, G., Rajamani, S., Torfason, B., Valgardsson, A. S., Sverrisson, J. T., Tragante, V., Holmen, O. L., Asselbergs, F. W., Roden, D. M., Darbar, D., Pedersen, T. R., Sabatine, M. S., Willer, C. J., Lochen, M. L., Halldorsson, B. V., Jonsdottir, I., Hveem, K., Arnar, D. O., Thorsteinsdottir, U., Gudbjartsson, D. F., Holm, H., & Stefansson, K. (2018). Coding variants in *RPL3L* and *MYZAP* increase risk of atrial fibrillation. *Commun Biol, 1*, 68.

Thrall, G., Lane, D., Carroll, D., & Lip, G. Y. (2006). Quality of life in patients with atrial fibrillation: a systematic review. *Am J Med, 119*(5), 448.e441-419.

Tomita-Mitchell, A., Stamm, K. D., Mahnke, D. K., Kim, M. S., Hidestrand, P. M., Liang, H. L., Goetsch, M. A., Hidestrand, M., Simpson, P., Pelech, A. N., Tweddell, J. S., Benson, D. W., Lough, J. W., & Mitchell, M. E. (2016). Impact of *MYH6* variants in hypoplastic left heart syndrome. *Physiol Genomics, 48*(12), 912-921.

Tsai, C. T., Hsieh, C. S., Chang, S. N., Chuang, E. Y., Ueng, K. C., Tsai, C. F., Lin, T. H., Wu, C. K., Lee, J. K., Lin, L. Y., Wang, Y. C., Yu, C. C., Lai, L. P., Tseng, C. D., Hwang, J. J., Chiang, F. T., & Lin, J. L. (2016). Genome-wide screening identifies a KCNIP1 copy number variant as a genetic predictor for atrial fibrillation. *Nat Commun, 7*, 10190.

Tung, P., Levitzky, Y. S., Wang, R., Weng, J., Quan, S. F., Gottlieb, D. J., Rueschman, M., Punjabi, N. M., Mehra, R., Bertisch, S., Benjamin, E. J., & Redline, S. (2017). Obstructive and Central Sleep Apnea and the Risk of Incident Atrial Fibrillation in a Community Cohort of Men and Women. *J Am Heart Assoc, 6*(7).

Tuteja, D., Rafizadeh, S., Timofeyev, V., Wang, S., Zhang, Z., Li, N., Mateo, R. K., Singapuri, A., Young, J. N., Knowlton, A. A., & Chiamvimonvat, N. (2010). Cardiac small conductance Ca2+-activated K+ channel subunits form heteromultimers via the coiled-coil domains in the C termini of the channels. *Circ Res, 107*(7), 851-859.

van der Harst, P., & Verweij, N. (2018). Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. *Circ Res, 122*(3), 433-443.

van der Linde, D., Konings, E. E., Slager, M. A., Witsenburg, M., Helbing, W. A., Takkenberg, J. J., & Roos-Hesselink, J. W. (2011). Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis. *J Am Coll Cardiol, 58*(21), 2241-2247.

van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., & Thermes, C. (2018). The Third Revolution in Sequencing Technology. *Trends Genet, 34*(9), 666-681.

Van Gelder, I. C., Rienstra, M., Crijns, H. J., & Olshansky, B. (2016). Rate control in atrial fibrillation. *Lancet, 388*(10046), 818-828.

Van Raay, T. J., Connors, T. D., Klinger, K. W., Landes, G. M., & Burn, T. C. (1996). A novel ribosomal protein L3-like gene (*RPL3L*) maps to the autosomal dominant polycystic kidney disease gene region. *Genomics, 37*(2), 172-176.

Villa, C. R., Ryan, T. D., Collins, J. J., Taylor, M. D., Lucky, A. W., & Jefferies, J. L. (2015). Left ventricular non-compaction cardiomyopathy associated with epidermolysis bullosa simplex with muscular dystrophy and *PLEC1* mutation. *Neuromuscul Disord, 25*(2), 165-168.

Visscher, P. M., Hill, W. G., & Wray, N. R. (2008). Heritability in the genomics era--concepts and misconceptions. *Nat Rev Genet, 9*(4), 255-266.

Wain, L. V., Shrine, N., Miller, S., Jackson, V. E., Ntalla, I., Soler Artigas, M., Billington, C. K., Kheirallah, A. K., Allen, R., Cook, J. P., Probert, K., Obeidat, M., Bosse, Y., Hao, K., Postma, D. S., Pare, P. D., Ramasamy, A., Magi, R., Mihailov, E., Reinmaa, E., Melen, E., O'Connell, J., Frangou, E., Delaneau, O., Freeman, C., Petkova, D., McCarthy, M., Sayers, I., Deloukas, P., Hubbard, R., Pavord, I., Hansell, A. L., Thomson, N. C., Zeggini, E., Morris, A. P., Marchini, J., Strachan, D. P., Tobin, M. D., & Hall, I. P. (2015). Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir Med, 3*(10), 769-781.

Wakili, R., Voigt, N., Kaab, S., Dobrev, D., & Nattel, S. (2011). Recent advances in the molecular pathophysiology of atrial fibrillation. *J Clin Invest, 121*(8), 2955-2968.

Walfridsson, H., Walfridsson, U., Nielsen, J. C., Johannessen, A., Raatikainen, P., Janzon, M., Levin, L. A., Aronsson, M., Hindricks, G., Kongstad, O., Pehrson, S., Englund, A., Hartikainen, J., Mortensen, L. S., & Hansen, P. S. (2015). Radiofrequency ablation as initial therapy in paroxysmal atrial fibrillation: results on health-related quality of life and symptom burden. The MANTRA-PAF trial. *Europace, 17*(2), 215-221.

Wall, J. D., & Pritchard, J. K. (2003). Haplotype blocks and linkage disequilibrium in the human genome. *Nat Rev Genet, 4*(8), 587-597.

Walter, K., Min, J. L., Huang, J., Crooks, L., Memari, Y., McCarthy, S., Perry, J. R., Xu, C., Futema, M., Lawson, D., Iotchkova, V., Schiffels, S., Hendricks, A. E., Danecek, P., Li, R., Floyd, J., Wain, L. V., Barroso, I., Humphries, S. E., Hurles, M. E., Zeggini, E., Barrett, J. C., Plagnol, V., Richards, J. B., Greenwood, C. M., Timpson, N. J., Durbin, R., & Soranzo, N. (2015). The UK10K project identifies rare variants in health and disease. *Nature, 526*(7571), 82-90.

Wang, J., Klysik, E., Sood, S., Johnson, R. L., Wehrens, X. H., & Martin, J. F. (2010). Pitx2 prevents susceptibility to atrial arrhythmias by inhibiting left-sided pacemaker specification. *Proc Natl Acad Sci U S A, 107*(21), 9753-9758.

Ward, K. E., Pryor, R. W., Matson, J. R., Razook, J. D., Thompson, W. M., & Elkins, R. C. (1990). Delayed detection of coarctation in infancy: implications for timing of newborn follow-up. *Pediatrics, 86*(6), 972-976.

Watson, T., Shantsila, E., & Lip, G. Y. (2009). Mechanisms of thrombogenesis in atrial fibrillation: Virchow's triad revisited. *Lancet, 373*(9658), 155-166.

Welsh, S., Peakman, T., Sheard, S., & Almond, R. (2017). Comparison of DNA quantification methodology used in the DNA extraction protocol for the UK Biobank cohort. *BMC Genomics, 18*(1), 26.

Weng, L. C., Choi, S. H., Klarin, D., Smith, J. G., Loh, P. R., Chaffin, M., Roselli, C., Hulme, O. L., Lunetta, K. L., Dupuis, J., Benjamin, E. J., Newton-Cheh, C., Kathiresan, S., Ellinor, P. T., & Lubitz, S. A. (2017). Heritability of Atrial Fibrillation. *Circ Cardiovasc Genet, 10*(6).

Weng, L. C., Preis, S. R., Hulme, O. L., Larson, M. G., Choi, S. H., Wang, B., Trinquart, L., McManus, D. D., Staerk, L., Lin, H., Lunetta, K. L., Ellinor, P. T., Benjamin, E. J., & Lubitz, S. A. (2018). Genetic Predisposition, Clinical Risk Factor Burden, and Lifetime Risk of Atrial Fibrillation. *Circulation, 137*(10), 1027-1038.

Wiche, G. (1998). Role of plectin in cytoskeleton organization and dynamics. *J Cell Sci, 111 (Pt 17)*, 2477-2486.

Wiche, G., Krepler, R., Artlieb, U., Pytela, R., & Denk, H. (1983). Occurrence and immunolocalization of plectin in tissues. *J Cell Biol, 97*(3), 887-901.

Willer, C. J., Li, Y., & Abecasis, G. R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics, 26*(17), 2190-2191.

Winter, L., & Wiche, G. (2013). The many faces of plectin and plectinopathies: pathology and mechanisms. *Acta Neuropathol, 125*(1), 77-93.

Wolf, L. (1943). Familial auricular fibrillation. *New Engl j Med, 229*, 396-398.

Wolf, P. A., Abbott, R. D., & Kannel, W. B. (1991). Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. *Stroke, 22*(8), 983-988.

Wright, G. E., Nowak, C. A., Goldberg, C. S., Ohye, R. G., Bove, E. L., & Rocchini, A. P. (2005). Extended resection and end-to-end anastomosis for aortic coarctation in infants: results of a tailored surgical approach. *Ann Thorac Surg, 80*(4), 1453-1459.

Wyse, D. G., Van Gelder, I. C., Ellinor, P. T., Go, A. S., Kalman, J. M., Narayan, S. M., Nattel, S., Schotten, U., & Rienstra, M. (2014). Lone atrial fibrillation: does it exist? *J Am Coll Cardiol, 63*(17), 1715-1723.

Xue, A., Wu, Y., Zhu, Z., Zhang, F., Kemper, K. E., Zheng, Z., Yengo, L., Lloyd-Jones, L. R., Sidorenko, J., Wu, Y., McRae, A. F., Visscher, P. M., Zeng, J., & Yang, J. (2018). Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat Commun, 9*(1), 2941.

Xue, S., & Barna, M. (2012). Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nat Rev Mol Cell Biol, 13*(6), 355-369.

Yang, J., Ferreira, T., Morris, A. P., Medland, S. E., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Weedon, M. N., Loos, R. J., Frayling, T. M., McCarthy, M. I., Hirschhorn, J. N., Goddard, M. E., & Visscher, P. M. (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet, 44*(4), 369-375, s361-363.

Yang, J., Zeng, J., Goddard, M. E., Wray, N. R., & Visscher, P. M. (2017). Concepts, estimation and interpretation of SNP-based heritability. *Nat Genet, 49*(9), 1304-1310.

Yang, Y. Q., Zhang, X. L., Wang, X. H., Tan, H. W., Shi, H. F., Fang, W. Y., & Liu, X. (2010). Familial aggregation of lone atrial fibrillation in the Chinese population. *Intern Med, 49*(22), 2385-2391.

Yates, A., Akanni, W., Amode, M. R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., Girón, C. G., Gordon, L., Hourlier, T., Hunt, S. E., Janacek, S. H., Johnson, N., Juettemann, T., Keenan, S., Lavidas, I., Martin, F. J., Maurel, T., McLaren, W., Murphy, D. N., Nag, R., Nuhn, M., Parker, A., Patricio, M., Pignatelli, M., Rahtz, M., Riat, H. S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S. P., Zadissa, A., Birney, E., Harrow, J., Muffato, M., Perry, E., Ruffier, M., Spudich, G., Trevanion, S. J., Cunningham, F., Aken, B. L., Zerbino, D. R., & Flicek, P. (2016). Ensembl 2016. *Nucleic acids research, 44*(D1), D710-D716.

Ye, J., Tucker, N. R., Weng, L. C., Clauss, S., Lubitz, S. A., & Ellinor, P. T. (2016). A Functional Variant Associated with Atrial Fibrillation Regulates PITX2c Expression through TFAP2a. *Am J Hum Genet, 99*(6), 1281-1291.

Yeh, Y. H., Wakili, R., Qi, X. Y., Chartier, D., Boknik, P., Kaab, S., Ravens, U., Coutu, P., Dobrev, D., & Nattel, S. (2008). Calcium-handling abnormalities underlying atrial arrhythmogenesis and contractile dysfunction in dogs with congestive heart failure. *Circ Arrhythm Electrophysiol, 1*(2), 93-102.

Yuan, F., Qiu, X. B., Li, R. G., Qu, X. K., Wang, J., Xu, Y. J., Liu, X., Fang, W. Y., Yang, Y. Q., & Liao, D. N. (2015). A novel *NKX2-5* loss-of-function mutation predisposes to familial dilated cardiomyopathy and arrhythmias. *Int J Mol Med, 35*(2), 478-486.

Zhang, X., Azhar, G., Chai, J., Sheridan, P., Nagano, K., Brown, T., Yang, J., Khrapko, K., Borras, A. M., Lawitts, J., Misra, R. P., & Wei, J. Y. (2001). Cardiomyopathy in transgenic mice with cardiac-specific overexpression of serum response factor. *Am J Physiol Heart Circ Physiol, 280*(4), H1782-1792.

Zhou, S. H., Helfenbein, E. D., Lindauer, J. M., Gregg, R. E., & Feild, D. Q. (2009). Philips QT interval measurement algorithms for diagnostic, ambulatory, and patient monitoring ECG applications. *Ann Noninvasive Electrocardiol, 14 Suppl 1*, S3-8.

Zhu, W., Yuan, P., Shen, Y., Wan, R., & Hong, K. (2016). Association of smoking with the risk of incident atrial fibrillation: A meta-analysis of prospective studies. *Int J Cardiol, 218*, 259-266.

Original publications

Paper I

JOURNAL OF THE AMERICAN COLLEGE OF CARDIOLOGY
© 2017 BY THE AMERICAN COLLEGE OF CARDIOLOGY FOUNDATION
PUBLISHED BY ELSEVIER

VOL. 70, NO. 17, 2017
ISSN 0735-1097/$36.00
http://dx.doi.org/10.1016/j.jacc.2017.09.005

A Missense Variant in *PLEC* Increases Risk of Atrial Fibrillation

Rosa B. Thorolfsdottir, MD,[a] Gardar Sveinbjornsson, MSc,[a] Patrick Sulem, MD, PhD,[a] Anna Helgadottir, MD, PhD,[a] Solveig Gretarsdottir, PhD,[a] Stefania Benonisdottir, MSc,[a] Audur Magnusdottir, PhD,[a] Olafur B. Davidsson, MSc,[a] Sridharan Rajamani, PhD,[a] Dan M. Roden, MD,[b] Dawood Darbar, MD,[c] Terje R. Pedersen, MD, PhD,[d] Marc S. Sabatine, MD, MPH,[e] Ingileif Jonsdottir, PhD,[a,f,g] David O. Arnar, MD, PhD, MPH,[a,f,h] Unnur Thorsteinsdottir, PhD,[a,f] Daniel F. Gudbjartsson, PhD,[a,i] Hilma Holm, MD,[a] Kari Stefansson, MD, PhD[a,f]

ABSTRACT

BACKGROUND Genome-wide association studies (GWAS) have yielded variants at >30 loci that associate with atrial fibrillation (AF), including rare coding mutations in the sarcomere genes *MYH6* and *MYL4*.

OBJECTIVES The aim of this study was to search for novel AF associations and in doing so gain insights into the mechanisms whereby variants affect AF risk, using electrocardiogram (ECG) measurements.

METHODS The authors performed a GWAS of 14,255 AF cases and 374,939 controls, using whole-genome sequence data from the Icelandic population, and tested novel signals in 2,002 non-Icelandic cases and 12,324 controls. They then tested the AF variants for effect on cardiac electrical function by using measurements in 289,297 ECGs from 62,974 individuals.

RESULTS The authors discovered 2 novel AF variants, the intergenic variant rs72700114, between the genes *LINC01142* and *METTL11B* (risk allele frequency = 8.1%; odds ratio [OR]: 1.26; $p = 3.1 \times 10^{-18}$), and the missense variant p.Gly4098Ser in *PLEC* (frequency = 1.2%; OR: 1.55; $p = 8.0 \times 10^{-10}$), encoding plectin, a cytoskeletal cross-linking protein that contributes to integrity of cardiac tissue. The authors also confirmed 29 reported variants. p.Gly4098Ser in *PLEC* significantly affects various ECG measurements in the absence of AF. Other AF variants have diverse effects on the conduction system, ranging from none to extensive.

CONCLUSIONS The discovery of a missense variant in *PLEC* affecting AF combined with recent discoveries of variants in the sarcomere genes *MYH6* and *MYL4* points to an important role of myocardial structure in the pathogenesis of the disease. The diverse associations between AF variants and ECG measurements suggest fundamentally different categories of mechanisms contributing to the development of AF. (J Am Coll Cardiol 2017;70:2157–68)

Listen to this manuscript's audio summary by *JACC* Editor-in-Chief Dr. Valentin Fuster.

From the [a]deCODE genetics/Amgen, Inc., Reykjavik, Iceland; [b]Departments of Medicine, Pharmacology, and Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee; [c]Division of Cardiology, Department of Medicine, University of Illinois at Chicago, Chicago, Illinois; [d]Center For Preventive Medicine, Oslo University Hospital and Medical Faculty, University of Oslo, Oslo, Norway; [e]TIMI Study Group, Division of Cardiovascular Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; [f]Faculty of Medicine, University of Iceland, Reykjavik, Iceland; [g]Department of Immunology, Landspitali University Hospital, Reykjavik, Iceland; [h]Department of Medicine, Landspitali University Hospital, Reykjavik, Iceland; and the [i]School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland. Dr. Thorolfsdottir, Mr. Sveinbjornsson, Dr. Sulem, Dr. Helgadottir, Dr. Gretarsdottir, Ms. Benonisdottir, Dr. Magnusdottir, Mr. Davidsson, Dr. Rajamani, Dr. Jonsdottir, Dr. Arnar, Dr. Thorsteinsdottir, Dr. Gudbjartsson, Dr. Holm, and Dr. Stefansson are employees of deCODE genetics/Amgen, Inc. Dr. Pedersen has received consulting and/or speakers honoraria from Amgen, Sanofi, and Merck. Dr. Sabatine has received research grant support through Brigham and Women's Hospital from Amgen, AstraZeneca, Daiichi-Sankyo, Eisai, GlaxoSmithKline, Intarcia, Janssen, MedImmune, Merck, Novartis, Pfizer, Poxel, and Takeda; and is a consultant for Amgen, CVS Caremark, Esperion, Intarcia, Ionis, Janssen, MedImmune, and Merck. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

Manuscript received May 11, 2017; revised manuscript received August 7, 2017, accepted September 5, 2017.

JACC VOL. 70, NO. 17, 2017
OCTOBER 24, 2017:2157–68

ABBREVIATIONS AND ACRONYMS

AF = atrial fibrillation

AVN = atrioventricular node

EBS = epidermolysis bullosa simplex

ECG = electrocardiogram

LD = linkage disequilibrium

SNP = single nucleotide polymorphism

SSS = sick sinus syndrome

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia in humans and a significant cause of morbidity and mortality (1-3). More than 30 rare sequence variants with large effects have been identified in familial AF, including coding variants in ion channels, signaling molecules, structural proteins, and transcription factors (4). These mutations explain only a small proportion of AF cases but can provide important insights into the mechanisms of AF. For example, mutations in the potassium channel gene *KCNQ1*, the first gene linked to familial AF, are thought to cause AF by reducing the atrial action potential duration and the effective refractory period (5). Over the last decade, genome-wide association studies (GWAS) have yielded variants at >30 loci that associate with nonfamilial AF as a primary or secondary trait, with the most notable association being at *PITX2*, a transcription factor gene that affects cardiac development (6). However, most of the AF variants discovered through GWAS are in noncoding regions, and the mechanisms by which they affect AF remains to be elucidated (6-19). Recently, GWAS in Iceland based on whole-genome sequencing have associated coding variants in the sarcomere genes *MYH6* and *MYL4* with AF (13,14).

SEE PAGE 2169

The use of electrocardiogram (ECG) measurements as intermediate traits of arrhythmia has proven valuable and has led to the discovery of variants that affect both the cardiac electrical function and the risk of arrhythmia, including variants that associate with both the PR interval and risk of AF (15-17). Assessing the effects of both coding and noncoding AF variants on detailed ECG traits could give valuable information about their modes of action and the pathophysiology of AF.

Here we describe a GWAS of AF in 14,255 Icelandic cases and 374,939 controls based on whole-genome sequencing and testing of the AF variants for association with closely related phenotypes including sick sinus syndrome (SSS) and ischemic stroke. We then tested the resulting AF variants and those previously described for associations with ECG traits.

METHODS

Detailed methods are available in the Online Appendix. The study was approved by the Icelandic Data Protection Authority and the National Bioethics Committee of Iceland. The study complies with tenets of the Declaration of Helsinki.

STUDY POPULATIONS. The Icelandic AF population consisted of all patients with the diagnosis of AF (International Classification of Diseases-10 [ICD-10] code I.48 and ICD-9 code 427.3) at Landspitali, the National University Hospital, in Reykjavik and Akureyri Hospital (the 2 largest hospitals in Iceland) from 1987 to 2015. All AF cases (N = 14,255) were included. Controls consisted of 374,939 Icelanders recruited through different genetic research projects at deCODE genetics (Reykjavik, Iceland). Individuals in the AF cohort were excluded from the control group (see Online Table 1 for subject and control characteristics). We followed up the novel AF variants in AF sample sets from the FOURIER (Further Cardiovascular Outcomes Research with PCSK9 Inhibition in Subjects with Elevated Risk) trial (20) (1,238 cases and 11,562 controls) and the Vanderbilt AF Registry (764 cases and 762 controls) (Online Methods). The deCODE genetics phenotype database contains extensive medical information on various diseases and traits. This includes sample sets from a pacemaker population (n = 3,578), SSS subjects (n = 3,310), ischemic stroke subjects (n = 5,626), and cardioembolic stroke subjects (n = 1,369). The controls used in the various case control analyses of this study consisted of individuals randomly drawn from the Icelandic genealogical database and individuals from other genetic studies at deCODE genetics.

ELECTROCARDIOGRAM DATA. ECG data were collected from Landspitali, the National University Hospital, in Reykjavik and included all ECGs obtained and digitally stored from 1998 to 2015, a total of 434,000 ECGs from 88,217 individuals. These were ECGs obtained from all hospital departments from both in- and outpatients. For the main analysis, sinus rhythm (heart rate 50 to 100 beats/min) ECGs from individuals without the diagnosis of AF were used. This was done to assess the effect of the AF variants on ECG measurements (e.g., P-wave morphology, PR intervals, and so forth) and thus cardiac electrical function in the absence of AF. Individuals with pacemakers also were excluded. This resulted in 289,297 sinus rhythm ECGs from 62,974 individuals. The analysis was also done using all ECGs regardless of rhythm and history of AF (Online Appendix).

GENERATION OF GENOTYPE DATA. This study was based on whole-genome sequence data from 15,220 Icelanders participating in various disease projects at deCODE genetics. The sequencing was done using standard TruSeq methodology (Illumina, San Diego, California) to a mean depth of 35 ± 8 x, as previously described (21). Genotypes of the 32.5 million sequence variants identified through sequencing (single nucleotide polymorphisms [SNPs] and indels) were

TABLE 1 Variants Associating Genome-Wide Significantly With AF (N = 13,471)

Reference SNP ID	Chr	Pos (hg38)	Closest Gene(s)	Risk Allele/ Other	RAF, %	Coding Effect	OR (95% CI)	p Value	p Value Threshold	Reported SNP ID (Ref. #)	r^2 With Reported SNP
Novel signals											
rs72700114	1q24	170,224,684	*METTL11B*, *LINC01142*	C/G	8.1	Intergenic	1.22 (1.15–1.29)	7.0×10^{-12}	2.3×10^{-9}	NA	NA
rs373243633	8q24	143,917,940	*PLEC*	T/C	1.2	Missense	1.55 (1.35–1.78)	8.0×10^{-10}	5.1×10^{-8}	NA	NA
Variants reported to associate with AF											
rs6843082	4q25	110,796,911	*PITX2*	G/A	20.1	Intergenic	1.43 (1.38–1.49)	1.6×10^{-73}	2.3×10^{-9}	rs6843082 (10)	1.00
rs387906656*	14q11	23,396,970	*MYH6*	A/G	0.3	Missense	2.43 (1.94–3.04)	1.1×10^{-14}	5.1×10^{-8}	rs387906656 (13)	1.00
rs34245846	1q21	154,858,667	*KCNN3*	G/A	32.7	Intronic	1.15 (1.11–1.19)	1.2×10^{-14}	7.9×10^{-10}	rs13376333 (10)	0.52
rs2359171	16q22	73,019,123	*ZFHX3*	A/T	18.8	Intronic	1.17 (1.13–1.22)	1.5×10^{-14}	2.3×10^{-9}	rs2106261 (8)	1.00
rs148321568	10q22	73,668,538	*SYNPO2L*, *MYOZ1*	A/T	85.2	Intergenic	1.19 (1.14–1.25)	3.7×10^{-13}	7.9×10^{-10}	rs10824026 (9)	0.95
rs11598047	10q24	103,582,915	*NEURL1*	G/A	17.9	Intronic	1.15 (1.10–1.19)	1.5×10^{-10}	2.3×10^{-9}	rs12415501 (11)	0.91
rs651386	1q24	170,622,169	*PRRX1*	A/T	57.4	Intergenic	1.11 (1.08–1.15)	1.9×10^{-10}	7.9×10^{-10}	rs3903239 (9)	0.61
rs1572226	6q22	118,313,916	*SLC35F1*, *PLN*	C/G	42.4	Intronic	1.11 (1.08–1.15)	3.1×10^{-10}	2.3×10^{-9}	rs4946333 (18)	0.56

*Locus previously associating with AF but not genome-wide significantly.

AF = atrial fibrillation; Chr = chromosome; CI = confidence interval; OR = odds ratio; NA = not applicable; Pos = position; RAF = risk allele frequency; SNP = single nucleotid polymorphism.

then imputed into 151,677 Icelanders chip typed using SNP chips (Illumina) and their close relatives (familial imputation). Imputation refers to the statistical inference of unobserved genotypes using known haplotypes in the population (see Online Methods for further details).

STATISTICAL ANALYSIS. Logistic regression was used to test for association between SNPs and AF and other traits, treating disease status as the response and allele counts from direct genotyping or expected genotype counts from imputation as covariates. To account for inflation in test statistics due to cryptic relatedness and stratification, we applied the method of linkage disequilibrium (LD) score regression (22). The estimated correction factor for AF based on LD score regression was 1.38 for the additive model. The threshold for genome-wide significance was corrected for multiple testing with a weighted Bonferroni adjustment, using as weights the enrichment of variant classes with predicted functional impact among association signals (see Online Methods for significance thresholds for specific groups of variants) (23). We tested 31 AF variants for association with 122 ECG measurements (Online Table 2), using linear regression, treating the ECG measurement as the response and the genotype as the covariate. The ECG measurements were adjusted for sex, year of birth, and age at measurement and were subsequently standardized to have a normal distribution. For individuals with multiple ECG measurements, the mean standardized value was used. The Benjamini-Hochberg false discovery rate procedure controlling the false discovery rate at 0.05 at each marker was used to account for multiple testing.

RESULTS

We tested ~32.5 million sequence variants for association with AF in 14,255 Icelandic cases and 374,939 controls. Variants were identified by whole-genome sequencing of 15,220 Icelanders and imputed into 151,677 long-range phased individuals and their relatives (21). We found 2 novel AF associations, an intergenic variant at the genes *METTL11B* and *LINC01142*, and a missense variant in the gene *PLEC*. We also replicated 29 loci previously associated with AF in GWAS, either as a primary or secondary trait, following association with SSS, heart rate, or PR interval (**Tables 1 and 2**, Online Figure 1). Variants at 10 loci reached genome-wide significance (**Table 1**). One of those had only been previously associated with AF as a secondary trait, a missense variant in *MYH6* reported in SSS (13) (see Online Figure 2 for loci plots). Two of the previously reported variants (rs12044963 at *KCND3* and rs2047036 at *SH3PXD2A*) were previously reported in Japanese subjects and nominally associated with AF in Europeans (19). Here we replicated the previously reported variants and provide further evidence that these are true AF variants. Additional variants reported in Japanese subjects were either not replicated (**Table 2**) or are very rare or nonexistent in the Icelandic population (Online Table 3).

A MISSENSE VARIANT IN *PLEC* INCREASES RISK OF AF. The novel AF association with *PLEC* on chromosome 8 is driven by a low-frequency missense variant p.Gly4098Ser (NP_958782.1; frequency = 1.2%) that associates with AF with an odds ratio (OR) of 1.55

TABLE 2 Replication of Loci Previously Reported to Associate With AF (N = 13,471)

Ref SNP (ID) (Ref. #)	Chr	Pos (hg38)	Closest Gene(s)	Risk Allele/ Other	RAF, %	Coding Effect	OR (95% CI)	p Value*
Variants reported to associate genome-wide significantly with AF								
rs3807989 (9)	7q31	116,546,187	*CAV1*	G/A	59.5	Intronic	1.11 (1.07–1.14)	**2.5×10^{-9}**
rs10821415 (9)	9q22	94,951,177	*C9orf3*	A/C	38.5	Intronic	1.10 (1.06–1.14)	**3.8×10^{-8}**
rs35176054 (18)	10q24	103,720,629	*SH3PXD2A*	A/T	13.6	Intronic	1.14 (1.08–1.19)	**9.3×10^{-8}**
rs72700118 (18)	1q24	170,225,682	*METTL11B, LINC01142*	A/C	13.4	Intergenic	1.12 (1.07–1.17)	**$2,9 \times 10^{-6}$**
rs10507248 (11)	12q24	114,359,288	*TBX5*	T/G	70.2	Intronic	1.09 (1.05–1.13)	**3.0×10^{-6}**
rs7164883 (9)	15q24	73,359,833	*HCN4*	G/A	14.7	Intronic	1.12 (1.07–1.17)	**1.1×10^{-6}**
rs7508 (18)	8p22	18,056,461	*ASAH1*	A/G	71.5	3′ UTR	1.09 (1.05–1.13)	**1.2×10^{-5}**
rs1152591 (9)	14q23	64,214,130	*SYNE2*	A/G	51.0	Upstream	1.07 (1.04–1.11)	**3.5×10^{-5}**
rs13216675 (11)	6q22	122,131,183	*GJA1*	T/C	66.2	Intergenic	1.08 (1.04–1.11)	**3.5×10^{-5}**
rs11047543 (18)	12p12	24,635,405	*SOX5*	G/A	86.7	Intergenic	1.11 (1.05–1.16)	**5.3×10^{-5}**
rs2540949 (18)	2p14	65,057,097	*CEP68*	A/T	59.3	Intronic	1.07 (1.03–1.10)	**1.5×10^{-4}**
rs12044963 (19)	1p13	111,849,738	*KCND3*	T/G	9.4	Intronic	1.11 (1.05–1.17)	**2.1×10^{-4}**
rs4642101 (11)	3p25	12,800,724	*CAND2*	G/T	65.8	Intronic	1.07 (1.03–1.10)	**3.1×10^{-4}**
rs6800541 (18)	3p22	38,733,341	*SCN10A*	T/C	63.6	Intronic	1.06 (1.03–1.10)	**4.3×10^{-4}**
rs3771537 (18)	2p13	69,811,660	*ANXA4*	A/C	53.1	Intronic	1.06 (1.03–1.10)	**4.6×10^{-4}**
rs75190942 (18)	11q24	128,894,676	*KCNJ5*	A/C	10.6	Downstream	1.09 (1.03–1.15)	**0.0016**
rs337711 (18)	5q22	114,412,874	*KCNN2*	T/C	34.1	Intronic	1.05 (1.01–1.08)	**0.0071**
rs2047036 (19)	10q24	103,717,405	*SH3PXD2A*	C/T	40.1	Intronic	1.04 (1.01–1.08)	**0.012**
rs2288327 (18)	2q31	178,546,938	*TTN*	G/A	14.0	Intronic	1.06 (1.01–1.11)	**0.017**
rs2967791 (18)	5q31	137,677,417	*KLHL3*	T/C	56.7	Intronic	1.04 (1.00–1.07)	**0.025**
rs7698692 (19)	4q34	173,682,953	*HAND2*	G/A	2.1	Intergenic	1.10 (0.99–1.23)	0.089
rs17461925 (19)	1q32	203,057,463	*PPFIA4*	A/G	66.2	Intronic	1.01 (0.97–1.04)	0.69
rs6490029 (11)	12q24	111,260,653	*CUX2*	A/G	29.8	Intronic	1.00 (0.98–1.02)	1.00
Variants reported to associate with AF as a secondary trait, following association with heart rate and/or PR interval								
rs6882776 (15)	5q35	173,237,160	*NKX2-5*	G/A	71.8	Upstream	1.09 (1.05–1.13)	**8.1×10^{-6}**
rs11708996 (16)	3p22	38,592,432	*SCN5A*	G/C	88.2	Intronic	1.04 (0.99–1.09)	0.14

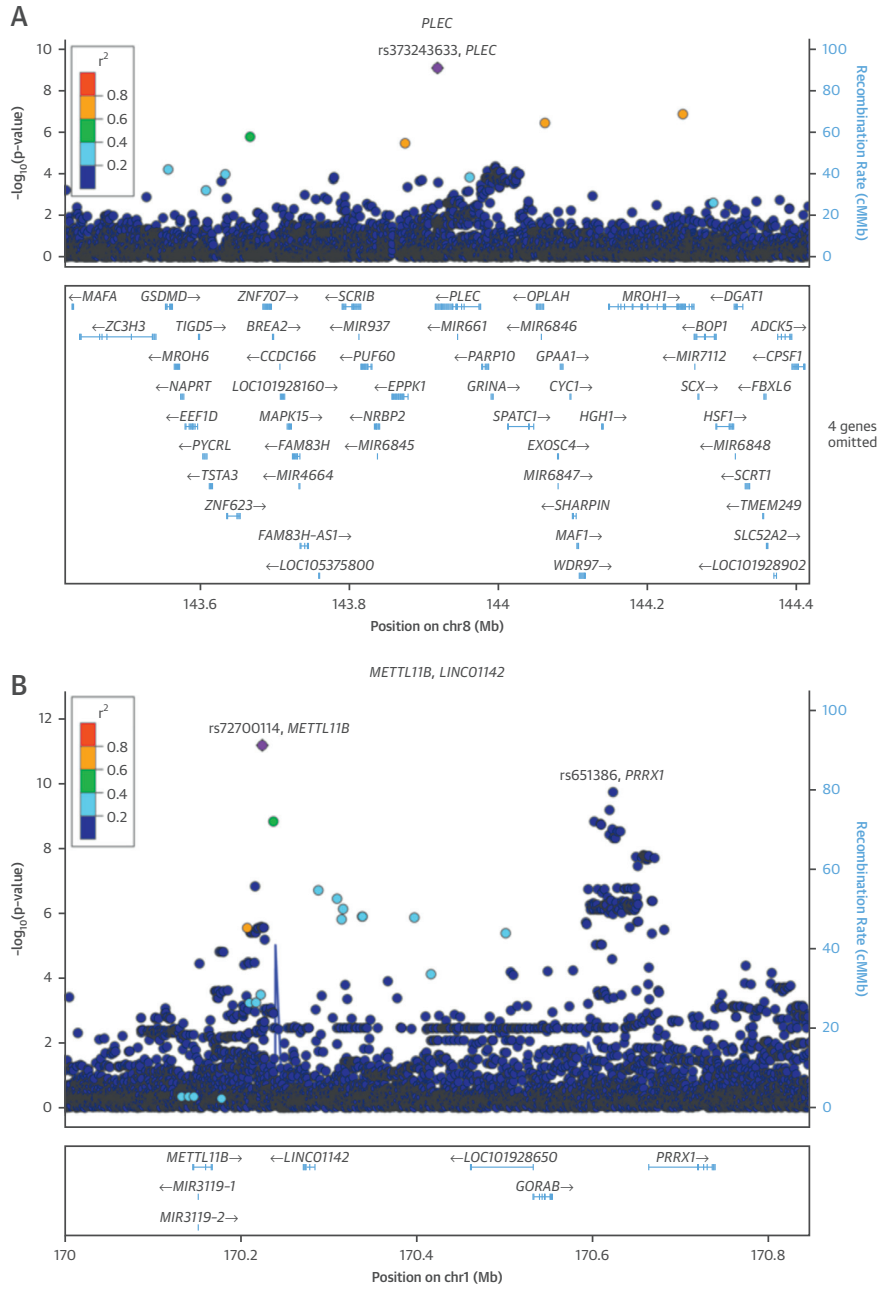
*Significant p values (<0.05) are in **bold**.
Abbreviations as in Table 1.

(95% confidence interval [CI]: 1.35 to 1.78; $p = 8.0 \times 10^{-10}$); (**Figure 1A**). It is the only variant in the region reaching genome-wide significance. Conditional analysis did not reveal additional signals in the region, and no other coding variant in *PLEC* associates independently with AF (Online Table 4). *PLEC* encodes plectin, a very large (>500 kDa) multidomain cytoskeletal linking protein with a role in maintaining tissue integrity in skin, striated muscle, and heart (24). It is expressed in many tissues, including the heart, with equal expression in atria and ventricles (25).

p.Gly4098Ser is a missense variant (C→T at position 143,917,940) in exon 32/32 of *PLEC* and is potentially damaging (Polymorphism Phenotyping v2 [PolyPhen-2]: 1; Sorting Intolerant From Tolerant [SIFT]: 0.02) (26,27). Generally, the *PLEC* gene is tolerant to missense mutations (constrained metric z-score: −0.38 [28]). However the region in question is conserved and has low tolerance to mutations according to Genomic Evolutionary Rate Profiling (GERP) score (4.2 [29]) and sub-region Residual Variation Intolerance Score (subRVIS) (<35th percentile [30]). Exon 32 encodes the C-terminal intermediate filament binding domain of plectin, and the variant results in a glycine to serine substitution on the surface, close to a suggested interaction site for intermediate filaments (Online Figure 3) (31).

The carrier frequency of p.Gly4098Ser among genotyped individuals was 3.08% (269 of 8,740 subjects) in cases and 2.09% (2,752 of 131,941 subjects) in controls. Seventeen of the 151,677 genotyped study participants (currently at a mean 46 years of age [range: 11 to 80 years]) were homozygous for p.Gly4098Ser. No homozygote had been diagnosed with AF according to our records. P.Gly4098Ser did not associate with AF ($p = 0.37$) or ECG traits (lowest p value = 0.0069 in T amplitude lead V_3; beta = −1.25) under a recessive mode of inheritance. The fact that none of the 17 homozygotes had been diagnosed with AF was unexpected but does not constitute a significant deviation from the additive model ($p = 0.093$). Furthermore, it is possible that some of them have AF that has not been diagnosed or who may develop AF later in life.

FIGURE 1 Regional Plots of Novel AF Loci



Regional plots of **(A)** the *PLEC* locus on chromosome 8q24 and **(B)** a signal, rs72700114, between *METTL11B* and *LINC01142* on chromosome 1q24, and a previously reported signal 400 kb away, rs651386, close to *PRRX1*. The plots depict association results (p values) with atrial fibrillation (N = 13,471). For each plot, the y-axis shows the $-\log_{10}$ p values, and the x-axis shows the genomic position (hg38). The leading variants of independent signals are labeled as **diamonds** and are **purple**. Other variants are colored according to correlation (r^2) with the leading marker.

The p.Gly4098Ser mutation is exceedingly rare outside Iceland. The variant was only found in 13 of the 140,842 individuals in the Genome Aggregation Database (12 Europeans, 1 South Asian) (28). The Icelandic population is a founder population, in that a small number of ancestors accounts for a relatively large proportion of the population. Hence, variants that are very rare in more outbred populations, like p.Gly4098Ser, may be more common in Iceland. Acknowledging the limited power to detect associations, we genotyped the variant in 2 AF sample sets, from the Vanderbilt AF registry (764 cases and 762 controls) and the FOURIER trial (1,238 cases and 11,562 controls). We found 3 carriers, and all had AF.

The second new AF variant, rs72700114 on chromosome 1q24 (risk allele frequency = 8.1%), is located between the genes *LINC01142* and *METTL11B* and is associated with AF with an OR of 1.22 ($p = 7.0 \times 10^{-12}$; 95% CI: 1.15 to 1.29) in the Icelandic GWAS. The association replicated well in the non-Icelandic AF sample sets and the combined OR was 1.26 (95% CI: 1.19 to 1.32; p value of 3.1×10^{-18}) (Online Table 5). Variant rs72700114 is located 400 kb from a previously reported AF locus, represented by rs651386, close to *PRRX1* (**Figure 1B**) (9). The 2 variants are not strongly correlated ($D' = 0.38$; $r^2 = 0.009$), and when conditioned on each other, they both associate with AF (Online Table 6). In fact, rs72700114 is the second variant to be associated with AF secondarily with this locus. The recently reported variant rs72700118 (18) is located only 1 kb away from rs72700114, but the 2 variants are not in LD ($D' = 1$; $r^2 = 0.014$). None of the 3 variants associates with the expression of nearby genes in our samples from blood (2,528 samples) or adipose tissue (686 samples).

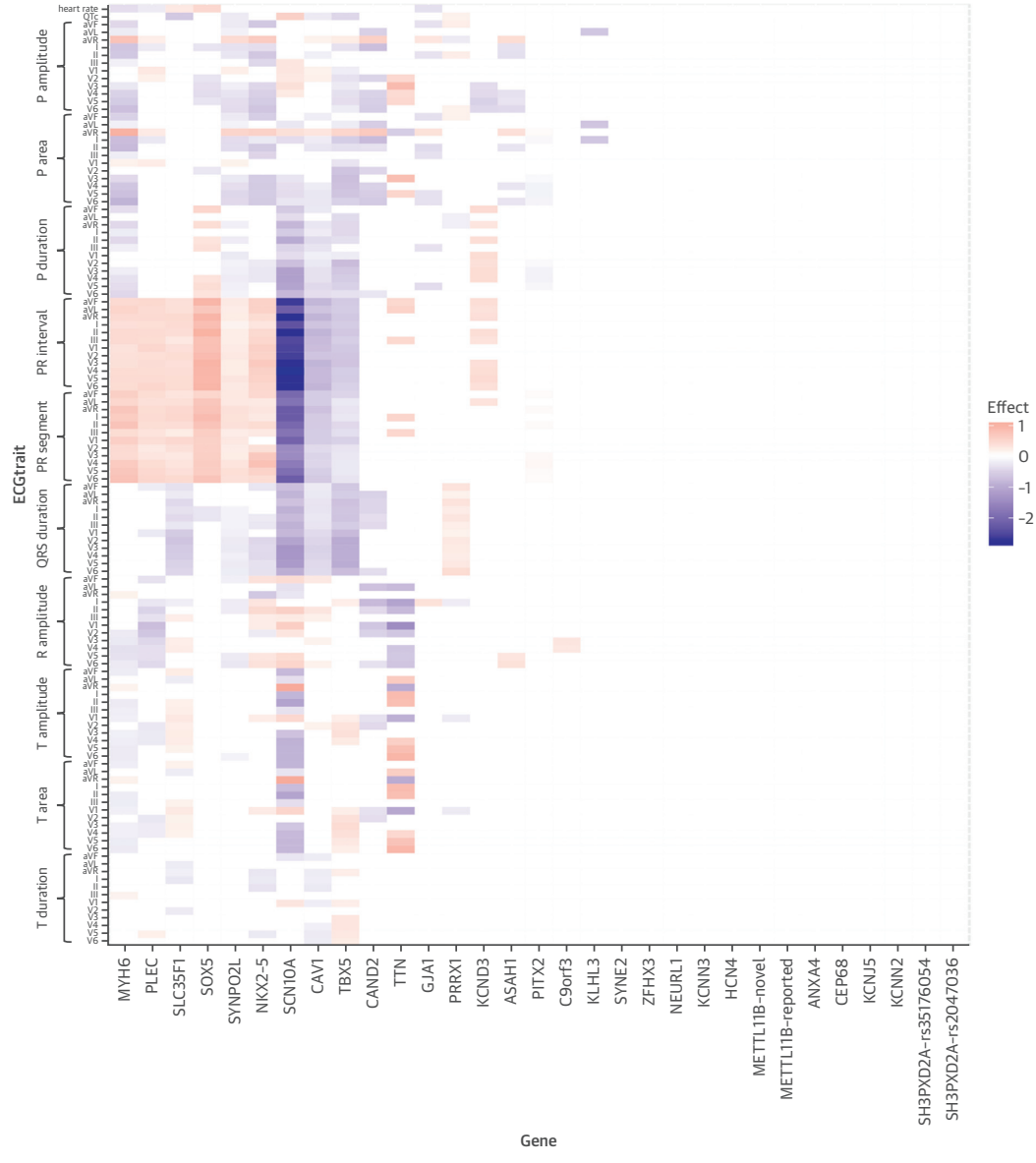
ASSOCIATION OF AF VARIANTS WITH ECG MEASUREMENTS IN THE ABSENCE OF AF. We tested the 2 novel and 29 previously reported AF variants for association with ECG traits that reflect electrophysiological functions of the heart in sinus rhythm (**Figure 2,** Online Table 7, Online Figure 4). A total of 289,297 sinus rhythm ECGs from 62,974 individuals without the diagnosis of AF were included in the analysis. We tested all variants for association with 122 ECG variables, some of them correlated (Online Table 2). Independent of AF diagnoses, the novel p.Gly4098Ser variant in *PLEC* associates with many ECG measurements. The variant affects P-wave amplitude and area, prolongs the PR segment representing atrioventricular node (AVN) conduction, and lowers the R wave amplitude. The coding AF variant in *MYH6*, another structural gene, also associated with SSS and a 6-fold risk of pacemaker placement, has a similar general effect on the cardiac electrical function, albeit a relatively stronger effect on the atria and AVN than on the ventricles.

Generally, the effects of the AF variants on ECG measurements range from none to extensive and vary in direction. For example, AF risk alleles can prolong, shorten, or have no detectable effect on the PR interval, an established predictor of AF development (32). Similarly, there is no clear relationship between the AF effect size and the effect on conduction. For example, the variant in *SCN10A* has extensive and strong ECG effects but small AF effect compared to the strongest common AF variant in *PITX2*. Furthermore, the AF variant with moderate effect in the potassium channel gene *KCNN3* has no association with ECG measurements in the absence of AF. The same applies to several other variants, for example, the novel intergenic AF variants at *LINC01142* /*METTL11B*.

TESTING FOR ASSOCIATION WITH SECONDARY PHENOTYPES. To further explore the p.Gly4098Ser *PLEC* variant, we tested it for association with all phenotypes in deCODE genetics phenotype/genotype database. Apart from AF, the most significant disease associations were with AF-related traits, SSS (OR: 1.64; 95% CI: 1.31 to 2.05; $p = 1.7 \times 10^{-5}$), pacemaker implantation (OR: 1.54; 95% CI: 1.24 to 1.92; $p = 9.5 \times 10^{-5}$), ischemic stroke (IS) (OR: 1.22; 95% CI: 1.01 to 1.47; $p = 0.035$) and the IS subphenotype cardioembolic stroke (OR: 1.53; 95% CI: 1.09 to 2.14; $p = 0.013$) (Online Table 8). Sinus node dysfunction is a common indication for pacemaker implantation and is frequently associated with AF through a complex causal relationship (33), and stroke is a well-known consequence of AF (3). To compare all AF variants, we tested them for associations with these AF-related traits, SSS, pacemaker implantation, IS, and cardioembolic stroke and then plotted the AF risk of each variant relative to the effect on the risk of the AF-related trait (**Figure 3**). The coding AF variant in *MYH6* that was originally discovered through its association to high risk of SSS (13) stood out with substantially greater risk of SSS and pacemaker implantation than predicted from its effect on AF risk. For all other variants, their effects on the AF-related traits were consistent with being proportional to their effects on AF.

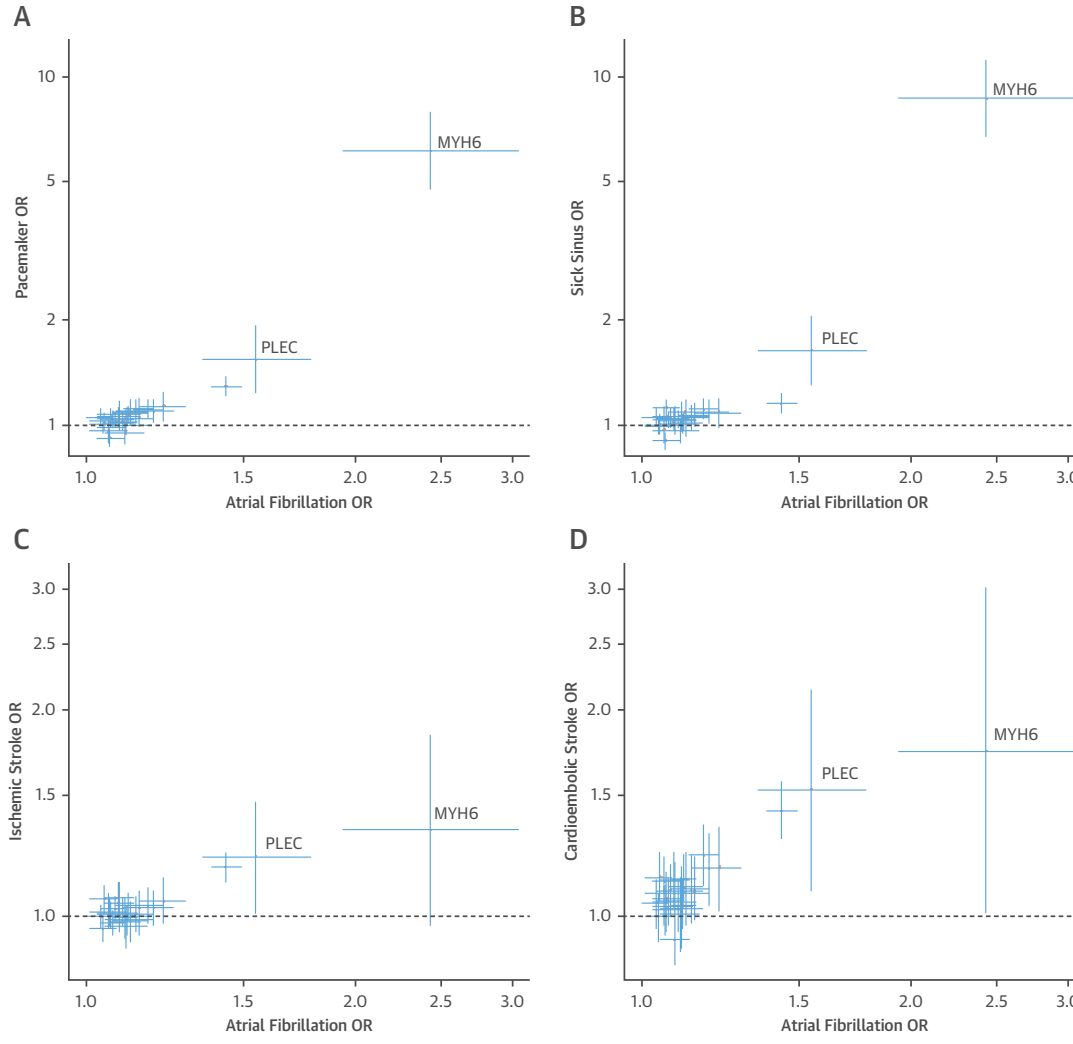
Mutations in *PLEC* have been linked to phenotypes involving skin, muscle, and heart. The association of p.Gly4098Ser in *PLEC*, with available relevant phenotypes including risk factors of AF and some traits described in plectinopathies, are listed in Online Table 8. The only other significant association

FIGURE 2 **Heatmap Shows the Effects of AF Variants on ECG Traits in Sinus Rhythm ECGs, Excluding AF Cases**



Sinus rhythm ECG measurements were available for 62,974 individuals without diagnosis of AF. Each column shows the estimated effect of the risk allele of an AF variant on various ECG traits. The effect of each variant, annotated with the corresponding gene name, is scaled with the $\log_{10}$-AF odds ratio. **Orange** represents a positive effect on the ECG variable, and **blue** represents negative effect. The effect is shown only for significant associations after adjusting for multiple testing with a false discovery rate procedure for each variant. Nonsignificant associations are **white** in the heatmap. AF = atrial fibrillation; ECG = electrocardiogram.

FIGURE 3 AF Effects of AF-Associated Variants on Risk of Pacemaker, SSS, and Stroke



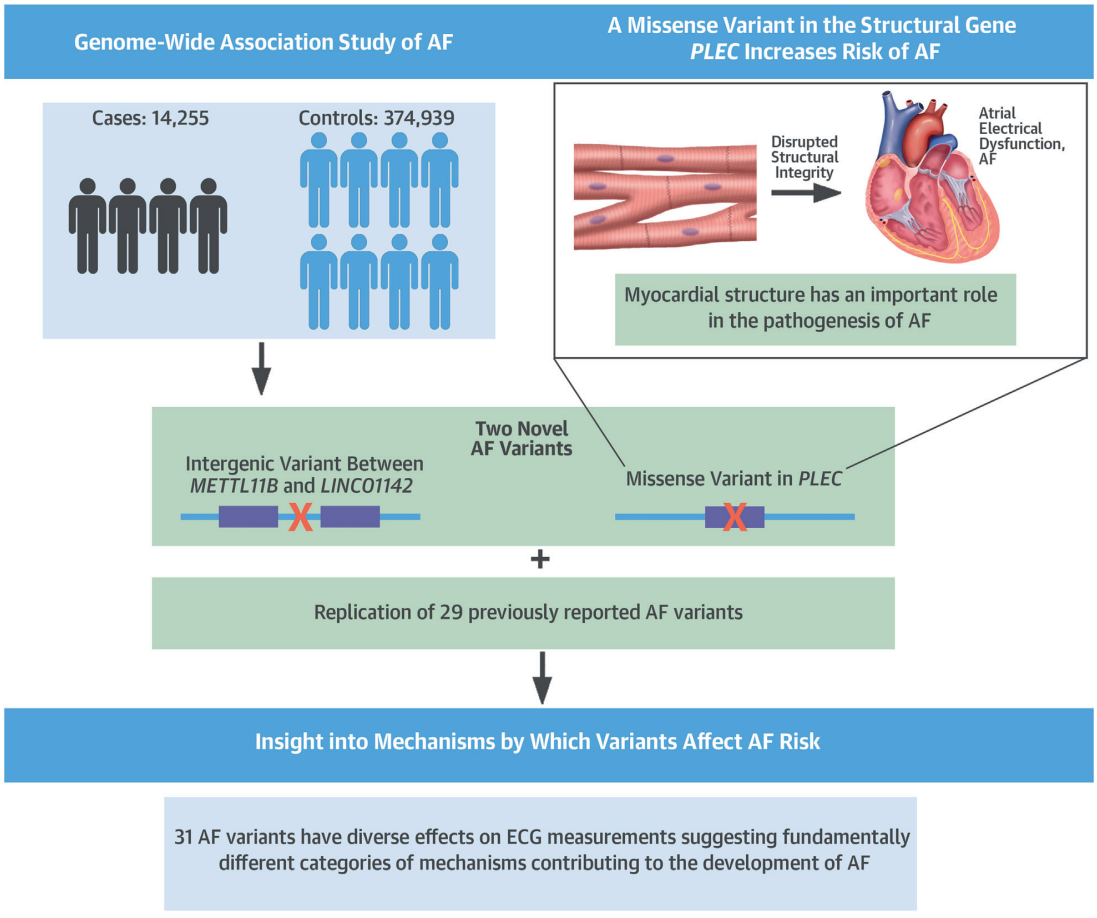
AF effects of AF-associated variants plotted against the effect of **(A)** pacemaker implantation (n = 3,578), **(B)** SSS (n = 3,310), **(C)** ischemic stroke (n = 5,626), and **(D)** cardioembolic stroke (n = 1,369), respectively. A logarithmic scale is used for the axis. AF = atrial fibrillation; OR = odds ratio; SSS = sick sinus syndrome.

($p < 0.05/16 = 0.003$) among these was with lower levels of creatine kinase (beta = −0.09; $p = 2.58 \times 10^{-4}$). Traits described in plectinopathies were not observed from a review of available data for the 17 homozygous carriers of p.Gly4098Ser. Apart from associating with AF and AF-related traits, the other novel AF variant at *LINC01142/METTL11B* was associated with risk of heart failure (OR: 1.10; 95% CI: 1.04 to 1.17; p = 0.0016) (Online Table 9).

DISCUSSION

We performed a large association study based on whole-genome sequencing and identified 2 novel AF variants, a common intergenic variant on chromosome 1q24 and a low-frequency missense variant in the gene *PLEC* (**Central Illustration**). The *PLEC* variant is associated with a 55% increased risk of AF and a 64% increased risk of SSS. It is the third coding

CENTRAL ILLUSTRATION **GWAS of AF and Insight Into Mechanisms by Which Variants Affect Risk for AF**



Thorolfsdottir, R.B. et al. J Am Coll Cardiol. 2017;70(17):2157–68.

This study used whole-genome sequence data to conduct a GWAS of AF (N = 13,471 sequences). We identified 2 novel AF variants and replicated 29 previously reported. Our discovery of a missense variant in *PLEC*, together with our previous findings of rare coding AF variants in 2 myosin genes, *MYH6* and *MYL4*, underscores the role of structural components of the myocardium in the pathogenesis of AF. Furthermore, the study demonstrates the fact that some AF variants have no effect on normal cardiac electrophysiological function, whereas others do. This suggests fundamentally different mechanisms in the development of AF. AF = atrial fibrillation; ECG = electrocardiogram; GWAS = genome-wide association study.

variant found to cause atrial fibrillation in a GWAS and, like the other 2 variants, is in a myocardial structure gene (13,14). This finding underscores the importance of the structural properties of the atria in the pathogenesis of AF.

We also replicated 29 previously reported AF variants and show that in general the variants affect SSS and IS in proportion to their AF effect. On the other hand, the AF variants have highly diverse effects on ECG traits in unaffected individuals, ranging from none or mild to extensive, demonstrating that sequence variants can affect the risk of arrhythmia without substantial disruption of normal cardiac electrical function.

There are several possible mechanisms by which a missense variant in the intermediate filament binding

site of plectin may have structural effects on the heart and cause electrophysiological abnormalities. Studies of plectin$^{-/-}$ mice revealed disintegration of intercalated discs and sarcomere disarrangement, thought to result from lack of plectin in Z-line and intercalated disc structures, to which it connects through intermediate filaments (31,34). Furthermore, the role of plectin in maintaining the integrity of the heart has been demonstrated by observations of cardiac involvement among individuals with homozygous or compound heterozygous *PLEC* mutations (35–39). Homozygous protein-truncating mutations in *PLEC* are known to cause a syndrome of skin fragility, epidermolysis bullosa simplex (EBS), commonly accompanied by muscular dystrophy (EBS-MD) (40,41), myasthenic syndrome (EBS-MDMyS) (42,43), limb-girdle muscular dystrophy type 2Q (LGMD2Q) (44), or pyloric atresia (EBS-PA) (45). Mutations in *PLEC* also cause EBS-Ogna, an autosomal dominant disease without muscular dystrophy (46). Cardiomyopathy and arrhythmias have been described in 5 patients with EBS due to *PLEC* mutations (35–39). An EBS phenotype was not observed in our data for homozygous carriers of p.Gly4098Ser, possibly reflecting the fact that the mutation is not protein-truncating.

p.Gly4098Ser in *PLEC* does not associate with cardiomyopathy in our data, nevertheless association with subclinical or mild cardiomyopathy cannot be excluded. Also, we cannot exclude the possibility that p.Gly4098Ser could cause subclinical or even overt atrial myopathy as this phenotype is not routinely assessed or diagnosed. Indeed, the mutation results in a widespread effect on cardiac electrical function in both atria and ventricles. Furthermore, the association with lower R amplitude and lower creatine kinase levels may reflect less skeletal and cardiac muscle mass of carriers.

Both the previously discovered AF genes, *MYH6* (13) and *MYL4* (14), encode parts of myosin, a major component of the sarcomere, the building block of the muscle contractile system (47). Discovery of their association with AF turned the attention to primary atrial cardiomyopathy as an important cause of this arrhythmia and other clinically significant electrophysiological abnormalities. The discovery of a missense variant in *PLEC* causing AF in the absence of clinical heart failure further supports the theory of clinical electrophysiological consequences of otherwise subclinical mechanical dysfunction. These findings parallel the increasing recognition of atrial cardiomyopathy as a salient entity (48) and an increased awareness of the role of atrial remodeling, defined as any change in atrial structure in the pathogenesis of AF (49). This is exemplified by *MYL4*, where individuals homozygous for a rare frameshift deletion develop AF as early as adolescence (14).

The association between AF variants and ECG traits varies substantially. The 2 coding variants, both in myocardial structural genes, *PLEC* and *MYH6*, have extensive effects on cardiac electrical function, but the *MYH6* variant has a comparatively stronger effect on the atria and AVN than on the ventricles. This is consistent with expression patterns of plectin and αMHC, the protein product of *MYH6*. αMHC expression is restricted to atrial tissue after birth, whereas plectin is equally expressed in atria and ventricles (25,50).

Of the 31 variants we assessed, 29 are in noncoding regions. For most, the genes and molecular mechanisms linking these noncoding variants to risk of AF have not been established. However, many of them are located in or close to genes with a known role in cardiovascular function or development, giving rise to theories on how they affect AF risk. For example, the intronic variant in *SCN10A*, encoding the alpha subunit of a voltage-gated sodium channel, is likely to act through a neighboring sodium channel gene, *SCN5A*. It is in LD with a common variant within *SCN10A* proven to affect expression of *SCN5A*, which is expressed in the adult human heart at 1,000-fold levels higher than its neighboring *SCN10A* (51). This variant widely affects cardiac electrical function. The similarly widespread effect on conduction by the variants in the 2 myocardial structural genes *PLEC* and *MYH6* strongly suggests an important role of cardiac mechanics in maintaining normal electrical conduction. Extensive effect on electrical function is also seen for variants in or close to the developmental genes *NKX2-5* and *TBX5*.

The strongest common AF variant is close to *PITX2* that has a role in cardiac development, and its dysfunction has been linked to atrial structural remodeling (52). Here, the variant only affects indices of atrial and AVN conduction with no observable effect on ventricular conduction. This is consistent with the fact that *PITX2* expression in the ventricles is downregulated during embryogenesis, whereas atrial expression remains high (53), and that atrial but not ventricular chamber-specific deletion of *PITX2* in mice results in atrial electric and structural remodeling linked to arrhythmogenesis (52).

Several AF variants have only mild or no effects on electrocardiogram traits, and there is no clear link between AF effect size and the effect on conduction, highlighting the complex relationship between normal cardiac electrophysiology and arrhythmia development. Clearly, variants can affect the risk of

arrhythmia without substantial disruption of the normal cardiac electrical function. For example, the novel intergenic variant close to *LINC01142* and *METTL11B* does not have an observable effect on conduction. It is unclear by which mechanism this locus affects the risk of AF; there are no obvious candidate genes in the region, and the associated SNPs do not associate with expression of either of the 2 flanking genes in our expression dataset or in that of the Genotype-Tissue Expression project (25).

STUDY LIMITATIONS. The p.Gly4098Ser mutation in *PLEC* is very rare in non-Icelanders and thus we have not been able to demonstrate an AF association for this mutation in populations outside of Iceland. However, this fact does not prevent the generalizability of the discovery to a broader population for several reasons. The *PLEC* gene is already strongly implicated in cardiac function and is the third gene encoding a cardiac structural protein to be linked with AF risk in a genome-wide study (13,14). Furthermore, the mutation associates genome-wide significantly with several ECG variables in the absence of AF, confirming its effects on cardiac conduction. Last, despite this particular variant being largely specific to the Icelandic population, other variants in the *PLEC* gene might be found to affect AF in other populations as was the case for *MYL4* (54).

CONCLUSIONS

We describe the discovery of the third rare or low-frequency coding variant conferring risk of AF in a GWAS. Of interest, all 3 variants were found in genes that affect myocardial structure. This underscores the role of structural components of the myocardium in the pathogenesis of AF and supports the notion that otherwise subclinical mechanical dysfunction may have clinical electrophysiological consequences. Coding variants in the structural genes *PLEC* and *MYH6* associate with AF, and both widely affect ECG measurements in the absence of AF. Other GWAS AF variants that we assess are in noncoding regions, and the mechanism by which they affect AF risk is largely unknown. Here we demonstrate a marked diversity in their effects on normal cardiac electrical function and, in some cases, a lack of such an effect, underscoring the extreme complexities in the pathogenesis of this common arrhythmia.

ACKNOWLEDGMENTS The authors thank the study subjects for their valuable participation and our colleagues, who contributed to data collection, sample handling, and genotyping.

ADDRESS FOR CORRESPONDENCE: Dr. Kari Stefansson, deCODE genetics/Amgen, Inc., Sturlugata 8, 101 Reykjavik, Iceland. E-mail: kstefans@decode.is. OR Dr. Rosa B. Thorolfsdottir, deCODE genetics/Amgen, Inc., Sturlugata 8, 101 Reykjavik, Iceland. E-mail: rosath@decode.is.

PERSPECTIVES

COMPETENCY IN MEDICAL KNOWLEDGE: Coding variants in myocardial structural genes increase the risk of developing atrial fibrillation, suggesting a role for cardiac mechanics in maintaining electrical homeostasis. Maintaining normal cardiac structure is important for prevention of AF, and sequence variants can raise the risk of arrhythmia with little or no substantial disruption of normal cardiac electrical function.

TRANSLATIONAL OUTLOOK: Further studies are needed to elucidate the mechanisms by which sequence variants raise the risk of developing AF.

REFERENCES

1. Go AS, Hylek EM, Phillips KA, et al. Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) study. JAMA 2001; 285:2370–5.

2. Benjamin EJ, Wolf PA, D'Agostino RB, Silbershatz H, Kannel WB, Levy D. Impact of atrial fibrillation on the risk of death: the Framingham Heart Study. Circulation 1998;98:946–52.

3. Stewart S, Hart CL, Hole DJ, McMurray JJ. A population-based study of the long-term risks associated with atrial fibrillation: 20-year follow-up of the Renfrew/Paisley study. Am J Med 2002;113:359–64.

4. Tucker NR, Ellinor PT. Emerging directions in the genetics of atrial fibrillation. Circ Res 2014; 114:1469–82.

5. Chen YH, Xu SJ, Bendahhou S, et al. KCNQ1 Gain-of-Function Mutation in Familial Atrial Fibrillation299. New York, NY: Science, 2003: 251–4.

6. Gudbjartsson DF, Arnar DO, Helgadottir A, et al. Variants conferring risk of atrial fibrillation on chromosome 4q25. Nature 2007;448:353–7.

7. Gudbjartsson DF, Holm H, Gretarsdottir S, et al. A sequence variant in ZFHX3 on 16q22 associates with atrial fibrillation and ischemic stroke. Nat Genet 2009;41:876–8.

8. Benjamin EJ, Rice KM, Arking DE, et al. Variants in ZFHX3 are associated with atrial fibrillation in individuals of European ancestry. Nat Genet 2009; 41:879–81.

9. Ellinor PT, Lunetta KL, Albert CM, et al. Meta-analysis identifies six new susceptibility loci for atrial fibrillation. Nat Genet 2012;44:670–5.

10. Ellinor PT, Lunetta KL, Glazer NL, et al. Common variants in KCNN3 are associated with lone atrial fibrillation. Nat Genet 2010;42:240–4.

11. Sinner MF, Tucker NR, Lunetta KL, et al. Integrating genetic, transcriptional, and functional analyses to identify five novel genes for atrial fibrillation. Circulation 2014;130:1225–35.

12. Lubitz SA, Lunetta KL, Lin H, et al. Novel genetic markers associate with atrial fibrillation risk in Europeans and Japanese. J Am Coll Cardiol 2014;63:1200–10.

13. Holm H, Gudbjartsson DF, Sulem P, et al. A rare variant in MYH6 is associated with high risk of sick sinus syndrome. Nat Genet 2011;43:316–20.

14. Gudbjartsson DF, Holm H, Sulem P, et al. A frameshift deletion in the sarcomere gene MYL4 causes early-onset familial atrial fibrillation. Eur Heart J 2016.

15. den Hoed M, Eijgelsheim M, Esko T, et al. Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. Nat Genet 2013;45:621–31.

16. Pfeufer A, van Noord C, Marciante KD, et al. Genome-wide association study of PR interval. Nat Genet 2010;42:153–9.

17. Holm H, Gudbjartsson DF, Arnar DO, et al. Several common variants modulate heart rate, PR interval and QRS duration. Nat Genet 2010;42:117–22.

18. Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. Nat Genet 2017.

19. Low SK, Takahashi A, Ebana Y, et al. Identification of six new genetic loci associated with atrial fibrillation in the Japanese population. Nat Genet 2017.

20. Sabatine MS, Giugliano RP, Keech A, et al. Rationale and design of the Further cardiovascular OUtcomes Research with PCSK9 Inhibition in subjects with Elevated Risk trial. Am Heart J 2016;173:94–101.

21. Gudbjartsson DF, Helgason H, Gudjonsson SA, et al. Large-scale whole-genome sequencing of the Icelandic population. Nat Genet 2015;47:435–44.

22. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 2015;47:291–5.

23. Sveinbjornsson G, Albrechtsen A, Zink F, et al. Weighting sequence variants based on their annotation increases power of whole-genome association studies. Nat Genet 2016;48:314–7.

24. Wiche G. Role of plectin in cytoskeleton organization and dynamics. J Cell Sci 1998;111 Pt 17:2477–86.

25. The Genotype-Tissue Expression (GTEx) project. Nat Genet 2013;45:580–5.

26. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248–9.

27. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4:1073–81.

28. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536:285–91.

29. Cooper GM, Stone EA, Asimenos G, Green ED, Batzoglou S, Sidow A. Distribution and intensity of constraint in mammalian genomic sequence. Genome Res 2005;15:901–13.

30. Gussow AB, Petrovski S, Wang Q, Allen AS, Goldstein DB. The intolerance to functional genetic variation of protein domains predicts the localization of pathogenic mutations within genes. Genome Biol 2016;17:9.

31. Castanon MJ, Walko G, Winter L, Wiche G. Plectin-intermediate filament partnership in skin, skeletal muscle, and peripheral nerve. Histochemistry and cell biology 2013;140:33–53.

32. Cheng S, Keyes MJ, Larson MG, et al. Long-term outcomes in individuals with prolonged PR interval or first-degree atrioventricular block. JAMA 2009;301:2571–7.

33. Chang HY, Lin YJ, Lo LW, et al. Sinus node dysfunction in atrial fibrillation patients: the evidence of regional atrial substrate remodelling. Europace 2013;15:205–11.

34. Andra K, Lassmann H, Bittner R, et al. Targeted inactivation of plectin reveals essential function in maintaining the integrity of skin, muscle, and heart cytoarchitecture. Gene Dev 1997;11:3143–56.

35. Schroder R, Kunz WS, Rouan F, et al. Disorganization of the desmin cytoskeleton and mitochondrial dysfunction in plectin-related epidermolysis bullosa simplex with muscular dystrophy. J Neuropathol Exp Neurol 2002;61:520–30.

36. Bolling MC, Pas HH, de Visser M, et al. PLEC1 mutations underlie adult-onset dilated cardiomyopathy in epidermolysis bullosa simplex with muscular dystrophy. J Invest Dermatol 2010;130:1178–81.

37. Villa CR, Ryan TD, Collins JJ, Taylor MD, Lucky AW, Jefferies JL. Left ventricular non-compaction cardiomyopathy associated with epidermolysis bullosa simplex with muscular dystrophy and PLEC1 mutation. Neuromuscul Disord 2015;25:165–8.

38. Celik C, Uysal H, Heper AO, Karaoglan B. Epidermolysis bullosa simplex associated with muscular dystrophy and cardiac involvement. Journal of clinical neuromuscular disease 2005;6:157–61.

39. Winter L, Türk M, Harter PN, et al. Downstream effects of plectin mutations in epidermolysis bullosa simplex with muscular dystrophy. Acta Neuropathol Commun 2016;4:44.

40. McLean WH, Pulkkinen L, Smith FJ, et al. Loss of plectin causes epidermolysis bullosa with muscular dystrophy: cDNA cloning and genomic organization. Genes Dev 1996;10:1724–35.

41. Pulkkinen L, Smith FJ, Shimizu H, et al. Homozygous deletion mutations in the plectin gene (PLEC1) in patients with epidermolysis bullosa simplex associated with late-onset muscular dystrophy. Hum Mol Genet 1996;5:1539–46.

42. Banwell BL, Russel J, Fukudome T, Shen XM, Stilling G, Engel AG. Myopathy, myasthenic syndrome, and epidermolysis bullosa simplex due to plectin deficiency. J Neuropathol Exp Neurol 1999;58:832–46.

43. Selcen D, Juel VC, Hobson-Webb LD, et al. Myasthenic syndrome caused by plectinopathy. Neurology 2011;76:327–36.

44. Gundesli H, Talim B, Korkusuz P, et al. Mutation in exon 1f of PLEC, leading to disruption of plectin isoform 1f, causes autosomal-recessive limb-girdle muscular dystrophy. Am J Hum Genet 2010;87:834–41.

45. Nakamura H, Sawamura D, Goto M, et al. Epidermolysis bullosa simplex associated with pyloric atresia is a novel clinical subtype caused by mutations in the plectin gene (PLEC1). J Mol Diagn 2005;7:28–35.

46. Koss-Harnes D, Hoyheim B, Anton-Lamprecht I, et al. A site-specific plectin mutation causes dominant epidermolysis bullosa simplex Ogna: two identical de novo mutations. J Invest Dermatol 2002;118:87–93.

47. Fuster V. Hurst's the Heart. New York, NY: McGraw Hill, 2004.

48. Goette A, Kalman JM, Aguinaga L, et al. EHRA/HRS/APHRS/SOLAECE expert consensus on atrial cardiomyopathies: definition, characterization, and clinical implication. Europace 2016;18:1455–90.

49. Nattel S, Harada M. Atrial remodeling and atrial fibrillation: recent advances and translational perspectives. J Am Coll Cardiol 2014;63:2335–45.

50. Franco D, Lamers WH, Moorman AF. Patterns of expression in the developing myocardium: towards a morphologically integrated transcriptional model. Cardiovasc Res 1998;38:25–53.

51. van den Boogaard M, Smemo S, Burnicka-Turek O, et al. A common genetic variant within SCN10A modulates cardiac SCN5A expression. J Clin Invest 2014;124:1844–52.

52. Chinchilla A, Daimi H, Lozano-Velasco E, et al. PITX2 insufficiency leads to atrial electrical and structural remodeling linked to arrhythmogenesis. Circ Cardiovasc Genet 2011;4:269–79.

53. Franco D, Chinchilla A, Daimi H, Dominguez JN, Aranega A. Modulation of conductive elements by Pitx2 and their impact on atrial arrhythmogenesis. Cardiovasc Res 2011;91:223–31.

54. Orr N, Arnaout R, Gula LJ, et al. A mutation in the atrial-specific myosin light chain gene (MYL4) causes familial atrial fibrillation. Nature Commun 2016;7:11303.

KEY WORDS arrhythmia, atrial fibrillation, electrocardiogram, plectin

APPENDIX For an expanded Methods section as well as supplemental figures and tables, please see the online version of this article.

Paper II

COMMUNICATIONS BIOLOGY

ARTICLE

DOI: 10.1038/s42003-018-0068-9 OPEN

Coding variants in *RPL3L* and *MYZAP* increase risk of atrial fibrillation

Rosa B. Thorolfsdottir et al.[#]

Most sequence variants identified hitherto in genome-wide association studies (GWAS) of atrial fibrillation are common, non-coding variants associated with risk through unknown mechanisms. We performed a meta-analysis of GWAS of atrial fibrillation among 29,502 cases and 767,760 controls from Iceland and the UK Biobank with follow-up in samples from Norway and the US, focusing on low-frequency coding and splice variants aiming to identify causal genes. We observe associations with one missense (OR = 1.20) and one splice-donor variant (OR = 1.50) in *RPL3L*, the first ribosomal gene implicated in atrial fibrillation to our knowledge. Analysis of 167 RNA samples from the right atrium reveals that the splice-donor variant in *RPL3L* results in exon skipping. We also observe an association with a missense variant in *MYZAP* (OR = 1.38), encoding a component of the intercalated discs of cardiomyocytes. Both discoveries emphasize the close relationship between the mechanical and electrical function of the heart.

[#]A full list of authors and their affliations appears at the end of the paper. Correspondence and requests for materials should be addressed to H.H. (email: Hilma.Holm@decode.is) or to K.S. (email: kstefans@decode.is)

Atrial fibrillation is the most common arrhythmia of clinical significance with an estimated number of 33.5 million individuals diagnosed with atrial fibrillation globally in the year 2010[1]. It is associated with increased mortality and morbidity, particularly stroke and heart failure, and is responsible for substantial health care costs[1]. Atrial fibrillation is a complex disease that is characterized by both mechanical and electrical abnormalities of the atria that may be detected prior to diagnosis of the arrhythmia itself. The role of atrial myopathy and fibrosis in the development of atrial fibrillation is increasingly recognized and it has been postulated that these processes may contribute to cardioembolic stroke in the absence of arrhythmia[2]. Thus, identification of the early stages of atrial myopathy may allow for therapy to prevent progression to atrial remodeling, atrial fibrillation, and stroke[3].

Genome-wide association studies (GWAS), assessing primarily common sequence variants, have yielded over 30 genetic loci that associate with atrial fibrillation[4]. Most of the associated variants are non-coding and the causative genes remain unknown but the closest genes reveal a polygenic process, implicating transcription factors, cardiac ion channels, myocardial, and cytoskeletal proteins in the pathogenesis of atrial fibrillation. In the pre-GWAS era, linkage mapping and candidate gene sequencing linked a number of rare sequence variants to atrial fibrillation, mostly in single cases or familial atrial fibrillation, including variants in cardiac ion channel genes[4]. These variants explain a small proportion of atrial fibrillation cases, and for many, the genetic evidence is not robust.

In the past few years, through GWAS based on whole-genome sequencing, we have identified three low-frequency coding variants that associate with atrial fibrillation[5–8]. All three variants are in structural genes, the myosin sarcomere genes *MYH6*[5] and *MYL4*[6,7] and the cytoskeletal gene *PLEC*[8]. These findings support the notion of an important relationship between myocardial mechanical integrity and the development of arrhythmias.

Here, we continue our search for variants associated with atrial fibrillation to shed further light on the pathophysiology of this common arrhythmia. We performed an atrial fibrillation GWAS using data from Iceland and the UK Biobank, focusing on rare and low-frequency coding and splice variants, with follow-up of the most significant variants in samples from Norway and the US.

Results

Associations with coding variants in *RPL3L* and *MYZAP*. We performed a meta-analysis on atrial fibrillation including 14,710 cases and 373,897 controls from Iceland and 14,792 cases and 393,863 controls from the UK Biobank[9], focusing on variants annotated as having moderate or high impact on protein function (including moderate: missense, in-frame indel, splice-region, and high impact: splice-acceptor, splice-donor, frameshift, stop-gained, and stop lost variants)[10]. To account for the expected impact, we applied the significance thresholds of $P < 5.1 \times 10^{-8}$ for moderate and $P < 2.6 \times 10^{-7}$ for high-impact variants[11].

We found two novel genome-wide significant atrial fibrillation associations in the gene *RPL3L* on chromosome 16, with the missense variant p.Ala75Val (allele frequency 3.65% in Iceland, OR: 1.19, $P = 3.4 \times 10^{-12}$) and the splice-donor variant c.1167+1G>A (allele frequency 0.61% in Iceland, OR: 1.52, $P = 8.2 \times 10^{-10}$). The two *RPL3L* variants are not correlated ($D' = 1$, $r^2 = 0.00024$), and when conditioned on each other, both associations with atrial fibrillation remained (Supplementary Table 1). To assess the relationship between *RPL3L* and atrial fibrillation further, we tested all 15 low-frequency coding variants in the gene for association with atrial fibrillation after conditioning on p.Ala75Val and c.1167+1G>A (significance threshold = 0.05/15 = 0.0033, Supplementary Table 2). One variant associated with atrial fibrillation with a *P*-value below this threshold but the association was not genome-wide significant in the meta-analysis. The *RPL3L* gene encodes a ribosomal protein (ribosomal protein like 3L) that is primarily expressed in skeletal muscle and heart unlike most ribosomal proteins, that are ubiquitously expressed[12].

We also observed a suggestive association with the missense variant p.Gln254Pro in the gene *MYZAP* on chromosome 15 (allele frequency 1.08% in Iceland, OR: 1.36, $P = 7.8 \times 10^{-8}$) (Table 1). No other coding variant in *MYZAP* associates independently with atrial fibrillation (Supplementary Table 3). *MYZAP* encodes myozap, myocardial zonula adherens protein, primarily expressed in the heart in man and its homolog in the mouse has been localized to the intercalated discs[13].

To further assess these associations, we tested the three sequence variants in four additional sample sets of 9204 cases and 76,161 controls combined, from the Nord-Trøndelag Health Study (HUNT), the Further Cardiovascular Outcomes Research with PCSK9 Inhibition in Subjects with Elevated Risk (FOURIER) trial, the Vanderbilt atrial fibrillation Registry, and the Tromsø Study. Joint analysis of all data sets yielded genome-wide significant association with atrial fibrillation of all three variants, *RPL3L* Ala75Val (OR: 1.20, $P = 1.7 \times 10^{-14}$), *RPL3L* c.1167+1G>A (OR: 1.50, $P = 5.0 \times 10^{-10}$), and *MYZAP* p.Gln254Pro (OR: 1.38, $P = 3.3 \times 10^{-10}$) (Table 1).

Three other moderate or high-impact coding variants in the genes *MYH6*, *PLEC*, and *MYL4* (recessive model), previously reported by us, were genome-wide significantly associated with atrial fibrillation in this data set[5–8].

The p.Ala75Val in *RPL3L* associates with electrocardiogram measures. We have previously demonstrated that the effects of reported atrial fibrillation variants on ECG traits measured in sinus rhythm range from none to extensive and there is no clear relationship between effects on atrial fibrillation and effects on ECG measures[8]. For example, a sequence variant associated with atrial fibrillation in the sodium channel gene *SCN10A* has extensive and strong effects on ECG measures but a relatively small atrial fibrillation effect compared to the most significant common atrial fibrillation variant near *PITX2* that has minimal effect on ECG measurements (Fig. 1). Figure 1 shows the effects of the *RPL3L* and *MYZAP* variants on ECG traits compared to the effects of 31 published atrial fibrillation variants. For the analysis we used 289,297 sinus rhythm ECGs from 62,974 individuals not diagnosed with atrial fibrillation and tested all variants for association with 122 ECG variables, some of which are correlated (Supplementary Table 4 and Supplementary Data 1). We used the Benjamini–Hochberg false discovery rate (FDR) procedure controlling the FDR at 0.05 at each marker to account for multiple testing. The *RPL3L* missense variant p.Ala75Val associates with measures of atrial conduction during sinus rhythm, both P wave amplitude and area, and with QRS duration. Neither of the other variants in *RPL3L* and *MYZAP* associated with ECG traits in sinus rhythm. When testing for association with ECG traits using all ECGs irrespective of rhythm and history of atrial fibrillation, p.Ala75Val in *RPL3L* associates more significantly with ECG measurements and p.Gln254Pro in *MYZAP* associates with various P wave indices, R amplitude, and T wave indices (Supplementary Fig. 1).

The p.Gln254Pro in *MYZAP* associates with sick sinus syndrome. Variants that associate with risk of atrial fibrillation also tend to associate with the related atrial arrhythmia sick sinus syndrome (SSS) and commonly with effects that are proportional to the atrial fibrillation effects[8]. One notable exception is the

Table 1 Meta-analysis results for atrial fibrillation variants

	MYZAP p.Gln254Pro missense			*RPL3L* p.Ala75Val missense			*RPL3L* c.1167+1G>A splice-donor		
Rs name	**rs147301839**			**rs140185678**			**rs140192228**		
Position (hg38)	**chr15:57632516**			**chr16:1953015**			**chr16:1945498**		
Data set (cases/controls)	**Freq %**	**OR (95% CI)**	***P***	**Freq %**	**OR (95% CI)**	***P***	**Freq %**	**OR (95% CI)**	***P***
Iceland (14,710/373,897)	1.08	1.38 (1.20–1.60)	9.0×10^{-6}	3.65	1.18 (1.09–1.28)	6.4×10^{-5}	0.61	1.37 (1.14–1.65)	8.7×10^{-4}
UK Biobank (14,792/393,863)	0.36	1.32 (1.10–1.57)	0.0023	3.37	1.20 (1.13–1.27)	1.2×10^{-8}	0.31	1.71 (1.40–2.07)	6.9×10^{-8}
Meta-analysis: Iceland and UK		**1.36 (1.21–1.52)**	$\mathbf{7.8 \times 10^{-8}}$		**1.19 (1.13–1.25)**	$\mathbf{3.4 \times 10^{-12}}$		**1.52 (1.33–1.74)**	$\mathbf{8.2 \times 10^{-10}}$
Fourier (1238/11,562)	0.21	1.61 (0.46–5.69)	0.46	4.88	0.93 (0.55–1.59)	0.79	0.12	1.59 (0.13–20.01)	0.72
Vanderbilt (759/759)	0.46	7.03 (1.29–38.23)	0.024	3.00	1.33 (0.88–2.02)	0.18	0.43	1.61 (0.46–5.69)	0.40
Tromsø (714/698)	0.49	1.14 (0.39–3.32)	0.81	3.90	1.29 (0.87–1.90)	0.20	0.073	1.60[a] (0.53–4.88)	0.30
HUNT (6493/63,142)	0.64	1.47 (1.14–1.89)	0.0027	2.87	1.22 (1.07–1.40)	0.0038	0.15	1.19 (0.71–2.00)	0.51
Replication only (Fourier, Vanterbilt, Tromsø, and HUNT)		**1.50 (1.18–1.91)**	$\mathbf{8.1 \times 10^{-4}}$		**1.22 (1.08–1.37)**	**0.0011**		**1.33 (0.88–2.00)**	**0.17**
Combined		**1.38 (1.25–1.53)**	$\mathbf{3.3 \times 10^{-10}}$		**1.20 (1.14–1.25)**	$\mathbf{1.7 \times 10^{-14}}$		**1.50 (1.32–1.70)**	$\mathbf{5.0 \times 10^{-10}}$

Freq frequency of minor allele, *OR* odds ratio, *95% CI* 95% confidence interval, *P P* value from case–control association analysis. Results from more than one cohort are in bold
Here we show association results for atrial fibrillation for the discovery and follow-up data sets and the joint analysis of all data sets (combined). The *P* value for heterogeneity analysis was 0.52 for p.Gln254Pro in *MYZAP*, 0.92 for p.Ala75Val in *RPL3L* and 0.65 for c.1167+1G>A in *RPL3L*
[a]No control was a carrier of the variant resulting in 2×2 table entry with a zero count. Effect and *P*-value was calculated with the R-package "metafor" by adding a small constant of 1/2 to the cells of the 2×2 table

missense mutation in *MYH6* that we originally discovered through its association with high risk of SSS and confers a substantially greater risk of SSS than predicted from its effect on atrial fibrillation risk[5]. We tested the three new atrial fibrillation variants in 3568 SSS cases and 346,025 controls from Iceland and 403 cases and 403,181 controls from the UK Biobank[9]. In the joint analysis, p.Gln254Pro in *MYZAP* associates with SSS (OR: 1.65, 95% CI: 1.33–2.05, $P = 5.0 \times 10^{-6}$) (Supplementary Table 5).

To gain a better understanding of the new atrial fibrillation variants, we tested them for association with other phenotypes in deCODE´s genotype/phenotype database under both additive and recessive models but found no other associations passing Bonferroni correction. Association results for available relevant phenotypes including risk factors of atrial fibrillation are listed in Supplementary Tables 6 and 7. Since mutations in ribosomal genes are commonly associated with bone marrow failure, we specifically queried the relationship between the *RPL3L* variants and blood cells and found no associations. Similarly, since mutations in intercalated disc genes, albeit not *MYZAP*, have been associated with cardiomyopathies in man[14] we assessed the link between the *MYZAP* variant and cardiomyopathies in our database, but found none.

The splice-donor variant in *RPL3L* causes exon skipping. We obtained RNA samples from cardiac atria of 167 Icelanders and used them to assess the effect of the splice-donor variant c.1167 +1G>A in *RPL3L*. Two of the 167 individuals carry this variant. Non-carriers only produce the primary *RPL3L* isoform, but both carriers also produce an alternative isoform that skips exon 9 ($P = 0.0052$, Fig. 2a). We also found that carriers express the two isoforms in approximately equal abundance. Exon 9 is the second to last exon in *RPL3L* and is 120 base pairs long, and therefore its deletion is in-frame (Fig. 2b).

Discussion

By performing a meta-analysis of atrial fibrillation using samples from deCODE and the UK biobank, focusing on rare and low-frequency coding and splice variants, with follow up in four sample sets from Norway and the US, we discovered three new atrial fibrillation variants in two genes, two in the ribosomal gene *RPL3L* and one in *MYZAP* that encodes a component of the cardiac intercalated discs. Risk of atrial fibrillation has not been associated with a ribosomal gene before.

The eukaryotic ribosome, composed of four different ribosomal RNAs and ~80 ribosomal proteins, is a complex cellular machine that translates messenger RNA into protein[15]. Only a few rare inherited diseases have been specifically linked to mutations in genes encoding ribosomal proteins. They include Diamond–Blackfan anemia and Shwachman–Diamond syndrome that are characterized by a distinct set of clinical features, including bone marrow failure and/or developmental abnormalities[16]. The ribosome has generally been considered to function in a housekeeping capacity but recent studies have revealed that ribosome activity may be regulated in a cell-specific manner, for example through changes in the protein composition of the ribosome[17,18]. One example is the *RPL3L* with expression restricted to skeletal muscle and the heart[12]. Ribosomes containing RPL3L instead of its ubiquitously expressed homolog, RPL3, have altered translational activity and it has been postulated that RPL3L may be a negative regulator of muscle growth[19].

The *RPL3L* missense variant associating with atrial fibrillation is p.Ala75Val, and Ala75 is highly conserved in both *RPL3L* and *RPL3* over a range of species (PROVEAN impact prediction scores <−2.5[20]) (Supplementary Table 8). Sequencing of RNA

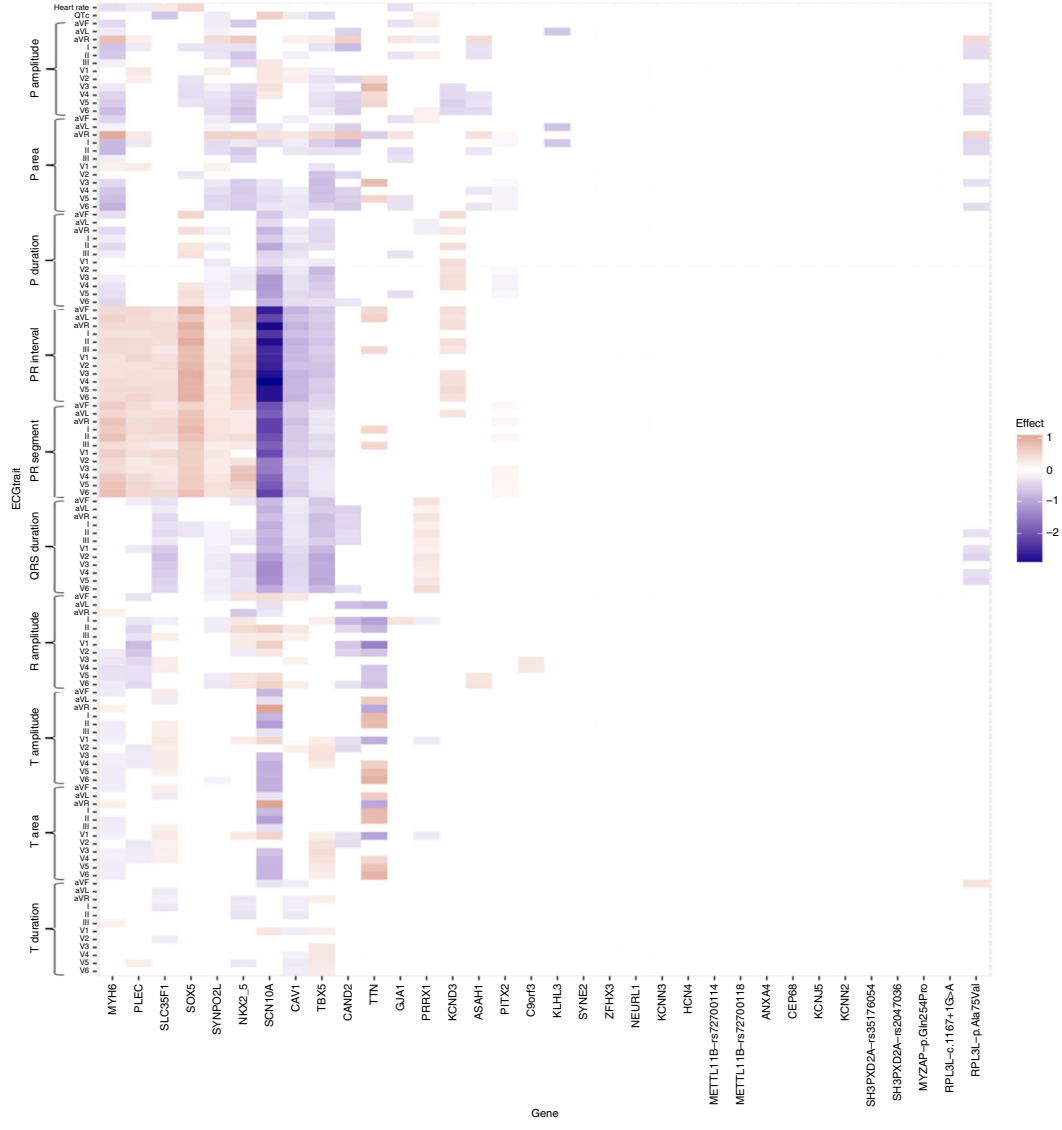


Fig. 1 Heatmap showing the effects of atrial fibrillation variants on ECG traits in sinus rhythm ECGs, excluding atrial fibrillation cases. See Thorolfsdottir et al[8]. ECG measurements were available for 62,974 individuals without atrial fibrillation. Each column shows the estimated effect of the risk allele of an atrial fibrillation variant on various ECG traits. The effect of each variant, annotated with the corresponding gene name, is scaled with the $\log_{10}$-atrial fibrillation odds ratio. Red color represents a positive effect on the ECG variable and blue color a negative effect. The effect is shown only for significant associations after adjusting for multiple testing with a false discovery rate procedure for each variant. Non-significant associations are white in the heatmap

samples from cardiac atria including from carriers of the splice-donor variant, c.1167+1G>A, demonstrated that the variant leads to skipping of *RPL3L* exon 9, the second to the last exon that encodes amino acid residues 350 to 389. These residues are 75% identical to the corresponding RPL3 residues. In yeast it has been shown that amino acids 373–380 in RPL3, corresponding to amino acids 382–389 in human RPL3L, form a part of the contact site of the ribosomes with the signal recognition particle that targets ribosomes to the endoplasmic reticulum membrane[21]. Based on functional similarities between RPL3 and RPL3L it is therefore possible that c.1167+1G>A disrupts engagement of RPL3L containing ribosomes with the endoplasmic reticulum and thus reducing ribosomal function. Since both *RPL3L* variants increase the risk of atrial fibrillation it could be predicted, based on the suggested effect of the splice-donor variant, that the variants are loss of, rather than gain of, function. The association of atrial fibrillation with a gene expressed in the atria that is involved with regulation of muscle growth is in line with the increasingly recognized tight link between mechanical myocardial integrity and the electrical function of the heart.

The *MYZAP* gene was recently discovered by Seeger et al. in an effort to find new components of the intercalated discs[13], a highly

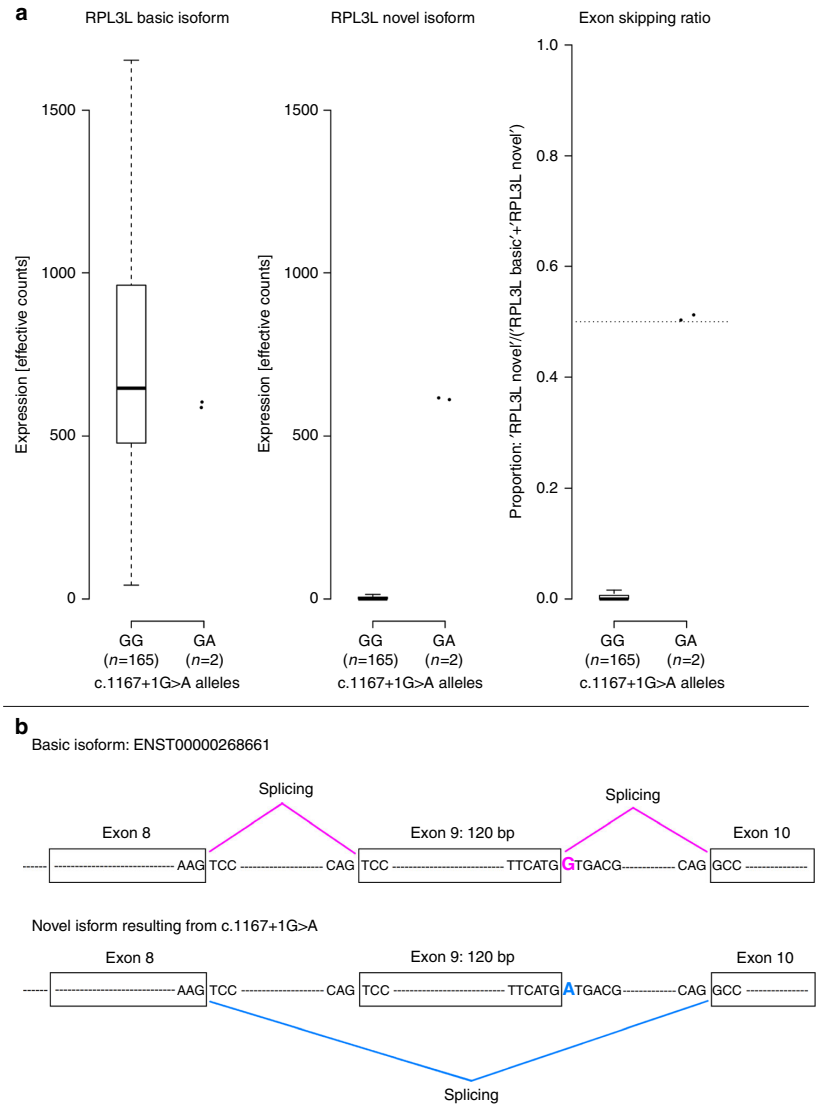


Fig. 2 The effect of the splice-donor variant c.1167+1G>A in *RPL3L* on splicing. **a** Quantification of two forms of RPL3L transcripts; the primary isoform, ENST00000268661, and a novel isoform with skipping of exon 9 resulting from c.1167+1G>A. It also shows the proportion of novel isoform among all transcripts. A total of 167 samples, all from the right atrium, where included in the analysis. Two of those came from carriers of c.1167+1G>A. The figure demonstrates that only the two carriers have the novel isoform with skipping of exon 9. Their exon skipping proportion is ~0.5 while it is zero in non-carriers. **b** A schematic illustration of the splicing of RPL3L among carriers and non-carriers of c.1167+G>A. The variant is in a splice-donor site by the second last exon and results in exon skipping. The skipped exon is 120 base pairs and therefore its deletion is in-frame

specialized cell–cell contact structure that enables mechanical, electrical and chemical communication between cardiomyocytes. Human Myozap mRNA is primarily expressed in the heart and in the mouse the protein was predominantly found at intercalated discs and sarcomeric Z-discs[13]. In vitro functional studies revealed a role in cardiac signal transduction as Myozap promotes serum response factor signaling to the nucleus[13]. A knockdown of the Myozap ortholog in zebrafish and cardiac overexpression of Myozap in the mouse both resulted in cardiomyopathy[13,22], suggesting an important role of the protein in maintaining cardiac integrity and function.

According to PROVEAN[20], Gln254 is conserved and the variant is predicted to be deleterious (Supplementary Table 8). The variant is located at the edge of the Myozap protein region associated with both activation of serum response factor-dependent transcription and actin colocalization (amino acids 91–250), and could therefore potentially affect either one or both of these protein functions[13]. An introduction of proline, a conformationally constrained amino acid, can lead to perturbations in local folding and therefore might interrupt the function of adjacent domains.

Mutations in intercalated disc genes cause cardiomyopathies, in particular arrhythmogenic right ventricular cardiomyopathy, characterized by a notable risk of both atrial fibrillation and ventricular arrhythmias, and one of the leading causes of sudden cardiac death in young people and athletes[23]. Interestingly, conduction abnormalities and arrhythmias in arrhythmogenic right ventricular cardiomyopathy are commonly encountered before the appearance of structural defects[14]. Atrial fibrillation variants have also been identified in and close to genes encoding components of intercalated discs[4], and the atrial fibrillation-associated gene *PITX2* has been shown to directly regulate intercalated disc genes[24]. P.Gln254Pro does not associate with cardiomyopathies, ventricular arrhythmias, or sudden cardiac death in our data, suggesting that it only affects the atria but we may lack power to identify a ventricular effect.

Like p.Gln254Pro in *MYZAP*, the three low-frequency missense and frameshift variants we have previously reported to increase the risk of atrial fibrillation, in *MYH6, MYL4*, and *PLEC*, also increase the risk of SSS[8]. Like *MYZAP*, all three genes encode structural components of the cardiomyocyte. In particular, *PLEC* encodes a multidomain cytoskeletal linking protein which, among other functions, connects with elements of the intercalated disc and has a role in maintaining its integrity[25,26].

In summary, we report the association of three low-frequency coding variants in *RPL3L* and *MYZAP* with increased risk of atrial fibrillation. Using RNA samples from cardiac tissue we show that a splice-donor variant in *RPL3L* causes exon skipping. These results add to previous discoveries of three low-frequency coding variants in structural genes associating with atrial fibrillation and highlight the intricate connection between myocardial structure and arrhythmogenesis. The association of a missense variant in *MYZAP* with atrial fibrillation and SSS emphasizes the role of the intercalated discs in maintaining normal cardiac rhythm. The fact that a coding variant in a ribosomal protein specifically expressed in skeletal muscle and the heart increases risk of atrial fibrillation is in line with the novel concept of ribosome specialization in muscle and underscores the importance of this specialization for normal function of the heart. GWAS have linked a number of common variants with risk of atrial fibrillation but emerging discoveries of low-frequency coding variants associating with atrial fibrillation continue to shed new light on the pathogenesis of the disease.

Methods

The study complies with the Declaration of Helsinki.

Icelandic atrial fibrillation study population. The Icelandic atrial fibrillation sample consisted of 15,552 Icelanders diagnosed with atrial fibrillation (International Classification of Diseases (ICD) 10 code I.48 and ICD 9 code 427.3) according to electronic medical records at Landspitali, The National University Hospital, in Reykjavik, Iceland, and Akureyri Hospital, the two largest hospitals in Iceland, between 1987 and 2017. In total, 14,710 out of the 15,552 cases had genotypes and were included in the analysis. Controls were 373,897 Icelanders recruited through different genetic research projects at deCODE genetics. All participating subjects who donated blood signed informed consent. Personal identities of the participants and biological samples were encrypted by a third party system. The study was approved by the Icelandic Data Protection Authority and the National Bioethics Committee of Iceland (no. VSNb2015030021).

UK Biobank atrial fibrillation study population. The UK Biobank project is a large prospective cohort study of ~500,000 individuals from across UK, aged between 40 and 69 at recruitment. The study has collected extensive phenotypic and genotypic information on its participants, including ICD coded diagnoses from inpatient and out-patient hospital episodes[9]. The atrial fibrillation population from UK Biobank consisted of 14,792 cases and 393,863 controls, all individuals of European ancestry recruited between 2006 and 2010[9]. Atrial fibrillation was ascertained based on ICD diagnoses. These are primary or secondary ICD-9 or ICD-10 diagnoses codes a participant has had recorded across all their episodes in hospital. Self-reported diagnoses were excluded from our analysis. Further details on the recruitment and variables collected in the UK Biobank study can be found in previous publications[9,27].

The Vanderbilt Atrial Fibrillation Registry. We genotyped novel atrial fibrillation variants in an atrial fibrillation sample set (759 cases and 759 controls) from the Vanderbilt Atrial Fibrillation Registry, a clinical and genetic registry at the Vanderbilt University Medical Center in Nashville, Tennessee. At enrollment into the registry, a detailed medical and drug history was obtained from all patients and patients were also asked to complete a symptom questionnaire. Patients with history of atrial fibrillation only associated with cardiac surgery were excluded from this study. Written informed consent was obtained from all patients under a protocol approved by the Vanderbilt University Institutional Review Board.

FOURIER replication cohort. We also followed-up novel atrial fibrillation variants in an atrial fibrillation sample set originating from the Further Cardiovascular Outcomes Research with PCSK9 Inhibition in Subjects with Elevated Risk (FOURIER) trial (1238 atrial fibrillation cases and 11,562 controls). FOURIER is a randomized placebo-controlled, double-blind, parallel-group, multinational trial testing the hypothesis that adding the drug evolocumab to statin therapy will reduce the incidence of major adverse cardiovascular events in patients with clinically evident cardiovascular disease (CVD). The whole study group consisted of 27,564 patients recruited at 1242 cities in 49 countries from 2013 to 2015. Eligible patients were between 40 and 85 years of age and had clinically evident atherosclerotic CVD. The design of the trial has been described in detail elsewhere[28,29]. A subset of over 12,000 participants of European descent from the FOURIER trial have been genotyped by us by whole-exome sequencing, chip-typing, and imputation. Detailed phenotypic information was gathered on all FOURIER study participants, including atrial fibrillation disease status. The Fourier atrial fibrillation sample set consists of 1238 cases and 11,562 controls of European descent, excluding all Icelandic participants.

Norwegian atrial fibrillation study population from the Tromsø Study. The Tromsø Study is a population-based prospective study with repeated health surveys in the municipality of Tromsø, Norway[30]. So far, more than 40,000 individuals have been examined.The population is being followed-up on an individual level with registration and validation of diseases and death and an endpoint registry has been established for CVD. Discharge diagnosis lists of CVD have been retrieved from the University Hospital of North Norway in Tromsø, and medical records for all individuals with a CV discharge diagnosis (including visits to out-patient clinics, out of hospital journals, autopsy records, and death certificates) have been reviewed. Atrial fibrillation has been registered based on ICD-9 and ICD-10 codes since 1986 as part of the ongoing CV endpoint registration in the Tromsø Study. People with postoperative atrial fibrillation only (≤28 days after the procedure) are registered, but are not included as cases. For the current project, one sex-matched and age-matched control for each case of atrial fibrillation from was drawn from the population-based Tromsø 4 survey. Participants in the Tromsø Study gave informed, written consent. The study was approved by the Regional Committee for Medical Research Ethics. The atrial fibrillation sample set consists of 714 cases and 698 controls.

The Nord-Trøndelag Health Study. The Nord-Trøndelag Health Study (HUNT) is a population-based health survey conducted in the county of Nord-Trøndelag, Norway. Individuals were included at three different time points during ~20 years (HUNT1 (1984–1986), HUNT2 (1995–1997), and HUNT3 (2006–2008))[31]. At each time point, the entire adult population (≥20 years) was invited to participate by completing questionnaires, attending clinical examinations, and interviews. Participation rates have generally been high: 89.4% ($n = 77,212$), 69.5% ($n = 65,237$), and 54.1% ($n = 50,807$) in HUNT1, HUNT2, and HUNT3, respectively[31]. Taken together, the health studies included information from over 120,000 different individuals from Nord-Trøndelag. Biological samples including DNA have been collected for ~70,000 participants. Atrial fibrillation was defined based on ICD-10 codes collected from local hospitals and out-patient clinics between 1999 and 2016. Cases (6493) were defined as individual with one or more ICD-9 or ICD-10 codes specific for atrial fibrillation ("I48" or "427.3") whereas controls (63,142) were all individuals without a code specific for atrial fibrillation.

Secondary phenotypes. Novel atrial fibrillation variants were tested for association with other phenotypes in the deCODE genetics phenotype database which contains extensive medical information on various diseases and other traits. The pacemaker population sample set includes 3578 individuals who received a pacemaker implantation (NCSP surgical codes FPE and FPF) at LUH between 1997 and 2015. The SSS sample set includes 3578 individuals who received the diagnosis of SSS (ICD-10 code I49.5, ICD 9 code 427.8) at LUH in Reykjavik between 1987 and 2015. Ischemic stroke cases were identified from a registry of individuals diagnosed with ischemic stroke or transient ischemic attack (TIA) at LUH during the years 1993 to 2014 ($n = 5626$). The ischemic stroke or TIA diagnoses were based on standard WHO criteria and imaging evidence (either CT or MRI), and were clinically confirmed by neurologists. A total of 1369 individuals with ischemic stroke were classified as having cardioembolic stroke based on a neurologist review

of medical records and classification according to the Trial of Org 10172 in Acute Stroke Treatment (TOAST)[32]. The controls used in the various case–control analyses of this study consisted of disease-free controls randomly drawn from the Icelandic genealogical database and individuals from other genetic studies at deCODE. We also assessed the association of novel atrial fibrillation variants with SSS among 403 cases and 403,181 controls in the UK Biobank.

Electrocardiogram data. Electrocardiograms (ECGs) obtained in Landspitali, The National University Hospital, the largest and only tertiary care hospital in Iceland, have been digitally stored since 1998. We have analysed 434,000 ECGs from 88,217 individuals obtained between 1998 and 2015. To assess the effect of atrial fibrillation variants on ECG traits, and thus cardiac electrical function, in the absence of atrial fibrillation, we excluded ECGs from individuals with atrial fibrillation and pacemakers and used 289,297 sinus rhythm (heart rate 50–100 beats per min) ECGs of 62,974 individuals for the primary analysis. The ECGs were digitally recorded with the Philips PageWriter Trim III, PageWriter 200, Philips Page Writer 50, and Phillips Page Writer 70 cardiographs and stored in the Philips TraceMasterVue ECG Management System. These were ECGs obtained in all hospital departments, from both in patients and outpatients. Digitally measured ECG waveforms and parameters were extracted from the database for analysis. The Philips PageWriter Trim III QT interval measurement algorithm has been previously described and shown to fulfill industrial ECG measurement accuracy standards[33]. The Philips PR interval and QRS complex measurements have been shown to fulfill industrial accuracy standards[34].

Whole-genome sequencing, variant calling, and imputation in Iceland. In Iceland, the study is based on whole-genome sequence data from the whole blood of 15,220 Icelanders participating in various disease projects at deCODE genetics. In addition, 151,677 Icelanders have been genotyped using Illumina SNP chips and genotype probabilities for untyped relatives has been calculated based on Icelandic genealogy. The sequencing was done using Illumina standard TruSeq methodology to a mean depth of 35 (SD 8). Autosomal SNPs and INDEL's were identified using the Genome Analysis Toolkit version 3.4.0[35]. Variants that did not pass quality control were excluded from the analysis according to GATK best practices[7]. Information about haplotype sharing was used to improve variant genotyping, taking advantage of the fact that all sequenced individuals had also been chip-typed and long-range phased[36].

The informativeness of genotype imputation (imputation information) was estimated by the ratio of the variance of imputed expected allele counts and the variance of the actual allele counts:

$$\frac{\mathrm{Var}(E(\theta|\text{chip data}))}{\mathrm{Var}(\theta)}$$

where θ is the allele count. Here, Var($E(\theta$/chip data)) is estimated by the observed variance in the imputed expected counts and $Var(\theta)$ was estimated by $p(1-p)$, where p is the allele frequency.

Variants were annotated using Ensembl release 80 and Variant Effect Predictor (VEP) version 2.8[10]. A total of 32.5 million variants passed the quality threshold and were imputed into 151,677 Icelanders who had been genotyped using Illumina chips.

To account for inflation in test statistics due to cryptic relatedness and stratification, we applied the method of LD score regression[37]. With a set of 1.1M variants we regressed the χ^2 statistics from our GWAS scan against LD score and used the intercept as a correction factor. The LD scores were downloaded from a LD score database (ftp://atguftp.mgh.harvard.edu/brendan/1k_eur_r2_hm3snps_se_weights.RDS; accessed 23.06.2015). The estimated correction factor for atrial fibrillation based on LD score regression was 1.39 for the additive model in the Icelandic sample and 1.04 in UK Biobank.

Genotyping in the UK biobank data. In the UK Biobank, genotyping was performed using a custom-made Affymetrix chip, UK BiLEVE Axiom[38], in the first 50,000 participants, and with Affymetrix UK Biobank Axiom array in the remaining participants[39]; 95% of the signals are on both chips. Imputation was performed by Wellcome Trust Centre for Human Genetics using a combination of 1000Genomes phase 3[40], UK10K[41], and Haplotype Reference Consortium (HRC) reference panels[42], for up to 92,693,895 SNPs[43].

Single-variant genotyping. For genotyping of single variants in atrial fibrillation sample sets from the Vanderbilt registry, FOURIER trial, and Norway, we used the Centaurus (Nanogen) or KASP platforms.

Statistical analysis. We performed a meta-analysis on 14,710 atrial fibrillation cases and 373,897 controls from Iceland and 14,792 atrial fibrillation cases and 393,868 controls from the UK Biobank. We used logistic regression to test for association between SNPs and atrial fibrillation and other phenotypes in the Icelandic study, treating phenotype status as the response and allele count as a covariate. We used allele counts from genotyping or integrated over possible genotype counts based on imputation. Other available individual characteristics that correlate with phenotype status were also included in the model as nuisance variables. In Iceland these covariates were: sex, county of birth, current age, or age at death (first and second order terms included), blood sample availability for the individual and an indicator function for the overlap of the lifetime of the individual with the time span of phenotype collection. In the UK biobank study 40 principal components were used to adjust for population stratification and age and sex were included as covariates in the logistic regression model. Only white British individuals were included in the study. For the meta-analysis we used a fixed-effects inverse variance method[44] based on effect estimates and standard errors from the Icelandic and the UK Biobank study. Only sequence variants from the Haplotype Reference Consortium panel (HRC)[42] were included in the meta-analysis and variants from deCODE and the UK Biobank imputation were matched on position and alleles. Standard errors were calculated in the following way:

For a P-value smaller than 1 we calculate the standard error as follows:

$$P = 2\Phi(z) = 2\Phi\left(\frac{\beta}{\sigma}\right).$$

Solving for σ gives

$$\sigma = \frac{\beta}{\Phi^{-1}\left(\frac{P}{2}\right)}$$

If $P = 1$, then $\Phi^{-1}\left(\frac{P}{2}\right) = 0$ and the above method breaks down. In this case we use data from other markers to estimate the relationship between allele frequency (f) and imputation information (I) and σ as follows:

$$Var(\beta) = \sigma^2 \propto \frac{1}{N} f(1-f) \propto \frac{1}{I} f(1-f)$$

Sample size (N) is proportional to imputation information (I) if we are always basing the analysis on the same set of individuals. Therefore, if we fit the following linear model:

$$\log(\sigma^2) = \gamma_1 + \gamma_I \log(I) + \gamma_f \log(f(1-f))$$

for a subset of 100,000 markers spread over the genome with MAF ranging close to uniformly between 0.1 and 50% and info between 0.9 and 1 and pick the subset of markers with $P < 0.9$ then we can predict σ for a marker with P close to 1.

We corrected the threshold for genome-wide significance for multiple testing with a weighted Bonferroni adjustment using as weights the enrichment of variant classes with predicted functional impact among association signals estimated from the Icelandic data[11].

With 32,463,443 sequence variants in the Icelandic data the weights given in Sveinbjornsson et. al. were rescaled to control the family-wise error rate. This yielded significance thresholds of 2.6×10^{-7} for high-impact variants ($N = 8464$) and 5.1×10^{-8} for moderate-impact variants ($N = 149{,}983$).

Conditional analysis of the region around novel atrial fibrillation variants, was performed by adding the top variant or variants as a covariate while testing every SNP in the region for association with atrial fibrillation in the Icelandic data.

We tested atrial fibrillation variants for association with 122 ECG measurements using linear regression, treating the ECG measurement as the response and the genotype as the covariate. Following the procedures described in Thorolfsdottir et al.[8], ECG measurements were adjusted for sex, year of birth, and age at measurement and were subsequently standardized to have a normal distribution. For individuals with multiple ECG measurements, the mean standardized value was used. We assume that the quantitative measurements follow a normal distribution with a mean that depends linearly on the expected allele at the variant and a variance-covariance matrix proportional to the kinship matrix[45]. The Benjamini–Hochberg FDR procedure controlling the FDR at 0.05 at each marker was used to account for multiple testing.

Expression analysis in cardiac tissue. RNA sequencing was performed on samples from cardiac right atrium of 167 Icelanders (see Supplementary Table 9, for subject characteristics). The samples were obtained during cardiothoracic surgery at Landspitali, The National University Hospital, in Reykjavik, Iceland. In the case of the splice-donor variant in *RPL3L* (c.1167+1G>A), the RNA samples from cardiac atria were used to identify a novel isoform and quantify expression at the transcript level. RNA sequencing libraries were inspected for sequencing and alignment integrity using parameters retrieved from RNA-SeQC[46], Picard CollectRnaSeqMetrics (http://broadinstitute.github.io/picard/), and FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Genotype concordance was determined by comparing imputed genotypes to those derived from RNA-seq. Genome alignments were found using STAR[47] aligning to GRCh38 with ensemble v87[48] gene annotations. Alignments of RNA-seq reads of carriers of the variant contained several reads that spliced over exon 9 in transcript ENST00000268661 of *RPL3L*. Neither of the two other transcripts of *RPL3L* showed any expression in all samples (See Supplementary Fig. 2). To assess quantitatively the effect of the variant on the isoform usage we created the transcript sequence for the novel isoform, added it to the ensemble v87 transcriptome, and re-quantified all samples using kallisto[49] and the modified transcriptome. The expression of the annotated and novel transcript was corrected w.r.t. the size factor computed from the gene

expression analysis. Finally the proportion of novel isoform usage was computed by dividing the estimated expression of the novel isoform with the sum of the expression of both isoforms. Due to the small number of carriers, two samples out of 167, we opted for a conservative test for computing the significance of the observed event, that the carriers have a ratio of 50% vs near 0% for non-carriers. The test used was the two-sided Mann–Whitney *U* test, which only takes the relative ranks of the samples into account and not the underlying values. The *P*-value computed was $P = 0.0052$, the lowest possible *P*-value that can be obtained using this statistical test with $n_1 = 165$ and $n_2 = 2$.

Data availability. The Icelandic population WGS data has been deposited at the European Variant Archive under accession code PRJEB8636. The authors declare that the data supporting the findings of this study are available within the article, its Supplementary Data files and upon request.

Received: 12 March 2018 Accepted: 11 May 2018
Published online: 12 June 2018

References

1. Chugh, S. S. et al. Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. *Circulation* **129**, 837–847 (2014).
2. Calenda, B. W., Fuster, V., Halperin, J. L. & Granger, C. B. Stroke risk assessment in atrial fibrillation: risk factors and markers of atrial myopathy. *Nat. Rev. Cardiol.* **13**, 549–559 (2016).
3. Goldberger, J. J. et al. Evaluating the atrial myopathy underlying atrial fibrillation: identifying the arrhythmogenic and thrombogenic substrate. *Circulation* **132**, 278–291 (2015).
4. Bapat, A., Anderson, C. D., Ellinor, P. T. & Lubitz, S. A. Genomic basis of atrial fibrillation. *Heart* **104**, 201–206 (2017).
5. Holm, H. et al. A rare variant in *MYH6* is associated with high risk of sick sinus syndrome. *Nat. Genet.* **43**, 316–320 (2011).
6. Gudbjartsson, D. F. et al. A frameshift deletion in the sarcomere gene *MYL4* causes early-onset familial atrial fibrillation. *Eur. Heart J.* **38**, 27–34 (2017).
7. Gudbjartsson, D. F. et al. Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* **47**, 435–444 (2015).
8. Thorolfsdottir, R. B. et al. A missense variant in *PLEC* increases risk of atrial fibrillation. *J. Am. Coll. Cardiol.* **70**, 2157–2168 (2017).
9. Sudlow, C. et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
10. McLaren, W. et al. Deriving the consequences of genomic variants with the Ensembl API and SNP effect predictor. *Bioinformatics* **26**, 2069–2070 (2010).
11. Sveinbjornsson, G. et al. Weighting sequence variants based on their annotation increases power of whole-genome association studies. *Nat. Genet.* **48**, 314–317 (2016).
12. Van Raay, T. J., Connors, T. D., Klinger, K. W., Landes, G. M. & Burn, T. C. A novel ribosomal protein L3-like gene (*RPL3L*) maps to the autosomal dominant polycystic kidney disease gene region. *Genomics* **37**, 172–176 (1996).
13. Seeger, T. S. et al. Myozap, a novel intercalated disc protein, activates serum response factor-dependent signaling and is required to maintain cardiac function in vivo. *Circ. Res.* **106**, 880–890 (2010).
14. Moncayo-Arlandi, J. & Brugada, R. Unmasking the molecular link between arrhythmogenic cardiomyopathy and Brugada syndrome. *Nat. Rev. Cardiol.* **14**, 744–756 (2017).
15. McCann, K. L. & Baserga, S. J. Genetics. Mysterious ribosomopathies. *Science* **341**, 849–850 (2013).
16. Raiser, D. M., Narla, A. & Ebert, B. L. The emerging importance of ribosomal dysfunction in the pathogenesis of hematologic disorders. *Leuk. Lymphoma* **55**, 491–500 (2014).
17. Sauert, M., Temmel, H. & Moll, I. Heterogeneity of the translational machinery: variations on a common theme. *Biochimie* **114**, 39–47 (2015).
18. Xue, S. & Barna, M. Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nat. Rev. Mol. Cell Biol.* **13**, 355–369 (2012).
19. Chaillou, T., Zhang, X. & McCarthy, J. J. Expression of muscle-specific ribosomal protein L3-like impairs myotube growth. *J. Cell. Physiol.* **231**, 1894–1902 (2016).
20. Choi, Y. & Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**, 2745–2747 (2015).
21. Ben-Shem, A. et al. The structure of the eukaryotic ribosome at 3.0 A resolution. *Science* **334**, 1524–1529 (2011).
22. Frank, D. et al. Mice with cardiac-restricted overexpression of Myozap are sensitized to biomechanical stress and develop a protein-aggregate-associated cardiomyopathy. *J. Mol. Cell. Cardiol.* **72**, 196–207 (2014).
23. Rampazzo, A., Calore, M., van Hengel, J. & van Roy, F. Intercalated discs and arrhythmogenic cardiomyopathy. *Circ. Cardiovasc. Genet.* **7**, 930–940 (2014).
24. Tao, Y. et al. Pitx2, an atrial fibrillation predisposition gene, directly regulates ion transport and intercalated disc genes. *Circ. Cardiovasc. Genet.* **7**, 23–32 (2014).
25. Castañón, M. J., Walko, G., Winter, L. & Wiche, G. Plectin–intermediate filament partnership in skin, skeletal muscle, and peripheral nerve. *Histochem. Cell. Biol.* **140**, 33–53 (2013).
26. Andra, K. et al. Targeted inactivation of plectin reveals essential function in maintaining the integrity of skin, muscle, and heart cytoarchitecture. *Genes Dev.* **11**, 3143–3156 (1997).
27. Elliott, P. & Peakman, T. C. The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int. J. Epidemiol.* **37**, 234–244 (2008).
28. Sabatine, M. S. et al. Rationale and design of the Further cardiovascular OUtcomes Research with PCSK9 Inhibition in subjects with Elevated Risk trial. *Am. Heart J.* **173**, 94–101 (2016).
29. Sabatine, M. S. et al. Evolocumab and clinical outcomes in patients with cardiovascular disease. *N. Engl. J. Med.* **376**, 1713–1722 (2017).
30. Jacobsen, B. K., Eggen, A. E., Mathiesen, E. B., Wilsgaard, T. & Njolstad, I. Cohort profile: the Tromso Study. *Int. J. Epidemiol.* **41**, 961–967 (2012).
31. Krokstad, S. et al. Cohort Profile: the HUNT Study, Norway. *Int. J. Epidemiol.* **42**, 968–977 (2013).
32. Adams, H. P. Jr et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in acute stroke treatment. *Stroke* **24**, 35–41 (1993).
33. Zhou, S. H., Helfenbein, E. D., Lindauer, J. M., Gregg, R. E. & Feild, D. Q. Philips QT interval measurement algorithms for diagnostic, ambulatory, and patient monitoring ECG applications. *Ann. Noninvasive Electrocardiol.* **14** (Suppl 1), S3–S8 (2009).
34. Lindauer, J., Gregg, R., Helfenbein, E., Shao, M. & Zhou, S. Global QT measurements in the Philips 12-lead algorithm. *J. Electrocardiol.* **38**, 90 (2005).
35. McKenna, A. et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
36. Kong, A. et al. Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat. Genet.* **40**, 1068–1075 (2008).
37. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
38. Wain, L. V. et al. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir. Med* **3**, 769–781 (2015).
39. Welsh, S., Peakman, T., Sheard, S. & Almond, R. Comparison of DNA quantification methodology used in the DNA extraction protocol for the UK Biobank cohort. *BMC Genomics* **18**, 26 (2017).
40. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
41. Walter, K. et al. The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90 (2015).
42. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
43. Bycroft, C. et al. Genome-wide genetic data on ~500,000 UK Biobank participants. https://doi.org/10.1101/166298. (2017).
44. Mantel, N. & Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl Cancer Inst.* **22**, 719–748 (1959).
45. Benonisdottir, S. et al. Epigenetic and genetic components of height regulation. *Nat. Commun.* **7**, 13490 (2016).
46. DeLuca, D. S. et al. RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* **28**, 1530–1532 (2012).
47. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
48. Yates, A. et al. Ensembl 2016. *Nucleic Acids Res.* **44**, D710–D716 (2016).
49. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).

Acknowledgements

We thank all the study subjects for their valuable participation as well as our colleagues that contributed to data collection, sample handling and genotyping. This research has been conducted using the UK Biobank Resource under Application Number '24711'. Dr. Darbar is supported by the NIH R01 HL092217, HL092217-07S1, R01 HL124935, and R01HL138737 grant.

Author contributions

R.B.T., G.S., P.S., D.O.A., H.H., D.F.G., U.T., and K.S. designed the study and interpreted the results. R.B.T., H.H., G.S. and S.J. drafted the manuscript. R.B.T., G.S., P.S., D.F.G., and E.V.I. performed bioinformatic and statistical analysis in the discovery study and

joint data analysis. R.B.T., O.B.D., R.P.K., G.T., A.H., S.G., J.T.S., I.J., U.T., and H.H. collected and characterized Icelandic phenotype data. J.B.N., V.T., O.L.H., F.W.A., D.M.R., D.D., T.R.P., M.S.S., C.J.W., M.L.L., and K.H. designed and managed individual studies and characterized phenotype data. B.T. and A.S.V. collected and managed samples from cardiac atrial tissue during cardiothoracic surgery. G.H.H., P.M. and B.V.H performed bioinformatic and statistical analysis of RNA sequencing data. R.B.T., G.S., P.S., J.B.N., S.J., G.H.H., P.M., E.V.I., O.B.D., R.P.K., G.T., A.H., S.G., G.N., S.R., B.T., A.S.V., J.T.S., V.T., O.L.H., F.W.A., D.M.R., D.D., T.R.P., M.S.S., C.J.W., M.L.L., B.V. H., I.J., K.H., D.O.A., U.T., D.F.G., H.H., and K.S. reviewed and approved the manuscript. K.S. supervised the study.

Additional information

Supplementary information accompanies this paper at https://doi.org/10.1038/s42003-018-0068-9.

Competing interests: The following authors affiliated with deCODE genetics/Amgen, Inc. are employed by the company: R.B.T., G.S., P.S., S.J., G.H.H., P.M., E.V.I., O.B.D., R.P.K., G.T., A.H., S.G., G.N., S.R., V.T., B.V.H., I.J., D.O.A., U.T., D.F.G., H.H., and K.S. T.R.P. has received consulting and/or speakers honoraria from Amgen, Sanofi, and Merck (all minor). M.S.S. has received research grant support through Brigham and Women's Hospital from Amgen, AstraZeneca, Daiichi-Sankyo, Eisai, GlaxoSmithKline, Intarcia, Janssen Research and Development, MedImmune, Merck, Novartis, Pfizer, Poxel, and Takeda and consulting honoraria from Amgen, CVS Caremark, Esperion, Intarcia, Ionis, Janssen Research and Development, MedImmune, and Merck. B.T., A.S.V., J.B.N., J.T.S., O.L.H., F.W.A., D.M.R., D.D., C.J.W., M.-L.L., and K.H. have no relationship with industry to disclose.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Rosa B. Thorolfsdottir[1,2], Gardar Sveinbjornsson[1], Patrick Sulem[1], Jonas B. Nielsen[3,4], Stefan Jonsson[1], Gisli H. Halldorsson[1], Pall Melsted[1,5], Erna V. Ivarsdottir[1], Olafur B. Davidsson[1], Ragnar P. Kristjansson[1], Gudmar Thorleifsson[1], Anna Helgadottir[1], Solveig Gretarsdottir[1], Gudmundur Norddahl[1], Sridharan Rajamani[1], Bjarni Torfason[2,6], Atli S. Valgardsson[6], Jon T. Sverrisson[7], Vinicius Tragante[1,8], Oddgeir L. Holmen[9,10,11], Folkert W. Asselbergs[8,12,13,14], Dan M. Roden[15], Dawood Darbar[16], Terje R. Pedersen[17], Marc S. Sabatine[18], Cristen J. Willer[3,4,19], Maja-Lisa Løchen[20], Bjarni V. Halldorsson[1,21], Ingileif Jonsdottir[1,2,22], Kristian Hveem[9,10,23], David O. Arnar[1,2,24], Unnur Thorsteinsdottir[1,2], Daniel F. Gudbjartsson[1,5], Hilma Holm[1] & Kari Stefansson[1,2]

[1]deCODE genetics/Amgen, Inc., Reykjavik, Iceland. [2]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [3]Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, MI, USA. [4]Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA. [5]School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland. [6]Department of Cardiothoracic Surgery, Landspitali University Hospital, Reykjavik, Iceland. [7]Department of Medicine, Akureyri Regional Hospital, Akureyri, Iceland. [8]Department of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, University of Utrecht, Utrecht, The Netherlands. [9]HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology, Levanger, Norway. [10]K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, Norwegian University of Science and Technology, Trondheim, Norway. [11]Department of Cardiology, St. Olav's University Hospital, Trondheim, Norway. [12]Durrer Center for Cardiovascular Research, Netherlands Heart Institute, Utrecht, The Netherlands. [13]Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, UK. [14]Farr Institute of Health Informatics Research and Institute of Health Informatics, University College London, London, UK. [15]Departments of Medicine, Pharmacology, and Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA. [16]Division of Cardiology, Department of Medicine, University of Illinois at Chicago, Chicago, IL, USA. [17]Center For Preventive Medicine, Oslo University Hospital and Medical Faculty, University of Oslo, Oslo, Norway. [18]TIMI Study Group, Division of Cardiovascular Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [19]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA. [20]Department of Community Medicine, UiT The Arctic University of Norway, Tromsø, Norway. [21]Reykjavik University, Reykjavik, Iceland. [22]Department of immunology, Landspitali University Hospital, Reykjavik, Iceland. [23]Department of Medicine, Levanger Hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway. [24]Department of Medicine, Landspitali University Hospital, Reykjavik, Iceland

Paper III

nature genetics

LETTERS

https://doi.org/10.1038/s41588-018-0171-3

Biobank-driven genomic discovery yields new insight into atrial fibrillation biology

Jonas B. Nielsen[1,2,23], Rosa B. Thorolfsdottir[3,4,23], Lars G. Fritsche[1,5,6,7,23], Wei Zhou[1,7,23], Morten W. Skov[8,23], Sarah E. Graham[1,2,23], Todd J. Herron[9,23], Shane McCarthy[10,23], Ellen M. Schmidt[11,23], Gardar Sveinbjornsson[3,23], Ida Surakka[1,2], Michael R. Mathis[12], Masatoshi Yamazaki[13], Ryan D. Crawford[7], Maiken E. Gabrielsen[5,6], Anne Heidi Skogholt[5,6], Oddgeir L. Holmen[5,6,14], Maoxuan Lin[1,2], Brooke N. Wolford[1,7], Rounak Dey[11], Håvard Dalen[15,16,17], Patrick Sulem[3], Jonathan H. Chung[10], Joshua D. Backman[10], David O. Arnar[3,4,18], Unnur Thorsteinsdottir[3,4], Aris Baras[10], Colm O'Dushlaine[10], Anders G. Holst[8], Xiaoquan Wen[11], Whitney Hornsby[1], Frederick E. Dewey[10], Michael Boehnke[11], Sachin Kheterpal[12], Bhramar Mukherjee[11], Seunggeun Lee[11], Hyun M. Kang[11], Hilma Holm[3], Jacob Kitzman[2], Jordan A. Shavit[19], José Jalife[1,9,20], Chad M. Brummett[12], Tanya M. Teslovich[10], David J. Carey[21], Daniel F. Gudbjartsson[3,22], Kari Stefansson[3,4], Gonçalo R. Abecasis[6,11*], Kristian Hveem[5,6,15*] and Cristen J. Willer[1,2,7*]

To identify genetic variation underlying atrial fibrillation, the most common cardiac arrhythmia, we performed a genome-wide association study of >1,000,000 people, including 60,620 atrial fibrillation cases and 970,216 controls. We identified 142 independent risk variants at 111 loci and prioritized 151 functional candidate genes likely to be involved in atrial fibrillation. Many of the identified risk variants fall near genes where more deleterious mutations have been reported to cause serious heart defects in humans (*GATA4, MYH6, NKX2-5, PITX2, TBX5*)[1], or near genes important for striated muscle function and integrity (for example, *CFL2, MYH7, PKP2, RBM20, SGCG, SSPN*). Pathway and functional enrichment analyses also suggested that many of the putative atrial fibrillation genes act via cardiac structural remodeling, potentially in the form of an 'atrial cardiomyopathy'[2], either during fetal heart development or as a response to stress in the adult heart.

We tested association between 34,740,186 genetic variants (minor allele frequency (MAF) $>2.5\times10^{-5}$) and atrial fibrillation, comparing a total of 60,620 cases and 970,216 controls of European ancestry from six contributing studies (The Nord-Trøndelag Health Study (HUNT), deCODE, the Michigan Genomics Initiative (MGI), DiscovEHR, UK Biobank, and the AFGen Consortium) (Supplementary Table 1). We identified 111 genomic regions with at least 1 genetic variant associated with atrial fibrillation ($P<5\times10^{-8}$). Of these, 80 loci have not previously been reported (Fig. 1, Supplementary Fig. 1, and Supplementary Tables 2 and 3). Based on approximate stepwise conditional analyses[3], we identified 31 additional genetic risk variants that demonstrated genome-wide statistically significant association with atrial fibrillation (Supplementary Table 4) that were nearby but independent of the 111 index variants (linkage disequilibrium (LD) r^2 <0.10). We applied the widely used genome-wide association study (GWAS) P value significance threshold of $P<5\times10^{-8}$. Had we applied a more stringent threshold of $P<5\times10^{-9}$ (ref. [4]), we would identify 94 loci, 63 of which have not been previously reported (Supplementary Table 2). We found that the total genome-wide genetic variation captured in this study explained 11.2% (s.e.m. 1.4%) of the variation in atrial fibrillation

[1]Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, MI, USA. [2]Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA. [3]deCODE genetics/Amgen, Inc., Reykjavik, Iceland. [4]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [5]HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology, Levanger, Norway. [6]K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, Norwegian University of Science and Technology, Trondheim, Norway. [7]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA. [8]Laboratory of Molecular Cardiology, Department of Cardiology, The Heart Centre, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark. [9]Department of Internal Medicine, Center for Arrhythmia Research, University of Michigan, Ann Arbor, MI, USA. [10]Regeneron Genetics Center, Tarrytown, NY, USA. [11]Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA. [12]Department of Anesthesiology, University of Michigan, Ann Arbor, MI, USA. [13]Medical Device Development and Regulation Research Center, The University of Tokyo, Tokyo, Japan. [14]Department of Cardiology, St. Olav's University Hospital, Trondheim, Norway. [15]Department of Medicine, Levanger Hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway. [16]Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim, Norway. [17]Department of Cardiology, St. Olav's University Hospital, Trondheim University Hospital, Trondheim, Norway. [18]Department of Medicine, Landspitali – National University Hospital, Reykjavik, Iceland. [19]Department of Pediatrics and Communicable Diseases, University of Michigan, Ann Arbor, MI, USA. [20]Fundación Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid, Spain. [21]Geisinger Health System, Danville, PA, USA. [22]School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland. [23]These authors contributed equally: Jonas B. Nielsen, Rosa B. Thorolfsdottir, Lars G. Fritsche, Wei Zhou, Morten W. Skov, Sarah E. Graham, Todd J. Herron, Shane McCarthy, Ellen M. Schmidt, Gardar Sveinbjornsson. *e-mail: goncalo@umich.edu; kristian.hveem@ntnu.no; cristen@umich.edu

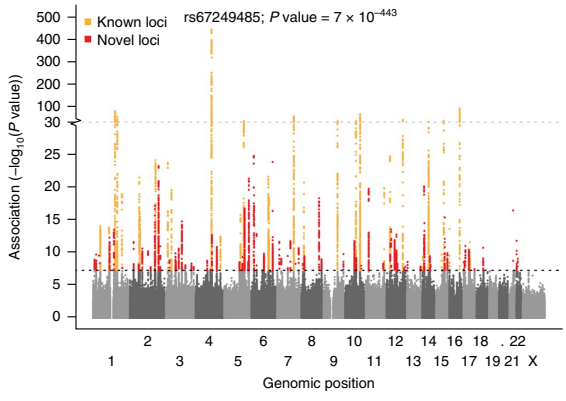


Fig. 1 | Manhattan plot showing known (orange) and novel (red) loci associated with atrial fibrillation. A total of 34,740,186 genetic variants (each represented by a dot) were tested, comparing 60,620 atrial fibrillation cases and 970,216 controls free of atrial fibrillation. The *x* axis represents the genome in physical order, and the *y* axis represents *P* values ($-\log_{10}$(*P* value)) of association. The black horizontal dotted line represents a Bonferroni-corrected threshold of statistical significance corresponding to 1,000,000 independent tests ($P<5\times10^{-8}$).

(h^2_{SNP} heritability). This is consistent with a recent report of 11.4%[5]. When combining the 111 locus index variants and the 31 additional genetic risk variants, we found that they explained 4.6% of the variation in atrial fibrillation.

Of the 35 loci previously reported for atrial fibrillation (Supplementary Table 3), we identified genome-wide significant association ($P<5\times10^{-8}$) at 31 (89%) after excluding results from the previously published AFGen Consortium (Supplementary Table 5)[6]. The four loci not captured comprised three loci discovered in East Asian populations (*KCNIP1*, *NEBL*, *CUX2*) and one missense variant (*PLEC*) for which we did not have data[7]. To further test the validity of our findings, we performed a heterogeneity test for the 111 index variants across the 6 contributing studies. Of the 111 index variants, only 2 index variants demonstrated evidence for heterogeneity of effect size across the 6 contributing studies when correcting for multiple testing ($P<0.05/111=4.5\times10^{-4}$) (Supplementary Table 2). Both of these index variants represent loci that have previously been established as associated with atrial fibrillation across multiple studies (near *PRRX1*, *PITX2*) (Supplementary Table 3). These findings demonstrate a high external validity of our results.

To understand the biology underlying the 111 atrial fibrillation-associated loci, we employed a number of approaches, including 'Data-driven Expression Prioritized Integration for Complex Traits' (DEPICT)[8] to identify cell types and tissues in which atrial fibrillation genes are likely to be expressed. Based on 37,427 human microarray expression samples from 209 different tissues and cell types, we observed a statistically significant enrichment for atrial ($P=2.4\times10^{-5}$), atrial appendage ($P=2.8\times10^{-5}$), heart ($P=5.2\times10^{-5}$), and ventricular tissues ($P=1.1\times10^{-4}$) (Fig. 2a and Supplementary Table 6). We further applied DEPICT to detect gene sets that were enriched for genes at atrial fibrillation-associated loci. Of the 14,461 gene sets we tested, 889 were enriched (false discovery rate (FDR) < 0.05; Fig. 2b and Supplementary Table 7). The highlighted gene sets point to biological processes related to cardiac development and morphology along with structural remodeling of the myocardium. These findings are consistent with recent reports that have linked atrial fibrillation with rare coding variants in the sarcomere genes *MYH6* and *MYL4* and in the multidomain cytoskeletal linking protein *PLEC*, along with more common coding variants in *TTN*, essential for the passive elasticity of heart and skeletal muscle[7,9–11].

Although we could identify protein-altering variants at 21 loci, comprising either the index variant ($n=2$ loci) or a variant in high LD (r^2) with the index variant ($n=19$ loci; Supplementary Table 8), we noted that most associated risk variants are in the non-coding genome. To assess the potential function of associated non-coding variants, we tested for enrichment of atrial fibrillation-associated variants with a variety of regulatory features, including DNase I hypersensitive sites, histone methylation marks, transcription factor binding sites, and chromatin states in a variety of cell and tissue types available from Roadmap Epigenomics[12] using 'Genomic Regulatory Elements and Gwas Overlap algoRithm' (GREGOR)[13]. This method tests whether the number of atrial fibrillation-associated index variants, or their LD proxies, overlap with the corresponding regulatory feature more often than expected when compared to control sets. Of 785 combinations of regulatory features and tissues examined (Supplementary Table 9), we found that atrial fibrillation-associated variants were most strongly associated with features in adult and fetal heart: active enhancers as indicated by H3K27ac in right atrium ($P=2\times10^{-33}$; 2.9×enrichment); H3K27ac in left ventricle ($P=3\times10^{-33}$; 2.6×enrichment); and in fetal heart tissue we found strong enrichment with H3K4me1 ($P=9\times10^{-27}$; 2.0×enrichment) and open chromatin as measured by DNase hypersensitivity ($P=2\times10^{-26}$; 2.1×enrichment) (Fig. 2c, Supplementary Fig. 2 and Supplementary Table 9). This suggests that some atrial fibrillation loci are important in transcriptional regulation in the adult heart, in development of the fetal heart, or both.

To further enhance the biological understanding of the atrial fibrillation-associated loci, we identified candidate functional genes. There were 3,048 genes or transcripts for which the transcribed region overlapped (see Online Methods) at least 1 variant in the 111 loci. We prioritized biological candidate genes that: (1) harbored a protein-altering index variant itself or in high LD (r^2 >0.80; Supplementary Table 8); (2) had expression levels that were associated and colocalized with atrial fibrillation-associated variants ($P<1.14\times10^{-9}$ in Genotype-Tissue Expression (GTEx) consortium data)[14]; (3) were highlighted by DEPICT (FDR <0.05; Supplementary Table 10); or (4) were nearest to the index variant in a locus. Using these criteria, we prioritized 151 candidate genes (Supplementary Tables 2 and 11).

To identify tissues in which the 151 prioritized candidate genes showed enhanced expression, we used 'Tissue Specific Expression Analysis' (TSEA)[15] and found enrichment in heart ($P=1\times10^{-16}$), blood vessel (9×10^{-13}) muscle tissues ($P=7\times10^{-11}$). To assess the empirical significance of these results, we performed 1,000 permutations of the same number of genes selected: (1) randomly from the genome, and (2) subsets of the 3,048 genes within the 111 atrial fibrillation loci. We determined that the observed TSEA *P* values were substantially more significant than expected by chance (Fig. 3). The finding of increased expression of these genes in heart support that the genes we prioritized are strong candidates for being involved in atrial fibrillation.

Interestingly, we identified as functional candidates at least 18 genes likely to be involved in cardiac and skeletal muscle function and integrity (*AKAP6*, *CFL2*, *MYH6*, *MYH7*, *MYO18B*, *MYO1C*, *MYOCD*, *MYOT*, *MYOZ1*, *MYPN*, *PKP2*, *RBM20*, *SGCA*, *SSPN*, *SYNPO2L*, *TTN*, *TTN-AS*, *WIPF1*); these included *SGCG*, which has been associated with muscular dystrophy[16], *RBM20*, which has been associated with dilated cardiomyopathy[17], and *PKP2*, which has been associated with arrhythmogenic right ventricular cardiomyopathy[18]. We identified at least 13 genes likely to be involved in mediation of developmental events (*ARNT2*, *EPHA3*, *FGF5*, *GATA4*, *GTF2I*, *HAND2*, *LRRC10*, *NAV2*, *NKX2-5*, *PITX2*, *SLIT3*,

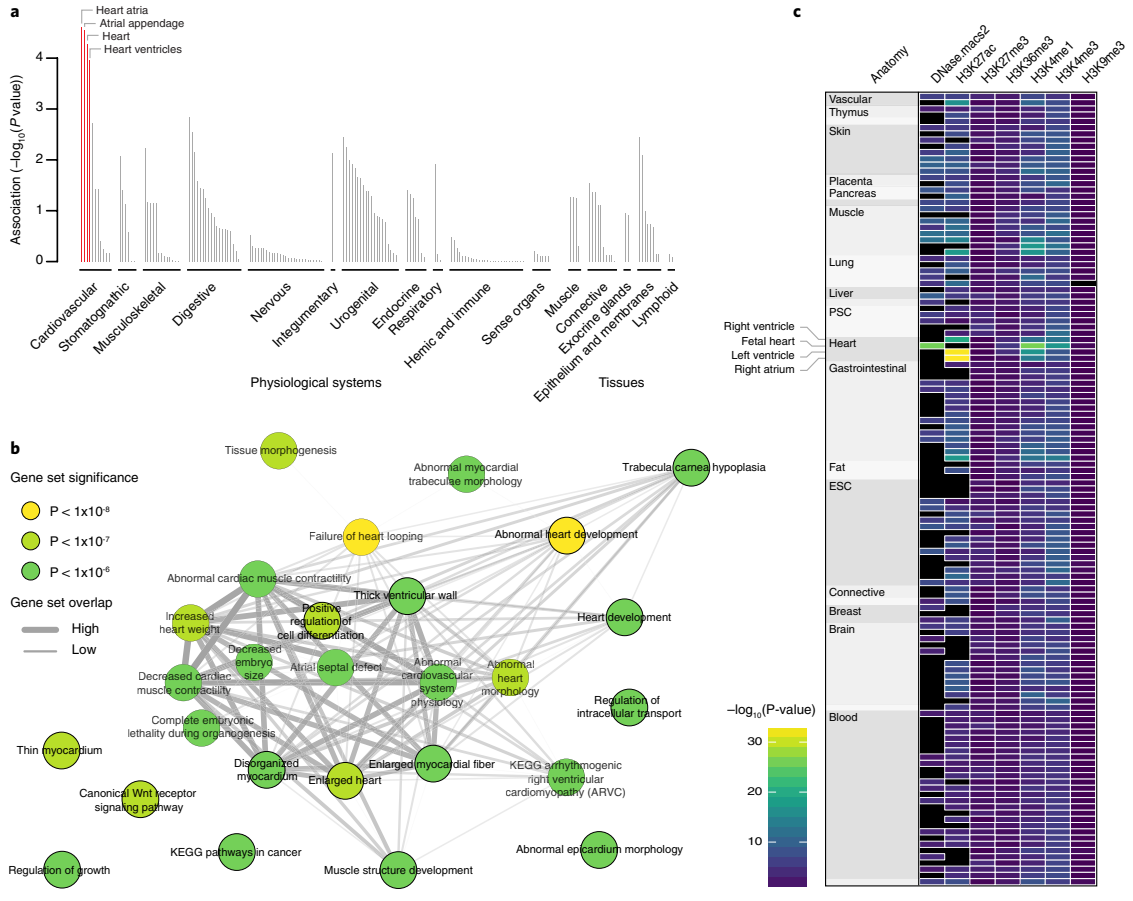


Fig. 2 | Tissues, reconstituted gene sets, and regulatory elements implicated in atrial fibrillation. a, Based on expression patterns across 37,427 human mRNA microarrays, DEPICT predicted genes within atrial fibrillation-associated loci to be highly expressed across various cardiac tissues. Tissues are grouped by type and significance. Red columns represent statistically significant tissues following Bonferroni correction ($P < 0.05/209 = 0.0002$). **b**, Top ($P < 1 \times 10^{-6}$) reconstituted gene sets (out of 826 with FDR < 0.05 and after exclusion of 'gene subnetworks') found by DEPICT to be significantly enriched for genes in atrial fibrillation-associated loci. Each node, colored according to the permutation *P* value, represents a gene set and the gray connecting lines represent pairwise overlap of genes within the gene sets. **c**, Heatmap indicating the overlap between atrial fibrillation-associated risk variants and regulatory elements across 127 Roadmap Epigenomics tissues (each represented by a row) using GREGOR. Black indicates no data. PSC, pluripotent stem cell; ESC, embryonic stem cell.

SOX15, *TBX5*) along with genes likely to be involved in intracellular calcium handling in the heart (*CALU*, *CAMK2D*, *CASQ2*, *PLN*), angiogenesis (*TNFSF12*, *TNFSF12-TNFSF13*), hormone signaling (*CGA*, *ESR2*, *IGF1R*, *NR3C1*, *THRB*), and function of cardiac ion channels (*HCN4, KCND3*, *KCNH2*, *KCNJ5*, *KCNN2*, *KCNN3*, *SCN10A*, *SCN5A*, *SLC9B1*).

We tested the 111 atrial fibrillation index variants for association with 123 electrocardiogram (ECG) parameters in 62,974 Icelanders in sinus rhythm, after exclusion of atrial fibrillation cases (Supplementary Fig. 3 and Supplementary Table 12). Sixty variants were associated with at least 1 ECG parameter when we controlled for an FDR of 0.05 at the variant level, 39 of which were novel atrial fibrillation variants, including many with substantial ECG effects, such as the variants near *NACA*, *THRB*, *CAMK2D*, *NKX2-5*, and *CDKN1A*. Many of the associated ECG parameters, including heart rate, P-wave duration, PR interval, and heart rate-corrected QT interval, are well-established intermediate phenotypes for atrial fibrillation[19–22]. Accordingly, several of the highlighted associations can be seen as indirect replications of the atrial fibrillation risk variants identified through GWASs. The results also indicate that many of the 111 atrial fibrillation risk variants act in the heart before atrial fibrillation. The type and direction of the ECG parameter associations might help inform the biology underlying the specific loci.

For the locus around index variant rs422068 on chromosome 14, our approach prioritized *MYH6* and *MYH7* as the most likely functional genes (Supplementary Table 2). A rare missense mutation in *MYH6* has in prior GWASs been associated with sick sinus syndrome[23], atrial fibrillation[7], and coarctation of the aorta[24], and several protein-altering variants in *MYH7* have been linked to hypertrophic cardiomyopathy[25]. *MYH6* and *MYH7* encode the molecular motors of cardiac muscle that transduce chemical energy from ATP hydrolysis into mechanical energy of each heartbeat. *MYH6* encodes α-myosin heavy chain (α-MyHC), which is the faster molecular motor of the thick filaments of the contractile

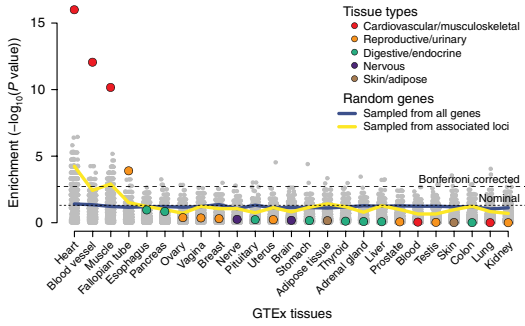


Fig. 3 | Significance of the expression enrichment for the atrial fibrillation candidate genes. This figure compares the tissue-specific gene expression enrichment for the 151 biological candidate genes (colored dots) to a null distribution derived by randomly selecting the same number of genes from the whole genome or from the associated loci. Tissue-specific gene expression data ($n = 25$ tissues) were obtained from Genotype-Tissue Expression (GTEx) consortium data. The gray dots represent the P values for each of the permutations for the randomized tests (1,000 for both sampling scenarios for each tissue), and the blue and yellow lines represent the per-tissue P value thresholds comparable to a false positive rate of 0.05.

apparatus in healthy adult atrial muscle[26]. On the other hand, *MYH7* encodes β-MyHC, a slower molecular motor[27], which is expressed only in the atria during cardiac development and not in the normal adult atria. It has been established that *MYH6* and *MYH7* are regulated in an inverse manner in the ventricles of the heart, and in heart failure and other cardiac disorders in humans, β-MyHC is upregulated, whereas α-MyHC is downregulated, resulting in diminution of cardiac performance[28]. Importantly, recent experiments have demonstrated that *MYH7* expression is elevated in atrial myocytes of patients with chronic atrial fibrillation as well as in an ovine model of chronic atrial fibrillation[29].

To explore potential mechanisms of *MYH6* and *MYH7* in atrial fibrillation, we developed an ischemic heart failure model for atrial fibrillation in rabbits. Ischemia was produced by chronic ligation of the left circumflex artery during thoracotomy with subsequent development of ischemic heart failure (>4 weeks postoperatively), profound left atrial dilation, and development of long-lasting atrial fibrillation following burst pacing (Fig. 4 and Supplementary Fig. 4). We found that MYH7 expression was only detectable in the heart failure remodeled left atrium (Fig. 4 and Supplementary Fig. 5). The control left atrium did not express detectable levels of MYH7 and exclusively expressed MYH6. More importantly, in the dilated left atrium, MYH7 expression was heterogeneously distributed (Fig. 4 and Supplementary Fig. 6), which thus resulted in contractile and metabolic heterogeneity, both of which are probably arrhythmogenic. Although the association between the rs422068 locus and atrial fibrillation could potentially be mediated through structural heart defects such as coarctation of the aorta or hypertrophic cardiomyopathy via genetic variation not captured in this study, it is likely that the heterogeneously distributed switch from the adult to the fetal isoform of myosin heavy chain that we observed in the dilated left atrium may predispose rabbit (and possibly human) hearts to developing long-lasting atrial fibrillation.

Next, we investigated whether any of the 151 biological candidate genes that we identified could potentially represent a novel drug target for already developed drugs or drugs undergoing development by querying the Drug Gene Interaction Database[30]. We found 1 or more potential drug or substance interactions for 32 of the 151 prioritized genes, totaling 475 drugs. Of these, 78 drugs targeting 14 genes are already known to be able to control or trigger atrial fibrillation or other cardiac arrhythmias (Supplementary Table 13). In addition to a number of drugs that could potentially impact atrial fibrillation via an effect on cardiac ion channels (for example, rottlerin, bepridil), including drugs already used for treating various neuropsychiatric disorders (for example, fosphenytoin, flunarizine), we also identified a number of anti-inflammatory drugs, including several glucocorticoids, and a cardiac-specific myosin activator (omecamtiv mecarbil), which is currently being tested for treatment of heart failure[31]. Whether the highlighted drugs can be used to treat or prevent atrial fibrillation requires further evaluation, but the findings can be used as a foundation for directing future functional experiments and clinical trials.

We constructed a polygenic risk score based on the 111 locus index variants and the 31 additional risk variants, identified through stepwise conditional analyses, weighted by effect estimates obtained from meta-analyses excluding the UK Biobank (Supplementary Table 14). We found that the risk score predicted prevalent atrial fibrillation in the UK Biobank with an unadjusted area under the receiver operator curve of 65%. We then used the polygenic risk score to test for association with 1,494 International Classification of Diseases (ICD) code-defined disease groups in UK Biobank participants of white British ancestry[32]. In addition to a strong association with atrial fibrillation ($P = 2 \times 10^{-920}$), we found association to additional mainly cardiovascular conditions ($P < 0.05/1{,}494 = 3.3 \times 10^{-5}$), including palpitations, heart valve disorders, heart failure, ischemic heart disease, and stroke (Supplementary Table 15 and Supplementary Fig. 7). However, when participants diagnosed with any type of cardiac arrhythmia ($n = 24{,}681$) were excluded from the analyses to avoid assessment bias (termed an exclusion phenome-wide association study)[33], the atrial fibrillation risk score was no longer associated with any ICD disease group (all $P > 3.3 \times 10^{-5}$). This suggests that the atrial fibrillation polygenic risk score is specific for atrial fibrillation and that the additional associations identified were mediated through atrial fibrillation, either as a result of a more thorough clinical examination (for example, heart valve disorders) or because atrial fibrillation is an intermediate step towards the disease (for example, stroke).

To examine the genetic impact on age of onset of atrial fibrillation, we generated a polygenic risk score ($n = 142$ markers) in which weights were based on information from all contributing studies (Supplementary Table 14) and tested for association with atrial fibrillation age of onset in the HUNT Study. In agreement with our previous report[11], we found that younger atrial fibrillation age of onset was associated with a higher genetic burden of atrial fibrillation (Supplementary Fig. 8). This finding supports previous epidemiological studies indicating that the risk of atrial fibrillation increases with decreasing atrial fibrillation age of onset in close relatives[34,35].

After acceptance of this manuscript, a GWAS of AF in 65,446 cases identifying 97 loci was published online[36]. We meta-analyzed the index variants from this report with samples from HUNT, deCODE, MGI, and DiscovEHR, comprising the largest possible independent dataset (up to 93,315 cases), and identified 24 genome-wide significant loci that are independent of the 111 we report here (Supplementary Table 16). We suggest that combining the initial 111 locus index variants, the 31 additional risk identified through approximate stepwise conditional analyses, and the 24 new locus index variants ($n = 166$ variants) would comprise the current optimal variant list for a polygenic risk score for atrial fibrillation (Supplementary Table 16).

In summary, we substantially increased the number of genome-wide significant risk variants for atrial fibrillation through a large GWAS meta-analysis. We highlighted genes important for function of cardiac ion channels and calcium signaling, along with cardiac transcription factors, which in turn could also affect the electrical properties of the myocardium[37,38], and in addition prioritized multiple atrial fibrillation functional candidate genes likely to be

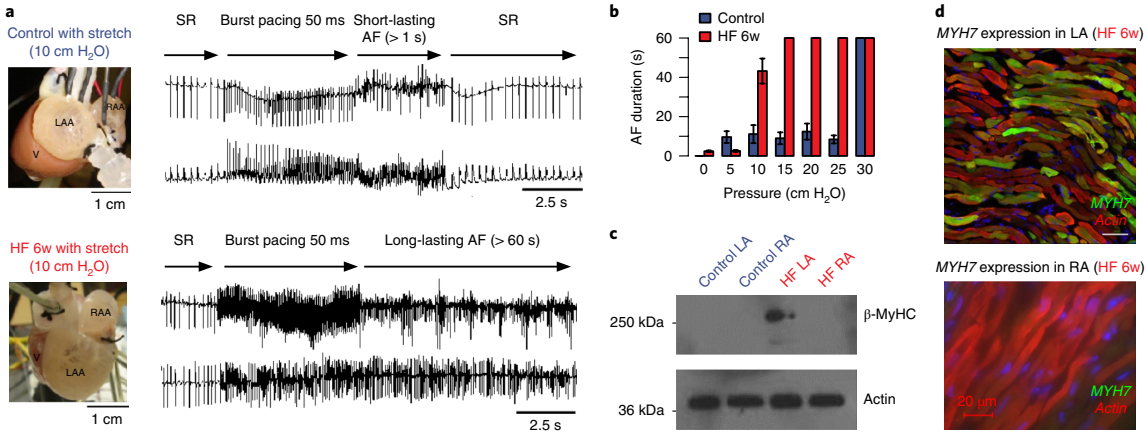


Fig. 4 | Atrial fibrillation is associated with heterogeneous changes in left atrial myosin isoform expression. a, Langendorff-perfused rabbit hearts from control (blue, top) or heart failure rabbits (red, bottom panel) were tested for atrial fibrillation inducibility and duration following burst pacing at 50 ms cycle length. Heart failure was induced by chronic left circumflex artery ligation and was allowed to develop over 6 weeks. During heart failure progression, severe left atrial hypertrophy occurred. **b**, Heart failure hearts ($n = 4$) developed long-lasting atrial fibrillation (>60 s) when intra-atrial pressure was increased to 10 cm H_2O. Control hearts ($n = 4$) did not develop long-lasting atrial fibrillation until intra-atrial pressure was increased to 30 cm H_2O. The colored bars represent mean atrial fibrillation duration and the black error bars represent the corresponding standard errors of the mean. All individual data points are shown in the more detailed Supplementary Fig. 4. **c**, Western blotting for MYH7 expression (β-MyHC protein) indicates MYH7 expression exclusively in the remodeled heart failure left atrium. The experiment was repeated for two independent heart failure animals and two control animals with similar results. An uncropped version of the western blot is shown as Supplementary Fig. 5. **d**, Immunostaining and confocal microscopy revealed heterogeneous MYH7 expression (green) in the heart failure left atrium. Consistent with western blotting data, the heart failure right atrium did not express MYH7. The experiment was repeated for two independent heart failure animals with similar results. Supplementary Figure 6 shows an additional image. LAA, left atrial appendage; RAA, right atrial appendage; V, ventricle; AF, atrial fibrillation; HF, heart failure; SR, sinus rhythm; w, week; LA, left atrium; RA, right atrium.

involved in structural integrity and function of heart and skeletal muscle. We performed pathway and functional enrichment analyses that highlighted fetal heart tissue and pathways related to cardiac development as important for developing atrial fibrillation. This might reflect that atrial fibrillation risk variants are acting in the developing heart or that the variants are important for reactivating fetal genes or pathways as a response to stress in the adult heart. We demonstrated an example of the latter; experiments in rabbits with heart failure and left atrial dilation identified a heterogeneous distributed molecular switch from the adult to the fetal isoform of myosin heavy chain, which resulted in contractile and functional heterogeneity that may predispose to initiation and maintenance of atrial fibrillation. These findings need confirmation but provide a foundation for directing future functional experiments to better understand the biology underlying atrial fibrillation.

URLs. GotCloud, https://genome.sph.umich.edu/wiki/GotCloud; Michigan Imputation Server, https://imputationserver.sph.umich.edu/index.html; METAL, http://genome.sph.umich.edu/wiki/METAL_Documentation; PLINK1.9, https://www.cog-genomics.org/plink/1.9; DEPICT, https://data.broadinstitute.org/mpg/depict/; COJO-GCTA software, http://cnsgenomics.com/software/gcta/; Roadmap Epigenomics project, http://www.roadmapepigenomics.org/; GTEx database, http://gtexportal.org; GREGOR, http://csg.sph.umich.edu/GREGOR/; Unified Medical Language System, https://www.nlm.nih.gov/research/umls/.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-018-0171-3.

Received: 2 January 2018; Accepted: 1 June 2018;
Published online: 30 July 2018

References

1. Jin, S. C. et al. Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nat. Genet.* **49**, 1593–1601 (2017).
2. Goette, A. et al. EHRA/HRS/APHRS/SOLAECE expert consensus on atrial cardiomyopathies: definition, characterization, and clinical implication. *EP Eur.* **18**, 1455–1490 (2016).
3. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
4. Wu, Y., Zheng, Z., Visscher, P. M. & Yang, J. Quantifying the mapping precision of genome-wide association studies using whole-genome sequencing data. *Genome Biol.* **18**, (2017).
5. Ge, T., Chen, C.-Y., Neale, B. M., Sabuncu, M. R. & Smoller, J. W. Phenome-wide heritability analysis of the UK Biobank. *PLoS Genet.* **13**, e1006711 (2017).
6. 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).
7. Thorolfsdottir, R. B. et al. A missense variant in *PLEC* increases risk of atrial fibrillation. *J. Am. Coll. Cardiol.* **70**, 2157–2168 (2017).
8. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
9. Gudbjartsson, D. F. et al. Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* **47**, 435–444 (2015).
10. Orr, N. et al. A mutation in the atrial-specific myosin light chain gene (*MYL4*) causes familial atrial fibrillation. *Nat. Commun.* **7**, 11303 (2016).
11. Nielsen, J. B. et al. Genome-wide study of atrial fibrillation identifies seven risk loci and highlights biological pathways and regulatory elements involved in cardiac development. *Am. J. Hum. Genet.* **102**, 103–115 (2018).
12. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
13. Schmidt, E. M. et al. GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. *Bioinformatics* **31**, 2601–2606 (2015).

14. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
15. Wells, A. et al. The anatomical distribution of genetic associations. *Nucleic Acids Res.* **43**, 10804–10820 (2015).
16. Noguchi, S. et al. Mutations in the dystrophin-associated protein γ-sarcoglycan in chromosome 13 muscular dystrophy. *Science* **270**, 819–822 (1995).
17. Brauch, K. M. et al. Mutations in ribonucleic acid binding protein gene cause familial dilated cardiomyopathy. *J. Am. Coll. Cardiol.* **54**, 930–941 (2009).
18. Gerull, B. et al. Mutations in the desmosomal protein plakophilin-2 are common in arrhythmogenic right ventricular cardiomyopathy. *Nat. Genet.* **36**, 1162–1164 (2004).
19. Skov, M. W. et al. Association between heart rate at rest and incident atrial fibrillation (from the Copenhagen Electrocardiographic Study). *Am. J. Cardiol.* **118**, 708–713 (2016).
20. Nielsen, J. B. et al. P-wave duration and the risk of atrial fibrillation: results from the Copenhagen ECG Study. *Heart Rhythm* **12**, 1887–1895 (2015).
21. Nielsen, J. B. et al. Risk of atrial fibrillation as a function of the electrocardiographic PR interval: Results from the Copenhagen ECG Study. *Heart Rhythm* **10**, 1249–1256 (2013).
22. Nielsen, J. B. et al. J-shaped association between QTc interval duration and the risk of atrial fibrillation: results from the Copenhagen ECG study. *J. Am. Coll. Cardiol.* **61**, 2557–2564 (2013).
23. Holm, H. et al. A rare variant in *MYH6* is associated with high risk of sick sinus syndrome. *Nat. Genet.* **43**, 316–320 (2011).
24. Bjornsson, T. et al. A rare missense mutation in *MYH6* confers high risk of coarctation of the aorta. Preprint at *bioRxiv* https://doi.org/10.1101/180794 (2017).
25. Maron, B. J. & Maron, M. S. Hypertrophic cardiomyopathy. *Lancet* **381**, 242–255 (2013).
26. England, J. & Loughna, S. Heavy and light roles: myosin in the morphogenesis of the heart. *Cell. Mol. Life Sci.* **70**, 1221–1239 (2013).
27. Herron, T. J., Korte, F. S. & McDonald, K. S. Loaded shortening and power output in cardiac myocytes are dependent on myosin heavy chain isoform expression. *Am. J. Physiol. Heart Circ. Physiol.* **281**, H1217–H1222 (2001).
28. Miyata, S., Minobe, W., Bristow, M. R. & Leinwand, L. A. Myosin heavy chain isoform expression in the failing and nonfailing human heart. *Circ. Res.* **86**, 386–390 (2000).
29. Cañón, S. et al. miR-208b upregulation interferes with calcium handling in HL-1 atrial myocytes: Implications in human chronic atrial fibrillation. *J. Mol. Cell. Cardiol.* **99**, 162–173 (2016).
30. Wagner, A. H. et al. DGIdb 2.0: mining clinically relevant drug–gene interactions. *Nucleic Acids Res.* **44**, D1036–D1044 (2016).
31. Teerlink, J. R. et al. Dose-dependent augmentation of cardiac systolic function with the selective cardiac myosin activator, omecamtiv mecarbil: a first-in-man study. *Lancet* **378**, 667–675 (2011).
32. Sudlow, C. et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
33. Fritsche, L. G. et al. Association of polygenic risk scores for multiple cancers in a phenome-wide study: results from The Michigan Genomics Initiative. *Am. J. Hum. Genet.* **102**, 1048–1061 (2018).
34. Lubitz, S. A. et al. Association between familial atrial fibrillation and risk of new-onset atrial fibrillation. *JAMA* **304**, 2263–2269 (2010).
35. Oyen, N. et al. Familial aggregation of lone atrial fibrillation in young persons. *J. Am. Coll. Cardiol.* **60**, 917–921 (2012).
36. Roselli, C. et al. Multi-ethnic genome-wide association study for atrial fibrillation. *Nat. Genet.* https://doi.org/10.1038/s41588-018-0133-9 (2018).
37. Costantini, D. L. et al. The homeodomain transcription factor Irx5 establishes the mouse cardiac ventricular repolarization gradient. *Cell* **123**, 347–358 (2005).
38. Veerman, C. C. et al. The Brugada Syndrome susceptibility gene *HEY2* modulates cardiac transmural ion channel patterning and electrical heterogeneity. *Circ. Res.* **121**, 537–548 (2017).

Acknowledgements

The Nord-Trøndelag Health Study (the HUNT Study) is a collaboration between the HUNT Research Centre (Faculty of Medicine, Norwegian University of Science and Technology (NTNU)), Nord-Trøndelag County Council, the Central Norway Health Authority, and the Norwegian Institute of Public Health. The K.G. Jebesen Center for Genetic Epidemiology is financed by Stiftelsen Kristian Gerhard Jebsen, the Faculty of Medicine and Health Sciences Norwegian University of Science and Technology (NTNU), and the Central Norway Regional Health Authority. This research has been conducted using the UK Biobank Resource under application number 24460. J.B.N. was supported by grants from the Danish Heart Foundation (16-R107-A6779) and the Lundbeck Foundation (R220-2016-1434). T.J.H. was supported by an American Heart Association Scientist Development Grant (0735464Z). J.A.S. was supported by National Institutes of Health grant R01-HL124232. C.J.W. was supported by National Institutes of Health grants R35-HL135824, R01-HL127564, R01-HL117626-02-S1, and R01-HL130705. To the best of our knowledge, this manuscript complies with all relevant ethical regulations.

Author contributions

J.B.N., R.B.T., L.G.F., W.Z., M.W.S., S.E.G., S.M., E.M.S., G.S., I.S., M.L., B.N.W., R.D., P.S., U.T., and X.W. performed the computational analyses. M.R.M., M.E.G., A.H.S., O.L.H., H.D., J.H.C., J.D.B., D.O.A., U.T., A.B., C.O., A.G.H., W.H., S.K., C.M.B., and T.M.T. conducted data acquisition. T.J.H., M.Y., R.D.C., J.K., J.A.S., and J.J. performed wet lab experiments. O.L.H., F.E.D., M.B., S.L., H.M.K., H.H., D.J.C., D.F.G., K.S., B.M., G.R.A., K.H., and C.J.W. designed and supervised the study. All authors contributed to manuscript preparation and read, commented on, and approved the manuscript.

Competing interests

R.B.T., G.S., D.O.A., P.S., U.T., D.F.G., H.H., and K.S. are employed by deCODE genetics/Amgen, Inc., Reykjavik, Iceland. A.G.H. is employed by Novo Nordisk A/S, Bagsværd, Denmark. S.M., J.H.C., J.D.B., A.B., C.O., F.E.D., G.R.A., and T.M.T. are employed by Regeneron Pharmaceuticals, Inc., Tarrytown, New York, USA. D.J.C. is employed by Geisinger Health System, Danville, Pennsylvania, USA.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0171-3.

Correspondence and requests for materials should be addressed to G.R.A. or K.H. or C.J.W.

Methods

Discovery cohorts. Additional details on selected cohorts are provided in the Supplementary Note.

HUNT. The Nord-Trøndelag Health Study (HUNT) is a population-based health survey conducted in the county of Nord-Trøndelag, Norway, since 1984[39]. We used a combination of hospital, out-patient, and emergency room discharge diagnoses (ICD-9 and ICD-10) to identify 6,493 atrial fibrillation cases and 63,142 atrial fibrillation-free controls with genotype data. Participation in the HUNT Study is based on informed consent, and the study has been approved by the Data Inspectorate and the Regional Ethics Committee for Medical Research in Norway.

deCODE. The Icelandic atrial fibrillation population consisted of all patients diagnosed with atrial fibrillation (ICD-10 code I48 and ICD-9 code 427.3) at Landspitali, The National University Hospital, in Reykjavik, and Akureyri Hospital (the two largest hospitals in Iceland) from 1987 to 2015. All atrial fibrillation cases, a total of 13,471, were included. Controls were 358,161 Icelanders recruited through different genetic research projects at deCODE genetics, excluding those in the atrial fibrillation cohort. The study was approved by the Icelandic Data Protection Authority and the National Bioethics Committee of Iceland (no. VSNb2015030021).

MGI. MGI is a hospital-based cohort collected at Michigan Medicine, USA. Atrial fibrillation cases ($n=1,226$) were defined as patients with ICD-9 billing code 427.31, and controls were individuals without atrial fibrillation, atrial flutter, or related phenotypes (ICD-9 426-427.99). MGI was reviewed and approved by the Institutional Review Board of the University of Michigan Medical School.

DiscovEHR. The DiscovEHR collaboration cohort is a hospital-based cohort including 58,124 genotyped individuals of European ancestry from the ongoing MyCode Community Health Initiative of the Geisinger Health System, USA[40]. Atrial fibrillation cases ($n=6,679$) were defined as DiscovEHR participants with at least one electronic health record problem list entry or at least two diagnosis code entries for two separate clinical encounters on separate calendar days for ICD-10 I48: atrial fibrillation and flutter. Corresponding controls ($n=41,803$) were defined as individuals with no electronic health record diagnosis code entries (problem list or encounter codes) for ICD-10 I48. The Study was approved by the Geisinger Institutional Review Board.

UK Biobank. The UK Biobank is a population-based cohort collected from multiple sites across the United Kingdom[32]. Cases of atrial fibrillation were selected using ICD-9 and ICD-10 codes for atrial fibrillation or atrial flutter (ICD-9 427.3 and ICD-10 I48). Controls were participants without any ICD-9 or ICD-10 codes specific for atrial fibrillation, atrial flutter, other cardiac arrhythmias, or conduction disorders.

AFGen Consortium. Published atrial fibrillation association summary statistics from 31 cohorts representing 17,931 atrial fibrillation cases and 115,142 controls were obtained from the authors[6].

Genotyping array, imputation, and association analysis. *HUNT.* Genotyping was performed at the Norwegian University of Science and Technology (NTNU) using the Illumina HumanCore Exome v1.0 and v1.1. Quality control was performed at the marker and sample level. A total of 2,201 individuals were whole-genome sequenced at low pass and genotype calls were generated using GotCloud pipeline (see URLs). Variants from the HUNT low-pass genomes were imputed into The Haplotype Reference Consortium (HRC) samples and vice versa to generate a single imputation reference panel of ~34,000 individuals including 2,201 HUNT study-specific samples. Imputation was performed using Minimac3, and variants with imputation $r^2>0.3$ were taken forward. We performed testing for association with atrial fibrillation using a generalized mixed model, including covariates birth year, sex, genotype batch, and principal components 1–4 as implemented in SAIGE[41].

deCODE. The study is based on whole-genome sequence data from 15,220 Icelanders participating in various disease projects at deCODE genetics. The sequencing was done using Illumina standard TruSeq methodology to a mean depth of 35× (s.d. 8)[9]. Autosomal SNPs and indels were identified using the Genome Analysis Toolkit version 3.4.0[42]. Variants that did not pass quality control were excluded from the analysis according to Genome Analysis Toolkit best practices. Variants identified through sequencing (SNPs and indels) were then imputed into 151,677 Icelanders genotyped using Illumina SNP chips and their close relatives (familial imputation)[43]. Variants for the meta-analysis were selected based on matching with either the 1000 Genomes Project reference panel (Phase 3) or the Haplotype Consortium reference panel[44] (based on allele, frequency, and correlation matching). Logistic regression was used to test for association between SNPs and atrial fibrillation, treating disease status as the response and allele counts from direct genotyping or expected genotype counts from imputation as covariates. Other available individual characteristics that correlate with phenotype status were also included in the model as nuisance variables. These characteristics were: sex, county of birth, current age or age at death (first- and second-order terms included), blood sample availability for the individual, and an indicator function for the overlap of the lifetime of the individual with the time span of phenotype collection. To account for inflation in test statistics due to cryptic relatedness and stratification, we applied the method of LD score regression[45]. The estimated correction factor for atrial fibrillation based on LD score regression was 1.38 for the additive model.

MGI. Genotyping was performed at the University of Michigan using the Illumina Human Core Exome v1.0 and v1.1. Quality control was performed at the marker and sample level. Imputation of variants from the HRC reference panel was performed using the Michigan Imputation Server (see URLs), and variants with imputation $r^2>0.3$ were included. Association with atrial fibrillation was determined using the Firth bias-corrected logistic likelihood ratio test[46] with adjustment for age, sex, and principal components 1–4.

DiscovEHR. Aliquots of DNA were sent to Illumina for genotyping on the Human OmniExpress Exome Beadchip. All individuals of European ancestry, as determined using principal component analysis, were imputed to the HRC reference panel using the Michigan Imputation Server. Markers with imputation $r^2>0.3$ and MAF >0.001 were carried forward for analysis. BOLT-LMM[47] was used to analyze BGEN dosage files, and variants were tested for association with atrial fibrillation under an additive genetic model, adjusting for sex, age, age^2, and the first four principal components of ancestry; additionally, a genetic relatedness matrix (calculated using variants with MAF >0.001, per-genotype missing data rate <1%, and Hardy–Weinberg equilibrium $P<10^{-15}$) was included as a random-effects variable in the model.

UK Biobank. Details on quality control, genotyping, and imputation can be found elsewhere[48]. In brief, study participants were genotyped using two very similar genotyping arrays (Applied Biosystems UK BiLEVE Axiom Array and UK BioBank Axiom Array) designed specifically for the UK Biobank. Phasing and imputation were done by the UK Biobank analysis team based on the HRC reference panel and the UK10K haplotype resource[48]. We restricted our analyses to HRC-imputed markers only as there have been reports of incorrect estimates for non-HRC markers in the first 500,000 people-release from UK Biobank. We performed testing for association with atrial fibrillation in people of white British ancestry using a generalized mixed model including covariates birth year, sex, genotype batch, and principal components 1–4 as implemented in SAIGE[41].

Meta-analysis. We included all markers that were available for analyses in any of the six contributing studies. For the DiscovEHR that applied the BOLT-LMM mixed model, we obtained an approximation of the allelic log-odds ratio and corresponding variance from the linear model as described previously[49]. Following this, we performed meta-analyses using the inverse variance method implemented in the software package METAL (see URLs)[50]. When estimating the cross-cohort allele frequencies, we only included participating studies where individuals were sampled independent of atrial fibrillation status (HUNT, deCODE, MGI, DiscovEHR, UK Biobank). This was done to avoid sampling bias. Heterogeneity tests were performed as implemented in METAL[50].

Definition of independent loci. Independent loci were defined as genetic markers >1 Mb and >0.25 cM apart in physical and genomic distance, respectively, with at least one genetic variant associated with atrial fibrillation at a genome-wide significance threshold of $P<5\times10^{-8}$. Loci borders were defined as the highest and lowest genomic positions within the locus reaching genome-wide significance plus an additional 1 Mb on either side.

LD estimation. We used 5,000 unrelated individuals that were randomly sampled among the HUNT Study participants for calculating LD r^2 by using the software PLINK1.9 (see URLs). We additionally used the 1000 Genomes Project phase 3 European (EUR) sample for LD estimation.

Approximate, stepwise conditional analyses. To identify independent risk variants within the identified atrial fibrillation-associated loci, we used the COJO-GCTA software (see URLs) to perform approximate, stepwise conditional analyses based on summary statistics from the meta-analyses and an LD matrix obtained from 5,000 unrelated individuals randomly sampled from the HUNT Study[3]. Only variants with MAF >0.01 were included in the analyses and variants were only considered truly independent if they were not in LD ($r^2<0.10$) with the locus index variant and any of the other independent risk variants.

Estimation of heritability. The genome-wide heritability explained by all markers was estimated based on GWAS summary statistics, LD-score regression, and European-ancestry LD information from the 1000 Genomes Project[45]. The heritability explained by atrial fibrillation-associated index variants and additional independent risk variants was calculated on the basis of odds ratios and risk allele frequencies as described previously[51].

Identifying candidate functional genes, gene sets, and tissues using DEPICT. We employed DEPICT (see URLs) to identify: (1) the most likely causal gene at associated loci, (2) reconstituted gene sets enriched for atrial fibrillation loci, and (3) tissues and cell types in which genes at associated loci that are preferentially expressed[8]. DEPICT uses gene expression data derived from a panel of 77,840 messenger RNA expression arrays[52] together with 14,461 existing gene sets defined based on molecular pathways derived from experimentally verified protein–protein interactions[53], genotype–phenotype relationships from the Mouse Genetics Initiative[54], Reactome pathways[55], KEGG pathways[56], and gene ontology terms[57]. Based on similarities across the microarray expression data, DEPICT reconstitutes the 14,461 existing gene sets by assigning each gene in the genome a likelihood of membership in each gene set. Using these precomputed gene sets and a set of trait-associated loci, DEPICT quantifies whether any of the 14,461 reconstituted gene sets are significantly enriched for genes in the associated loci and prioritizes genes that share predicted functions with genes from the other associated loci more often than expected by chance. Additionally, DEPICT uses a set of 37,427 human mRNA microarrays to identify tissues and cell types in which genes from associated loci are highly expressed (all genes residing within an LD of $r^2 > 0.5$ from index variant).

We ran DEPICT using all atrial fibrillation-associated index variants and all variants identified through stepwise conditional analyses, regardless of LD structure. For the gene sets significantly enriched for atrial fibrillation-associated loci ($P < 1 \times 10^{-6}$, FDR <0.05), we computed a weighted pairwise similarity based on the number of overlapping genes for genes with a Z score <4.75 (corresponding to $P < 1 \times 10^{-6}$) for being part of the gene set. For gene sets with no genes with a Z score <4.75, we included the three most significant genes as suggested previously[58].

Identification of regulatory elements using GREGOR. We tested for enrichment of index variants with functional domains using the software GREGOR (see URLs)[13]. This method tests for an increase in the number of atrial fibrillation-associated index variants, or their LD proxies, overlapping with the regulatory feature more often than expected by chance by comparing to permuted control sets where the variants are matched for frequency, number of LD proxies, and distance to the nearest gene. We use a saddle-point approximation to estimate the P value by comparing to the distribution of permuted statistics[13]. We ran GREGOR using all atrial fibrillation-associated index variants and all variants identified through stepwise conditional analyses, regardless of LD structure, and narrow peak BED-files from the Roadmap Epigenomics project (see URLs)[12].

Identification of expression quantitative trait loci (eQTLs) using GTEx data. We performed eQTL look-up using the GTEx database (see URLs)[14] version 6p, which holds *cis*-eQTL expression data of up to 190,000,000 single nucleotide variants across 44 tissues, by searching for all atrial fibrillation-associated loci index variants, all independent risk variants identified from the stepwise conditional analyses, and any variants in strong LD ($r^2 > 0.80$) with these variants using an eQTL significance threshold of $P < 1.14 \times 10^{-9}$ ($5 \times 10^{-8}/44$ tissues). For all statistically significant genes, we queried all markers in the GTEx database that affected the expression of the affected genes and tested whether the eQTL markers colocalized with the GWAS signal as described previously[59].

Ischemic heart failure model of atrial fibrillation susceptibility. Ischemic heart failure was modeled using a previously described rabbit model of left circumflex artery ligation[60]. In this model, the left atrium progressively dilates following the ischemic insult as heart failure develops. Figure 4a shows images of Langendorff perfused hearts of control (sham operated) and heart failure animals highlighting the overt dilation of the left atrium in heart failure. With equivalent left atrial pressure, atrial fibrillation was induced in each condition with high frequency burst pacing as shown in the ECG traces and as described previously[61]. Protein expression analysis was performed using western blot. Use of animals was reviewed and approved by the Care and Use Committee of the University of Michigan. All pre-, intra-, and postoperative surgical procedures were developed in collaboration with veterinarians on staff in the Unit for Laboratory Animal Medicine of the University of Michigan.

Tissue-specific expression analysis (TSEA). The TSEA analyses were performed using the R software pSI package (see URLs)[15]. For the calculations, predefined pSI values provided by the pSI package creators were used. To get null distributions for the P values for the prioritized genes, we performed two sets of permutations: randomly selected from the entire human genome and randomly selected from the associated loci (also matching the number of genes picked in each of the loci), as done previously[62]. In both scenarios, 1,000 permutations were performed.

ECG-wide association analyses. ECG data were collected from Landspitali University Hospital in Reykjavik and included all ECGs obtained and digitally stored from 1998 to 2015, including a total of 434,000 ECGs from 88,217 individuals. A total of 289,297 ECGs of 62,974 individuals were sinus rhythm (heart rate 50–100 beats per minute) ECGs of individuals without the diagnosis of atrial fibrillation. The ECGs were digitally recorded with the Philips PageWriter Trim III, PageWriter 200, Philips Page Writer 50, and Phillips Page Writer 70 cardiographs and stored in the Philips TraceMasterVue ECG Management System. These were ECGs obtained in all hospital departments, from both inpatients and outpatients. Digitally measured ECG waveforms and parameters were extracted from the database for analysis. The Philips PageWriter Trim III QT interval measurement algorithm has been previously described and shown to fulfill industrial ECG measurement accuracy standards[63]. The Philips PR interval and QRS complex measurements have been shown to fulfill industrial accuracy standards[64].

We tested the 111 genome-wide significant atrial fibrillation index variants for association with 123 ECG parameters using a linear mixed effects model implemented in the Bolt software package[47], treating the ECG measurement as the response and the genotype as the covariate. All measures except heart rate and QT interval are presented for all 12 ECG leads. For this analysis, we used 289,297 sinus rhythm ECGs (heart rate 50–100 beats per minute) from 62,974 individuals who have not been diagnosed with atrial fibrillation according to our databases. This was done to assess the effect of the atrial fibrillation variants on ECG measures and cardiac electrical function in the absence of atrial fibrillation. Individuals with pacemakers were also excluded. The ECG measurements were adjusted for sex, year of birth, and age at measurement and were subsequently quantile standardized to have a normal distribution. For individuals with multiple ECG measurements, the mean standardized value was used. We assume that the quantitative measurements follow a normal distribution with a mean that depends linearly on the expected allele at the variant and a variance–covariance matrix proportional to the kinship matrix[65]. Since 123 traits were tested, the Benjamini–Hochberg FDR procedure controlling the FDR at 0.05 at each marker was used to account for multiple testing.

Polygenic risk scores. For each study participant in the UK Biobank and in the HUNT Study, we constructed an inverse normal-transformed polygenic risk score for atrial fibrillation using summarized dosage-weighted risk estimates from the list of 142 independent risk variants. For the UK Biobank risk score, risk estimates (beta coefficients) were obtained by meta-analyzing the risk variants across all contributing studies excluding the UK Biobank. To explore the association between the genetic burden of atrial fibrillation and the age of onset of atrial fibrillation, which we assumed was independent of the case status used for obtaining the risk estimates, we obtained risk estimates from meta-analyses of the full sample size.

Phenome-wide association analyses in the UK Biobank. We used a previously published scheme to define disease-specific binary phenotypes by combining hospital ICD-9 codes into hierarchical PheCodes, each representing a particular disease group[66]. ICD-10 codes were mapped to PheCodes using a combination of available maps through the Unified Medical Language System (see URLs), string matching, and manual review. UK Biobank study participants were labeled with a PheCode if they had one or more of the PheCode-specific ICD codes. Cases were all UK Biobank study participants with the PheCode of interest and controls were all UK Biobank study participants without the PheCode of interest or any related PheCodes. Sex checks were performed, so PheCodes specific for one sex could not mistakenly be assigned to the other sex. The associations between the polygenic risk score and each of the defined phenotypes were tested using a logistic regression adjusted for sex and birth year.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability. The meta-analysis summary association statistics that support the findings of this study are available for download at http://csg.sph.umich.edu/willer/public/afib2018.

References

39. Krokstad, S. et al. Cohort Profile: The HUNT Study, Norway. *Int. J. Epidemiol.* **42**, 968–977 (2013).
40. Carey, D. J. et al. The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research. *Genet. Med.* **18**, 906–913 (2016).
41. Zhou, W. et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Preprint at *bioRxiv* https://doi.org/10.1101/212357 (2017).
42. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
43. Kong, A. et al. Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat. Genet.* **40**, 1068–1075 (2008).
44. 1000 Genomes Project Consortium et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
45. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

46. Ma, C., Blackwell, T., Boehnke, M. & Scott, L. J., GoT2D investigators. Recommended joint and meta-analysis strategies for case-control association testing of single low-count variants. *Genet. Epidemiol.* **37**, 539–550 (2013).
47. Loh, P.-R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
48. Bycroft, C. et al. Genome-wide genetic data on ~500,000 UK Biobank participants. Preprint at *bioRxiv* https://doi.org/10.1101/166298 (2017).
49. Cook, J. P., Mahajan, A. & Morris, A. P. Guidance for the utility of linear models in meta-analysis of genetic association studies of binary phenotypes. *Eur. J. Hum. Genet.* **25**, 240–245 (2017).
50. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
51. So, H.-C., Gui, A. H. S., Cherny, S. S. & Sham, P. C. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genet. Epidemiol.* **35**, 310–317 (2011).
52. Fehrmann, R. S. N. et al. Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat. Genet.* **47**, 115–25 (2015).
53. Lage, K. et al. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat. Biotechnol.* **25**, 309–316 (2007).
54. Bult, C. J. et al. Mouse genome informatics in a new age of biological inquiry. in *Proc. IEEE International Symposium on Bio-Informatics and Biomedical Engineering* 29–32 (IEEE, Piscataway, New Jersey, USA, 2000).
55. Croft, D. et al. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* **39**, D691–D697 (2011).
56. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40**, D109–D114 (2012).
57. Raychaudhuri, S. et al. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5**, e1000534 (2009).
58. Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
59. Wen, X., Pique-Regi, R. & Luca, F. Integrating molecular QTL data into genome-wide genetic association analysis: probabilistic assessment of enrichment and colocalization. *PLoS Genet.* **13**, e1006646 (2017).
60. Herron, T. J. et al. Ca2+-independent positive molecular inotropy for failing rabbit and human cardiac muscle by alpha-myosin motor gene transfer. *FASEB J.* **24**, 415–424 (2010).
61. Yamazaki, M., Filgueiras-Rama, D., Berenfeld, O. & Kalifa, J. Ectopic and reentrant activation patterns in the posterior left atrium during stretch-related atrial fibrillation. *Prog. Biophys. Mol. Biol.* **110**, 269–277 (2012).
62. Ferreira, M. A. et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat. Genet.* **49**, 1752–1757 (2017).
63. Zhou, S. H., Helfenbein, E. D., Lindauer, J. M., Gregg, R. E. & Feild, D. Q. Philips QT interval measurement algorithms for diagnostic, ambulatory, and patient monitoring ECG applications. *Ann. Noninvasive Electrocardiol* **14**(Suppl.), S3–S8 (2009).
64. Lindauer, J., Gregg, R., Helfenbein, E., Shao, M. & Zhou, S. Global QT measurements in the Philips 12-lead algorithm. *J. Electrocardiol.* **38**, 90 (2005).
65. Benonisdottir, S. et al. Epigenetic and genetic components of height regulation. *Nat. Commun.* **7**, 13490 (2016).
66. Denny, J. C. et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat. Biotechnol.* **31**, 1102–1110 (2013).

Paper IV

ESC European Society of Cardiology

European Heart Journal (2018) **0**, 1–7
doi:10.1093/eurheartj/ehy142

CLINICAL RESEARCH
Congenital heart disease

A rare missense mutation in *MYH6* associates with non-syndromic coarctation of the aorta

Thorsteinn Bjornsson[1†], Rosa B. Thorolfsdottir[1†], Gardar Sveinbjornsson[1], Patrick Sulem[1], Gudmundur L. Norddahl[1], Anna Helgadottir[1], Solveig Gretarsdottir[1], Audur Magnusdottir[1], Ragnar Danielsen[2], Emil L. Sigurdsson[3,4], Berglind Adalsteinsdottir[5,6], Sverrir I. Gunnarsson[7], Ingileif Jonsdottir[1,6,8], David O. Arnar[1,2,6], Hrodmar Helgason[9], Tomas Gudbjartsson[6,10], Daniel F. Gudbjartsson[1,11], Unnur Thorsteinsdottir[1,6], Hilma Holm[1]*, and Kari Stefansson[1,6]*

[1]deCODE genetics/Amgen, Inc., Sturlugata 8, 101 Reykjavik, Iceland; [2]Department of Medicine, Landspitali—The National University Hospital of Iceland, Hringbraut, 101 Reykjavik, Iceland; [3]Department of Family Medicine, University of Iceland, Vatnsmyrarvegur 16, 101 Reykjavik, Iceland; [4]Department of Development, Primary Health Care of the Capital Area, Alfabakki 16, 109 Reykjavik, Iceland; [5]Department of Cardiology, Haukeland University Hospital, Jonas Lies vei 83, 5021 Bergen, Norway; [6]Faculty of Medicine, University of Iceland, Vatnsmyrarvegur 16, 101 Reykjavik, Iceland; [7]Division of Cardiovascular Medicine, Department of Medicine, University of Wisconsin, 600 Highland Ave, Madison, WI 53792, USA; [8]Department of Immunology, Landspitali—The National University Hospital of Iceland, Hringbraut, 101 Reykjavik, Iceland; [9]Children's Hospital, Landspitali—The National University Hospital of Iceland, Hringbraut, 101 Reykjavik, Iceland; [10]Department of Cardiothoracic Surgery, Landspitali—The National University Hospital of Iceland, Hringbraut, 101 Reykjavik, Iceland; and [11]School of Engineering and Natural Sciences, University of Iceland, Hjardarhagi 4, 107 Reykjavik, Iceland

Received 3 November 2017; revised 5 January 2018; editorial decision 28 February 2018; accepted 2 March 2018

Aims Coarctation of the aorta (CoA) accounts for 4–8% of congenital heart defects (CHDs) and confers substantial morbidity despite treatment. It is increasingly recognized as a highly heritable condition. The aim of the study was to search for sequence variants that affect the risk of CoA.

Methods and results We performed a genome-wide association study of CoA among Icelanders (120 cases and 355 166 controls) based on imputed variants identified through whole-genome sequencing. We found association with a rare (frequency = 0.34%) missense mutation p.Arg721Trp in *MYH6* (odds ratio = 44.2, $P = 5.0 \times 10^{-22}$), encoding the alpha-heavy chain subunit of cardiac myosin, an essential sarcomere protein. Approximately 20% of individuals with CoA in Iceland carry this mutation. We show that p.Arg721Trp also associates with other CHDs, in particular bicuspid aortic valve. We have previously reported broad effects of p.Arg721Trp on cardiac electrical function and strong association with sick sinus syndrome and atrial fibrillation.

Conclusion Through a population approach, we found that a rare missense mutation p.Arg721Trp in the sarcomere gene *MYH6* has a strong effect on the risk of CoA and explains a substantial fraction of the Icelanders with CoA. This is the first mutation associated with non-familial or sporadic form of CoA at a population level. The p.Arg721Trp in *MYH6* causes a cardiac syndrome with highly variable expressivity and emphasizes the importance of sarcomere integrity for cardiac development and function.

Keywords Coarctation of the aorta • Genetics • Sarcomere • *MYH6*

* Corresponding author. Tel: +354 570 1900, Fax: +354 570 1901, Email: hilma.holm@decode.is; kstefans@decode.is

† These authors contributed equally to this work.

Introduction

Coarctation of the aorta (CoA) is the most common birth defect of the aorta with an incidence of about one per 2500 live births.[1] It is defined by local narrowing of the proximal descending aorta and/or aortic arch, accompanied by bicuspid aortic valve (BAV) in more than 50% of cases and generally presenting as either neonatal heart failure or hypertension later in life.[2] Surgical or interventional treatment considerably improves outcome but risk of premature cardiovascular morbidity and mortality remains despite appropriate therapy.[3]

Coarctation of the aorta is primarily a non-familial or sporadic disease.[4] However, it has been shown to cosegregate in families with left-ventricular outflow tract obstruction (LVOTO) malformations, a mechanistically defined subgroup of congenital heart defects (CHDs) including CoA, BAV, congenital aortic stenosis, and hypoplastic left heart syndrome (HLHS).[5] As a group, the LVOTO malformations are markedly heritable (0.71–0.90) and have a high relative risk for first-degree relatives (36.9).[6] In addition, around 15% of individuals with CoA occur as part of a recognized genetic syndrome (e.g. 45, X, or Turner).[7]

Not much is known about genetic causes of non-syndromic CoA. Several studies have found mutations in families with LVOTO malformations and a few instances of sporadic CoA, both with and without concomitant CHDs. The most strongly implicated gene is *NOTCH1*,[8–10] encoding a transmembrane receptor that regulates cell fate during development. Mutations in other genes, including *MYH6*[11,12] *SMAD6*,[13] *NKX2-5*,[14] and *GATA5*,[15] have been found in one or few individuals with CoA. The *MYH6* mutations were found in two families, one with predisposition to atrial septal defect (ASD)[11] and the other to HLHS.[12] Some individuals in these families presented with CoA. In addition, knockout in mice of several genes found within copy number variants in individuals with CoA, including *MCTP2*,[16] *MATR3*,[17] and *FOXC1*,[18] have resulted in CoA-like phenotypes.

Methods

GWAS study design

Study samples

The CoA sample set included 120 Icelanders who received the discharge diagnosis of CoA at Landspitali, The National University Hospital (LUH) in Reykjavik, the only tertiary referral centre in Iceland, between 1984 and 2016. The individuals were diagnosed with CoA between the years 1950 and 2016, with most individuals (75%) diagnosed after 1990. The individuals diagnosed with CoA were identified either through diagnostic codes of CoA (ICD-9 code 747.1, ICD-10 code Q25.1) registered between 1990 and 2016 or procedure codes of CoA (WHO codes 1-273, 5-369, 5-382, and 5-387, NOMESCO codes FDJ 00, FDJ 10, FDJ 20, FDJ 30, FDJ 42, and FDJ 96) registered between 1984 and 2016. The diagnoses of CoA were confirmed and detailed phenotypic characteristics (Supplementary material online, Table S1) established through review of electronic and paper medical records at LUH. Coarctation of the aorta was defined as a non-syndromic congenital narrowing of the aorta, the diagnosis of which was confirmed by a cardiologist with echocardiography and/or cardiac catheterization. The individuals used as controls in the CoA GWAS analyses consisted of disease-free individuals randomly drawn from the Icelandic genealogical database and individuals from other genetic studies at deCODE.

In addition to CoA, we included in the study the following samples from the deCODE phenotype database: BAV, ASD, ventricular septal defect (VSD), patent ductus arteriosus (PDA), late onset aortic valve stenosis (AVS), sick sinus syndrome (SSS), atrial fibrillation (AF), heart failure (HF), ischaemic stroke (IS), hypertension (HTN), coronary artery disease (CAD), left atrial diameter (LAD), aortic root diameter (ARD), left ventricular end-diastolic diameter (LVEDD), electrocardiogram (ECG) data, thoracic aortic aneurysm, high-degree atrioventricular block, and hypertrophic cardiomyopathy (Supplementary material online, Supplementary methods).

All DNA samples used in the study are part of deCODE's biobank established in 1996 and built up since then through various genetic studies at deCODE.

The study was approved by the Icelandic Data Protection Authority and the National Bioethics Committee of Iceland. Study approval numbers were VSN-15-053, VSN-15-016, VSN-15-056, VSN-15-058, VSN-15-114, VSN-15-057, and 10-009-S1. Written informed consent was obtained from all study participants. The study complies with the declaration of Helsinki.

Genotyping, whole-genome sequencing, and imputation

For chip genotyping, 151 677 samples were typed with the Illumina HumanHap300, HumanCNV370, HumanHap610, HumanHap1M, HumanHap660, Omni-1, Omni 2.5, or Omni Express bead chips at deCODE. Long range phasing of all chip-genotyped individuals was performed with methods previously described[19,20] (Supplementary material online, Supplementary methods).

The whole genomes of 15 220 Icelanders were sequenced using Illumina technology to a median depth of 35X (Supplementary material online, Supplementary methods). The sequence variants identified in the 15 220 sequenced Icelanders were then imputed into 151 677 Icelanders who had been genotyped with various Illumina single nucleotide polymorphism chips and their genotypes phased using long-range phasing.[19,20] The imputation of the sequence variants, identified thorough whole-genome sequencing (WGS), into the chip typed long-range phased individuals was performed with the same model as used by IMPUTE.[21] The utilization of long-range phased haplotypes enables accurate imputation of variants with frequency down to approximately 0.02% in this data set. Using genealogic information on Icelanders from The Book of Icelanders,[22] the sequence variants were imputed into first and second-degree relatives of chip genotyped individuals (genealogical imputation),[23] to further increase the sample size for association analysis and to increase the power to detect associations. We identified 32.5 million high quality sequence variants (all with imputation information >0.8 that mapped to build hg38) that were tested for association with CoA under the multiplicative model.

Association analysis

In the association analysis were 120 individuals diagnosed with CoA and 355 116 individuals as controls, all with imputed genotypes. The sequence variants imputed were identified through WGS of 15 220 Icelanders ($n = 33$ individuals with CoA and $n = 15\,187$ individuals as controls). Of the individuals diagnosed with CoA, 39 were chip typed and long-range phased and of the individuals who were controls 140 661 were chip typed and long-range phased. These were imputed with the same model as used by IMPUTE. The remaining individuals ($n = 81$ CoA cases and $n = 214\,412$ controls), were not chip typed themselves but were first or second degree relatives of the chip typed individuals and imputed using genealogical imputation as described in Ref.[23]

To account for inflation in test statistics due to cryptic relatedness and stratification, we applied the method of linkage disequilibrium score

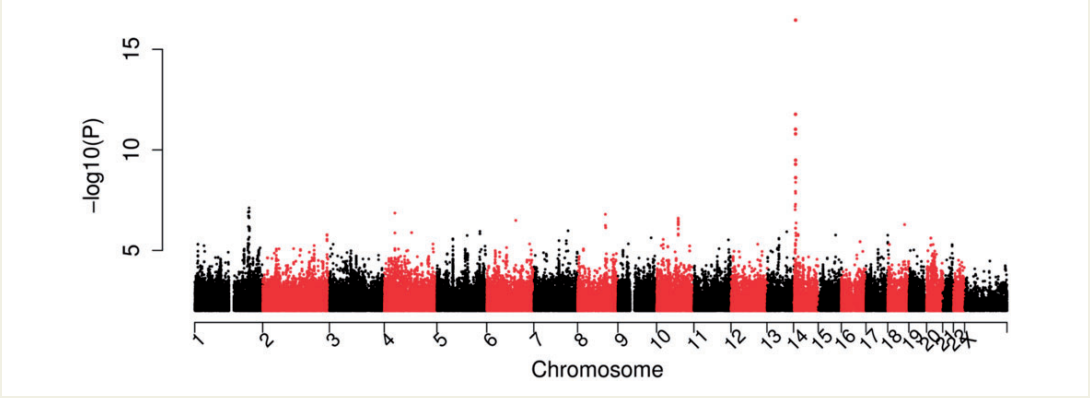


Figure 1 Manhattan plot of coarctation of the aorta genome-wide association study in Iceland. The *P* values ($-\log_{10}$) are plotted against their respective positions on each chromosome.

regression[24] (Supplementary material online, Supplementary methods). The estimated correction factor was 1.04 for the multiplicative model of the CoA association. To correct for multiple testing we used the weighted Holm–Bonferroni method[25] to allocate family wise error rate of 0.05 equally between four annotation-based classes of sequence variants (Supplementary material online, Supplementary methods). When testing the association of p.Arg721Trp with several other cardiac phenotypes, the individuals that served as controls consisted of disease-free individuals randomly drawn from the Icelandic genealogical database and individuals from other genetic studies at deCODE.

Phenotypic differences between carriers and non-carriers of p.Arg721Trp

To analyse if CoA carriers of the p.Arg721Trp mutation differed clinically from non-carrier individuals with CoA, we evaluated the frequencies of various clinical characteristics in these two groups with CoA (see Supplementary material online, Table S2). Fisher's exact test was used to test for significant difference in the mean frequency of the variants between non-carriers and carriers, and the odds ratio (OR) was calculated as [pa/(1-pa)]/[pc/(1-pc)], where pa and pc are the mean frequencies of the variants in non-carriers and carriers, respectively.

Results

A rare missense mutation in *MYH6* associates with coarctation of the aorta

To search for sequence variants that associate with non-syndromic CoA, we performed a GWAS including 120 Icelanders with CoA and 355 116 Icelanders who served as population controls. We observed a genome-wide significant association with CoA at chromosome 14q11 (*Figure 1*), explained by a rare (allele frequency = 0.34%) missense mutation p.Arg721Trp (c.2161C>T) in *MYH6*, encoding the alpha myosin heavy chain subunit (αMHC) in cardiac muscle. Alpha myosin heavy chain subunit is a main component of the sarcomere, the basic contractile unit of cardiac muscle.[26] P.Arg721Trp associates with CoA with an OR of 44.2 (95% confidence interval 20.5–95.5) and $P = 5.01 \times 10^{-22}$ (genome-wide significance threshold for missense variants was set at 6.5×10^{-8}, see Methods)[27] (*Figure 2*, *Table 1*). None of the genotyped individuals ($N = 151\ 677$) were homozygous for the mutation, consistent with its low frequency (1.8 homozygotes expected under Hardy–Weinberg equilibrium). Since we observed no homozygotes, we could not discriminate between the dominant and the multiplicative modes of inheritance.

The p.Arg721Trp mutation is located in exon 18 (out of 39 exons) of *MYH6* and leads to an arginine to tryptophan alteration at amino acid 721 (full-length protein 1939 amino acid) (see Supplementary material online, Figure S1). It is located in the converter domain of αMHC (see Supplementary material online, Figure S1 and S2), a small domain crucial in conveying a conformational change from the active site to the lever arm upon adenosine triphosphate (ATP) hydrolysis.[28] It is considered likely that the mutation alters protein function (SIFT = 0, PolyPhen = 0.99, MutationTaster = 0.93), probably by altering the folding of the converter domain.

There were 987 carriers of p.Arg721Trp among the 151 677 chip-typed Icelanders and eight of those (one per 123 carriers) were diagnosed with CoA. In line with low penetrance of the mutation for CoA, p.Arg721Trp carriers diagnosed with CoA did not cluster in families. However, 20% of the 39 chip-typed individuals with CoA carried p.Arg721Trp. Thus, while the penetrance of the mutation for CoA is low, it accounts for a large proportion of individuals with CoA in the Icelandic population.

The p.Arg721Trp mutation is not present in the Exome Variant Server, containing sequence data from 6503 individuals [Exome Variant Server, NHLBI Exome Sequencing Project (ESP), Seattle, WA, USA] (http://evs.gs.washington.edu/EVS/) (August 2016) and one copy was found in The Genome Aggregation Database (gnomAD), holding data from 126 216 exome sequences and 15 136 WGS unrelated individuals.[29] The p.Arg721Trp mutation thus appears to be absent from or present at a very low frequency in other populations.

We show the phenotypic characteristics of individuals with CoA in Supplementary material online, Table S1. About half were diagnosed

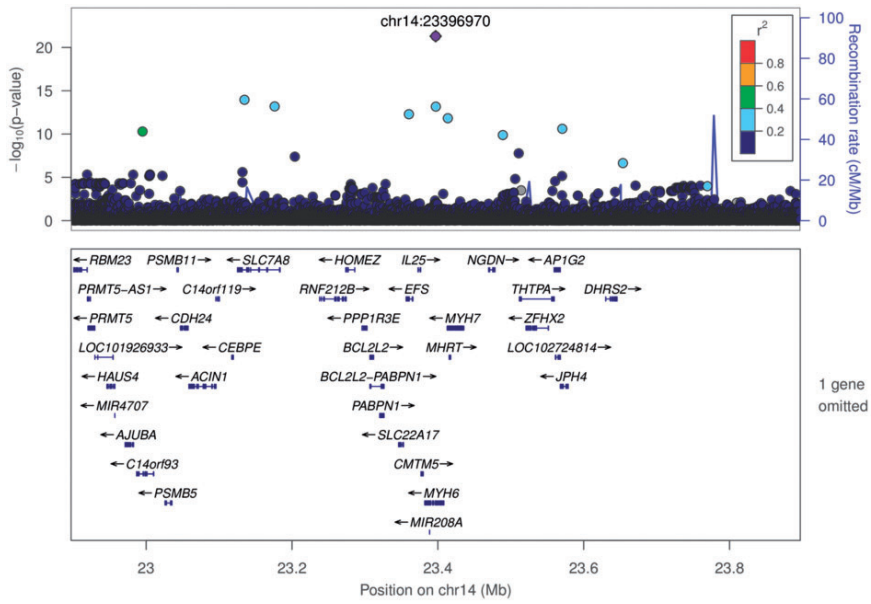


Figure 2 Region plot for the association of variants on 14q11 with coarctation of the aorta. Shown is a 1 Mb region on chromosome 14. The strongest association is with the missense variant p.Arg721Trp in *MYH6* located at position 23 396 970 on chromosome 14 (chr14: 23 396 970). The nine other variants shown are weakly correlated with p.Arg721Trp, r^2 between 0.6–0.4 (green) and 0.4–0.2 (light blue).

during the first month of life and three quarters during the first year of life. As expected,[3] CoA was more common in males than in females (1.6:1). About three quarters of individuals with CoA had other CHDs, most commonly BAV and VSD. Similar to other studies,[30] the aortic valve was bicuspid in about half of those with CoA.

To determine whether there are phenotypic differences between carriers ($n = 24$) and non-carriers ($n = 96$) of the p.Arg721Trp mutation within the CoA sample set, we evaluated the frequencies of various clinical characteristics in the two groups (see Supplementary material online, Table S2). Carriers were nominally more likely to present with mild rather than more critical and complex forms of CoA (OR = 4.2 and $P = 0.023$). We observed no other differences.

Association of p.Arg721Trp in *MYH6* with other cardiac diseases

We have previously demonstrated that p.Arg721Trp associates strongly with SSS and AF, atrial arrhythmias that are common in the elderly and frequently coexist.[31] With larger sample sizes, these associations have become stronger; however, previously reported association with thoracic aortic aneurysm is no longer significant (*Table 1*). In the context of assessing effects of AF risk variants on cardiac conduction, we have also recently shown that p.Arg721Trp associates with many ECG measures corresponding to a widespread effect on electrical function of the heart[32] (see Supplementary material online, Figure S3).

To further explore the effect of the p.Arg721Trp *MYH6* mutation, we tested it for association with additional cardiac phenotypes, including other CHDs, common heart diseases, and several echocardiogram variables (*Table 1*, Supplementary material online, Table S3 and Figure S3; significance threshold set at $P < 0.003$ (0.05/17 individual phenotypes tested). The p.Arg721Trp mutation associates with increased risk of several CHDs: BAV, VSD, ASD, and PDA (*Table 1*). As expected, the strongest association was with BAV (OR = 10.5 and $P = 7.3 \times 10^{-8}$). In addition, the mutation associates with late onset AVS. To assess if p.Arg721Trp associates with CHDs in the absence of diagnosed CoA, we re-tested for association after removing individuals with CoA from the analysis. Although the effect of p.Arg721Trp is consistently weaker, the associations remain (see Supplementary material online, Table S3). We cannot exclude the existence of undiagnosed CoA in these individuals. The mutation also associates with HF and IS and with LAD but not with other variables derived from the echocardiographic data such as ARD or LVEDD (*Table 1*). The p.Arg721Trp mutation did not associate with HTN or CAD.

Discussion

Through GWAS based on variants identified through WGS, we found a rare missense variant in the sarcomere gene *MYH6* that has a strong effect on the risk of CoA in the Icelandic population and explains a substantial fraction of CoA in Icelanders. The same mutation

Table I Association of p.Arg721Trp with congenital heart defects and various cardiac phenotypes

	N_{aff}	N_{contr}	OR/effect (95% CI)[a]	*P*-value
Congenital heart defects				
Coarctation of the aorta	120	355 116	44.2 (20.5 to 95.5)	5.0×10^{-22}
Bicuspid aortic valve	208	293 346	10.5 (2.6 to 38.0)	7.3×10^{-8}
Ventricular septal defect	715	357 641	4.4 (1.9 to 10.0)	3.7×10^{-4}
Patent ductus arteriosus	594	357 762	4.9 (2.1 to 11.6)	2.3×10^{-4}
Atrial septal defect	657	353 096	3.3 (1.5 to 7.1)	0.0026
Cardiac conditions				
Sick sinus syndrome	3310	346 082	8.7 (6.8 to 11.2)	6.2×10^{-64}
Atrial fibrillation	13 471	374 939	2.4 (1.9 to 3.0)	1.1×10^{-14}
Aortic valve stenosis	2457	349 342	2.7 (1.8 to 4.0)	1.8×10^{-6}
Heart failure	10 480	353 508	1.8 (1.4 to 2.3)	2.3×10^{-6}
Ischaemic stroke	8948	369 624	1.5 (1.1 to 2.0)	0.0029
High degree atrioventricular block	1303	361 919	2.1 (1.2 to 3.5)	0.0092
Coronary artery disease	37 782	318 845	1.2 (1.0 to 1.5)	0.056
Hypertrophic cardiomyopathy	163	239 293	0.0 (0.0 to 4.5)	0.15
Thoracic aortic aneurysm	353	302 458	1.8 (0.6 to 5.3)	0.31
Hypertension	54 974	324 803	1.1 (0.9 to 1.3)	0.44
Echocardiogram				
Left atrial diameter	19 380		0.3 (0.1 to 0.5)	2.6×10^{-4}
Aortic root diameter	19 506		-0.1 (-0.2 to 0.1)	0.41
LVEDD[b]	5701		0.0 (-0.3 to 0.3)	0.93

Shown are the number of affected individuals and control individuals used in the association analysis for each of the traits.
[a]Estimated odds ratio (OR) or the effect in standard deviation and the 95% confidence interval (CI) for the association with p.Arg721Trp.
[b]Left ventricular end-diastolic diameter.

also associates with other CHDs, in particular BAV. It has a widespread effect on cardiac electrical function and associates strongly with atrial arrhythmias, both SSS and AF. This is the first mutation shown to associate with non-familial or sporadic form of CoA at a population level. The p.Arg721Trp mutation appears to be absent from other populations or if present, at a very low frequency. The Icelandic population is a founder population in that a small number of ancestors account for a relatively large proportion of genetic diversity in the current population. Hence, sequence variants that are very rare in more outbred populations, like p.Arg721Trp, may thus be more frequent in Icelanders.[33]

Myosin is a major component of the sarcomere, the building block of the contractile system of cardiac muscle. Myosin is an ATPase cellular motor protein composed of two heavy chains and two pairs of light chains. The two heavy chains are αMHC and beta myosin heavy chain (βMHC) encoded by *MYH6* and *MYH7*, respectively. Both αMHC and βMHC are expressed throughout the heart during embryonic cardiogenesis and βMHC continues to do so in the adult heart whereas αMHC expression becomes restricted to the atrium.[34] Expression of *MYH6* has not been detected in the aorta.[35]

The pathogenesis of CoA is not well understood. One of the main models of CoA pathogenesis, the haemodynamic theory,[2,36] maintains that cardiac lesions resulting in decreased left ventricular outflow promote development of CoA by reducing blood flow through the Foetal aorta. The p.Arg721Trp mutation could predispose to CoA by reducing blood flow through the Foetal aorta because of diminished contraction of the developing heart. This hypothesis is supported by overexpression studies in rat cardiomyocytes showing that the p.Arg721Trp mutation impairs sarcomeric structure[37] and by our ECG data demonstrating widespread effect of p.Arg721Trp on cardiac electrical function, including in the ventricles. Our hypothesis is compatible with the fact that *MYH6* is expressed in the ventricles during the development of the heart but not in the aorta.

Very rare mutations in *MYH6*, other than p.Arg721Trp, have been linked to various CHDs,[11,12,38] particularly familial ASD[39,40] and both dilated and hypertrophic cardiomyopathy.[41] In all instances, these mutations have been restricted to a few sporadic cases or too few families. In two of these families, one with predisposition to ASD[11] and the other to HLHS,[12] some of the affected family members had other cardiac defects, including CoA. p.Arg721Trp in *MYH6* differs from these rare familial mutations in that it associates with CoA at the population level and explains about 20% of individuals with CoA in Iceland.

The main limitation of the study is the small size of the CoA sample set. A larger set might have facilitated detection of more variants associating with CoA with weaker effects than observed for p.Arg721Trp, and allowed a better estimate of the effect (OR), penetrance and the fraction of CoA cases explained by the mutation.

Conclusion

In conclusion, our findings give insights into the pathophysiology of CoA, supporting the haemodynamic theory of the pathogenesis.

Moreover, the pleiotropic effect of p.Arg721Trp in *MYH6* suggests it causes a cardiac syndrome with highly variable expressivity that is difficult to understand clinically without sequence information. Furthermore, these data emphasize the importance of sarcomere integrity for cardiac development and function.

Supplementary material

Supplementary material is available at *European Heart Journal* online.

Acknowledgements

We thank all the study subjects for their valuable participation as well as our colleagues who contributed to data collection, sample handling, and genotyping.

Funding

deCODE genetics/Amgen, Inc.

Conflict of interest: The following authors affiliated with deCODE Genetics/Amgen, Inc. are employed by the company: T.B., R.B.T., G.S., P.S., G.L.N., A.H., S.G., A.M., I.J., D.O.A., D.F.G., U.T., H.H., K.S.

References

1. Hoffman JI, Kaplan S. The incidence of congenital heart disease. *J Am Coll Cardiol* 2002;**39**:1890–1900.
2. Allen HD, Driscoll DJ, Shaddy RE, Feltes TF. *Moss & Adams' Heart Disease in Infants, Children, and Adolescents: Including the Fetus and Young Adult*. 8th ed. Philadelphia: Lippincott Wilkins & Wilkins; 2013.
3. Aboulhosn J, Child JS. Left ventricular outflow obstruction: subaortic stenosis, bicuspid aortic valve, supravalvar aortic stenosis, and coarctation of the aorta. *Circulation* 2006;**114**:2412–2422.
4. Keane J, Fyler D, Lock J. *Nadaś Pediatric Cardiology*. 2nd ed. Philadelphia: Saunders; 2006.
5. Loffredo CA, Chokkalingam A, Sill AM, Boughman JA, Clark EB, Scheel J, Brenner JI. Prevalence of congenital cardiovascular malformations among relatives of infants with hypoplastic left heart, coarctation of the aorta, and d-transposition of the great arteries. *Am J Med Genet A* 2004;**124a**:225–230.
6. McBride KL, Pignatelli R, Lewin M, Ho T, Fernbach S, Menesses A, Lam W, Leal SM, Kaplan N, Schliekelman P, Towbin JA, Belmont JW. Inheritance analysis of congenital left ventricular outflow tract obstruction malformations: segregation, multiplex relative risk, and heritability. *Am J Med Genet A* 2005;**134a**:180–186.
7. van der Bom T, Zomer AC, Zwinderman AH, Meijboom FJ, Bouma BJ, Mulder BJ. The changing epidemiology of congenital heart disease. *Nat Rev Cardiol* 2011;**8**:50–60.
8. McBride KL, Riley MF, Zender GA, Fitzgerald-Butt SM, Towbin JA, Belmont JW, Cole SE. NOTCH1 mutations in individuals with left ventricular outflow tract malformations reduce ligand-induced signaling. *Hum Mol Genet* 2008;**17**:2886–2893.
9. Kerstjens-Frederikse WS, van de Laar IM, Vos YJ, Verhagen JM, Berger RM, Lichtenbelt KD, Klein Wassink-Ruiter JS, van der Zwaag PA, Du Marchie Sarvaas GJ, Bergman KA, Bilardo CM, Roos-Hesselink JW, Janssen JH, Frohn-Mulder IM, van Spaendonck-Zwarts KY, van Melle JP, Hofstra RM, Wessels MW. Cardiovascular malformations caused by *NOTCH1* mutations do not keep left: data on 428 probands with left-sided CHD and their families. *Genet Med* 2016;**18**:914–923.
10. Freylikhman O, Tatarinova T, Smolina N, Zhuk S, Klyushina A, Kiselev A, Moiseeva O, Sjoberg G, Malashicheva A, Kostareva A. Variants in the *NOTCH1* gene in patients with aortic coarctation. *Congenit Heart Dis* 2014;**9**:391–396.
11. Arrington CB, Bleyl SB, Matsunami N, Bonnell GD, Otterud BE, Nielsen DC, Stevens J, Levy S, Leppert MF, Bowles NE. Exome analysis of a family with pleiotropic congenital heart disease. *Circ Cardiovasc Genet* 2012;**5**:175–182.
12. Tomita-Mitchell A, Stamm KD, Mahnke DK, Kim MS, Hidestrand PM, Liang HL, Goetsch MA, Hidestrand M, Simpson P, Pelech AN, Tweddell JS, Benson DW, Lough JW, Mitchell ME. Impact of *MYH6* variants in hypoplastic left heart syndrome. *Physiol Genomics* 2016;**48**:912–921.
13. Tan HL, Glen E, Topf A, Hall D, O'Sullivan JJ, Sneddon L, Wren C, Avery P, Lewis RJ, ten Dijke P, Arthur HM, Goodship JA, Keavney BD. Nonsynonymous variants in the *SMAD6* gene predispose to congenital cardiovascular malformation. *Hum Mutat* 2012;**33**:720–727.
14. McElhinney DB, Geiger E, Blinder J, Benson DW, Goldmuntz E. *NKX2.5* mutations in patients with congenital heart disease. *J Am Coll Cardiol* 2003;**42**:1650–1655.
15. Bonachea EM, Chang SW, Zender G, LaHaye S, Fitzgerald-Butt S, McBride KL, Garg V. Rare *GATA5* sequence variants identified in individuals with bicuspid aortic valve. *Pediatr Res* 2014;**76**:211–216.
16. Lalani SR, Ware SM, Wang X, Zapata G, Tian Q, Franco LM, Jiang Z, Bucasas K, Scott DA, Campeau PM, Hanchard N, Umana L, Cast A, Patel A, Cheung SW, McBride KL, Bray M, Craig Chinault A, Boggs BA, Huang M, Baker MR, Hamilton S, Towbin J, Jefferies JL, Fernbach SD, Potocki L, Belmont JW. *MCTP2* is a dosage-sensitive gene required for cardiac outflow tract development. *Hum Mol Genet* 2013;**22**:4339–4348.
17. Quintero-Rivera F, Xi QJ, Keppler-Noreuil KM, Lee JH, Higgins AW, Anchan RM, Roberts AE, Seong IS, Fan X, Lage K, Lu LY, Tao J, Hu X, Berezney R, Gelb BD, Kamp A, Moskowitz IP, Lacro RV, Lu W, Morton CC, Gusella JF, Maas RL. *MATR3* disruption in human and mouse associated with bicuspid aortic valve, aortic coarctation and patent ductus arteriosus. *Hum Mol Genet* 2015;**24**:2375–2389.
18. Sanchez-Castro M, Eldjouzi H, Charpentier E, Busson PF, Hauet Q, Lindenbaum P, Delasalle-Guyomarch B, Baudry A, Pichon O, Pascal C, Lefort B, Bajolle F, Pezard P, Schott JJ, Dina C, Redon R, Gournay V, Bonnet D, Le Caignec C. Search for rare copy-number variants in congenital heart defects identifies novel candidate genes and a potential role for *FOXC1* in patients with coarctation of the aorta. *Circ Cardiovasc Genet* 2016;**9**:86–94.
19. Kong A, Masson G, Frigge ML, Gylfason A, Zusmanovich P, Thorleifsson G, Olason PI, Ingason A, Steinberg S, Rafnar T, Sulem P, Mouy M, Jonsson F, Thorsteinsdottir U, Gudbjartsson DF, Stefansson H, Stefansson K. Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat Genet* 2008;**40**:1068–1075.
20. Kong A, Steinthorsdottir V, Masson G, Thorleifsson G, Sulem P, Besenbacher S, Jonasdottir A, Sigurdsson A, Kristinsson KT, Jonasdottir A, Frigge ML, Gylfason A, Olason PI, Gudjonsson SA, Sverrisson S, Stacey SN, Sigurgeirsson B, Benediktsdottir KR, Sigurdsson H, Jonsson T, Benediktsson R, Olafsson JH, Johannsson OT, Hreidarsson AB, Sigurdsson G, Ferguson-Smith AC, Gudbjartsson DF, Thorsteinsdottir U, Stefansson K. Parental origin of sequence variants associated with complex diseases. *Nature* 2009;**462**:868–874.
21. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007;**39**:906–913.
22. Gudbjartsson DF, Sulem P, Helgason H, Gylfason A, Gudjonsson SA, Zink F, Oddson A, Magnusson G, Halldorsson BV, Hjartarson E, Sigurdsson GT, Kong A, Helgason A, Masson G, Magnusson OT, Thorsteinsdottir U, Stefansson K. Sequence variants from whole genome sequencing a large group of Icelanders. *Sci Data* 2015;**2**:150011.
23. Styrkarsdottir U, Thorleifsson G, Sulem P, Gudbjartsson DF, Sigurdsson A, Jonasdottir A, Jonasdottir A, Oddsson A, Helgason A, Magnusson OT, Walters GB, Frigge ML, Helgadottir HT, Johannsdottir H, Bergsteinsdottir K, Ogmundsdottir MH, Center JR, Nguyen TV, Eisman JA, Christiansen C, Steingrimsson E, Jonasson JG, Tryggvadottir L, Eyjolfsson GI, Theodors A, Jonsson T, Ingvarsson T, Olafsson I, Rafnar T, Kong A, Sigurdsson G, Masson G, Thorsteinsdottir U, Stefansson K. Nonsense mutation in the *LGR4* gene is associated with several human diseases and other traits. *Nature* 2013;**497**:517–520.
24. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, Daly MJ, Price AL, Neale BM. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015;**47**:291–295.
25. Holm S. A simple sequentially rejective multiple test procedure. *Scand J Stat* 1979;**6**:65–70.
26. Fuster V, Harrington RA, Narula J, Eapen ZJ. *Hurst's the Heart*. 14th ed. New York City: McGraw-Hill Education; 2017.
27. Sveinbjornsson G, Albrechtsen A, Zink F, Gudjonsson SA, Oddson A, Masson G, Holm H, Kong A, Thorsteinsdottir U, Sulem P, Gudbjartsson DF, Stefansson K. Weighting sequence variants based on their annotation increases power of whole-genome association studies. *Nat Genet* 2016;**48**:314–317.
28. Sweeney HL, Houdusse A. Myosin VI rewrites the rules for myosin motors. *Cell* 2010;**141**:573–582.
29. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J,

Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;**536**:285–291.
30. Teo LL, Cannell T, Babu-Narayan SV, Hughes M, Mohiaddin RH. Prevalence of associated cardiovascular abnormalities in 500 patients with aortic coarctation referred for cardiovascular magnetic resonance imaging to a tertiary center. *Pediatr Cardiol* 2011;**32**:1120–1127.
31. Holm H, Gudbjartsson DF, Sulem P, Masson G, Helgadottir HT, Zanon C, Magnusson OT, Helgason A, Saemundsdottir J, Gylfason A, Stefansdottir H, Gretarsdottir S, Matthiasson SE, Thorgeirsson GM, Jonasdottir A, Sigurdsson A, Stefansson H, Werge T, Rafnar T, Kiemeney LA, Parvez B, Muhammad R, Roden DM, Darbar D, Thorleifsson G, Walters GB, Kong A, Thorsteinsdottir U, Arnar DO, Stefansson K. A rare variant in *MYH6* is associated with high risk of sick sinus syndrome. *Nat Genet* 2011;**43**:316–320.
32. Thorolfsdottir RB, Sveinbjornsson G, Sulem P, Helgadottir A, Gretarsdottir S, Benonisdottir S, Magnusdottir A, Davidsson OB, Rajamani S, Roden DM, Darbar D, Pedersen TR, Sabatine MS, Jonsdottir I, Arnar DO, Thorsteinsdottir U, Gudbjartsson DF, Holm H, Stefansson K. A missense variant in *PLEC* increases risk of atrial fibrillation. *J Am Coll Cardiol* 2017;**70**:2157–2168.
33. Gudbjartsson DF, Helgason H, Gudjonsson SA, Zink F, Oddson A, Gylfason A, Besenbacher S, Magnusson G, Halldorsson BV, Hjartarson E, Sigurdsson GT, Stacey SN, Frigge ML, Holm H, Saemundsdottir J, Helgadottir HT, Johannsdottir H, Sigfusson G, Thorgeirsson G, Sverrisson JT, Gretarsdottir S, Walters GB, Rafnar T, Thjodleifsson B, Bjornsson ES, Olafsson S, Thorarinsdottir H, Steingrimsdottir T, Gudmundsdottir TS, Theodors A, Jonasson JG, Sigurdsson A, Bjornsdottir G, Jonsson JJ, Thorarensen O, Ludvigsson P, Gudbjartsson H, Eyjolfsson GI, Sigurdardottir O, Olafsson I, Arnar DO, Magnusson OT, Kong A, Masson G, Thorsteinsdottir U, Helgason A, Sulem P, Stefansson K. Large-scale whole-genome sequencing of the Icelandic population. *Nat Genet* 2015;**47**:435–444.
34. Franco D, Lamers WH, Moorman AF. Patterns of expression in the developing myocardium: towards a morphologically integrated transcriptional model. *Cardiovasc Res* 1998;**38**:25–53.
35. GTEx Consortium. The genotype-tissue expression (GTEx) project. *Nat Genet* 2013;**45**:580–585.
36. Rudolph AM, Heymann MA, Spitznas U. Hemodynamic considerations in the development of narrowing of the aorta. *Am J Cardiol* 1972;**30**:514–525.
37. Ishikawa T, Jou CJ, Nogami A, Kowase S, Arrington CB, Barnett SM, Harrell DT, Arimura T, Tsuji Y, Kimura A, Makita N. Novel mutation in the alpha-myosin heavy chain gene is associated with sick sinus syndrome. *Circ Arrhythm Electrophysiol* 2015;**8**:400–408.
38. Granados-Riveron JT, Ghosh TK, Pope M, Bu'Lock F, Thornborough C, Eason J, Kirk EP, Fatkin D, Feneley MP, Harvey RP, Armour JA, David Brook J. Alpha-cardiac myosin heavy chain (*MYH6*) mutations affecting myofibril formation are associated with congenital heart defects. *Hum Mol Genet* 2010;**19**:4007–4016.
39. Ching YH, Ghosh TK, Cross SJ, Packham EA, Honeyman L, Loughna S, Robinson TE, Dearlove AM, Ribas G, Bonser AJ, Thomas NR, Scotter AJ, Caves LS, Tyrrell GP, Newbury-Ecob RA, Munnich A, Bonnet D, Brook JD. Mutation in myosin heavy chain 6 causes atrial septal defect. *Nat Genet* 2005;**37**:423–428.
40. Posch MG, Waldmuller S, Müller M, Scheffold T, Fournier D, Andrade-Navarro MA, De Geeter B, Guillaumont S, Dauphin C, Yousseff D, Schmitt KR, Perrot A, Berger F, Hetzer R, Bouvagnet P, Özcelik C. Cardiac alpha-myosin (*MYH6*) is the predominant sarcomeric disease gene for familial atrial septal defects. *PLoS One* 2011;**6**:e28872.
41. Carniel E, Taylor MR, Sinagra G, Di Lenarda A, Ku L, Fain PR, Boucek MM, Cavanaugh J, Miocic S, Slavov D, Graw SL, Feiger J, Zhu XZ, Dao D, Ferguson DA, Bristow MR, Mestroni L. Alpha-myosin heavy chain: a sarcomeric gene associated with dilated and hypertrophic phenotypes of cardiomyopathy. *Circulation* 2005;**112**:54–59.

Appendix 1

Top 20 associations of coding variants in *PLEC* with AF and results of conditional analysis (paper I)

Rs name	Position (hg38)	Amaj/ Amin[a]	MAF[b] (%)	Coding change	Coding effect	Info	OR[c]	*P*	OR[c] adjusted	*P* adjusted	Covariate
rs373243633	chr8:143917940	C/T	1.2	p.Gly4098Ser	missense	0.99	1.55	7.8×10^{-10}	1.00	1	p.Gly4098Ser
rs11136336	chr8:143933019	G/A	37.6	p.Ala531Val	missense	1.00	0.95	0.0025	0.96	0.028	p.Gly4098Ser
rs77303974	chr8:143919896	G/A	0.6	p.Arg3336Cys	missense	0.99	1.38	0.0027	1.38	0.0035	p.Gly4098Ser
rs200206105	chr8:143917631	C/T	0.02	p.Glu4091Lys	missense	0.99	3.39	0.013	3.71	0.0087	p.Gly4098Ser
rs201051109	chr8:143917634	C/T	0.2	p.Val4090Met	missense	0.99	1.48	0.014	1.50	0.012	p.Gly4098Ser
rs774270631	chr8:143924598	C/G	0.2	p.Glu1804Asp	missense	0.99	0.51	0.014	0.47	0.0072	p.Gly4098Ser
rs200272804	chr8:143923147	T/C	0.1	p.Asn2288Ser	missense	0.99	1.71	0.015	1.79	0.0099	p.Gly4098Ser
rs201430180	chr8:143925331	C/T	0.1	p.Arg1560Gln	missense	0.99	0.50	0.015	0.50	0.019	p.Gly4098Ser
rs35261863	chr8:143919763	C/G	1.9	p.Gly3380Ala	missense	1.00	1.16	0.016	1.18	0.0068	p.Gly4098Ser
rs6558407	chr8:143921326	C/T	39.8	p.Arg2859His	missense	1.00	0.96	0.017	0.97	0.12	p.Gly4098Ser
rs7002002	chr8:143923759	G/A	39.8	p.Ala2084Val	missense	1.00	0.96	0.018	0.97	0.12	p.Gly4098Ser
rs11136333	chr8:143927341	T/C	39.8	c.3838-6A>G	splice region	1.00	0.96	0.018	0.97	0.12	p.Gly4098Ser
rs7002152	chr8:143925888	T/C	39.8	c.4126-4A>G	splice region	1.00	0.96	0.018	0.97	0.12	p.Gly4098Ser
rs11136334	chr8:143927420	C/T	39.8	p.Arg1276Gln	missense	1.00	0.96	0.018	0.97	0.12	p.Gly4098Ser
rs58308209	chr8:143921001	G/C	0.8	p.Asp2967Glu	missense	1.00	1.23	0.024	1.23	0.028	p.Gly4098Ser
rs768113059	chr8:143920742	C/T	0.1	p.Ala3054Thr	missense	0.98	1.69	0.024	1.68	0.031	p.Gly4098Ser
rs28526657	chr8:143921809	C/T	0.8	p.Arg2698Gln	missense	1.00	1.22	0.027	1.22	0.030	p.Gly4098Ser
rs7833924	chr8:143921861	A/G	40.6	p.Ser2681Pro	missense	1.00	0.97	0.049	0.98	0.24	p.Gly4098Ser
rs55895668	chr8:143926863	T/C	40.6	p.His1349Arg	missense	1.00	0.97	0.049	0.98	0.24	p.Gly4098Ser
rs763436354	chr8:143919232	C/T	0.1	p.Cys3557Tyr	missense	0.99	1.57	0.066	1.66	0.044	p.Gly4098Ser

[a]Amaj/Amin = major allele/minor allele; [b]MAF = minor allele frequency; [c]OR = odds ratio

Appendix 2

Association of rs72700114 at *LINC01142/METTL11B* with AF in Iceland and follow up sample sets (paper I)

Icelandic GWAS			
N cases	**N controls**	**OR[a] (CI[b])**	***P* value**
13,471	374,939	1.22 (1.15-1.29)	7.0×10^{-12}
FOURIER samples			
N cases	**N controls**	**OR[a] (CI[b])**	***P* value**
1,238	11,562	1.30 (1.13-1.49)	1.9×10^{-4}
Vanderbilt AF registry			
N cases	**N controls**	**OR[a] (CI[b])**	***P* value**
764	762	1.99 (1.55-2.55)	6.5×10^{-8}
Combined results			
N cases	**N controls**	**OR[a] (CI[b])**	***P* value**
15,473	387,263	1.26 (1.19-1.32)	3.1×10^{-18}

Significant heterogeneity was observed between the non-Icelandic and Icelandic samples ($P = 7.8\times10^{-4}$). The heterogeneity is driven by the high OR observed in the Vanderbilt sample set.

[a]OR = odds ratio; [b]CI = 95% confidence interval

Appendix 3

Results of conditional analysis between rs72700114 at *METTL11B/LINC01142* and a previously reported signal 400 kb away, rs651386, close to *PRRX1* in the Icelandic GWAS on AF (paper I)

RS name	Position (hg38)	Risk allele/ other	RAF (%)[a]	OR[b]	*P*	OR adjusted	*P* adjusted	Covariate
rs7270011	Chr1:170224684	C/G	8.1	1.22	7.04×10^{-12}	1.22	4.27×10^{-11}	Chr1: 170622169
rs651386	Chr1: 170622169	A/T	57.4	1.11	1.94×10^{-10}	1.10	6.05×10^{-08}	Chr1:170224684

[a]RAF = risk allele frequency; [b]OR = odds ratio

Appendix 4

Associations of low frequency (<5%) imputed moderate and high impact variants in *MYZAP* with AF in Iceland (N = 14,710) and results of conditional analysis with p.Gln254Pro as covariate (paper II)

Rs name	Position (hg38)	Amaj/ Amin[a]	MAF[b] (%)	Coding change	Coding effect	Info	OR[c]	*P*	OR[c] adjusted	*P* adjusted	Covariate
rs147301839	chr15:57632516	A/C	1.08	p.Gln254Pro	Missense	1.00	1.38	9.03×10^{-6}	1.00	1.00	p.Gln254Pro
rs117361082	chr15:57637771	A/G	1.38	p.Glu337Gly	Missense	1.00	1.31	4.17×10^{-5}	1.06	0.69	p.Gln254Pro
rs147173331	chr15:57618043	T/C	0.54	p.Leu58Pro	Missense	1.00	1.15	0.19	1.17	0.15	p.Gln254Pro
rs146453491	chr15:57633609	G/T	0.002	c.805-4G>T	Splice region	0.96	6.67	0.25	7.4	0.24	p.Gln254Pro
rs142253131	chr15:57604284	C/A	0.14	p.Leu31Ile	Missense	0.99	0.80	0.31	0.81	0.36	p.Gln254Pro
rs138712430	chr15:57618135	G/A	0.08	p.Val89Met	Missense	0.98	0.83	0.56	0.82	0.53	p.Gln254Pro
rs140229874	chr15:57637729	A/G	0.2	p.His323Arg	Missense	0.99	0.97	0.87	0.97	0.87	p.Gln254Pro

[a]Amaj/Amin = major allele/minor allele; [b]MAF = minor allele frequency; [c]OR = odds ratio

Appendix 5

Association of three novel AF variants (paper II) with SSS in Iceland and the UK Biobank

		MYZAP p.Gln254Pro		*RPL3L* p.Ala75Val		*RPL3L* c.1167+1G>A	
	Number of cases/controls	**OR[a] (CI[b])**	***P* value**	**OR[a] (CI[b])**	***P* value**	**OR[a] (CI[b])**	***P* value**
Iceland	3,568/346,025	1.51 (1.21-1.90)	3.5×10^{-4}	1.01 (0.88-1.17)	0.83	0.97 (0.69- 1.37)	0.85
UK	403/403,181	3.80 (1.89-7.62)	1.8×10^{-4}	0.91 (0.62-1.34)	0.62	2.45 (0.93-6.48)	0.070
Combined	3,971/749,206	1.65 (1.33-2.05)	5.0×10^{-6}	1.00 (0.88-1.14)	0.97	1.07 (0.78-1.48)	0.67

[a]OR = odds ratio; [b]CI = 95% confidence interval

Appendix 6

Conditional analysis of two *RPL3L* variants on each other with regards to AF association in Iceland (paper II)

Variant tested	Variant conditioned on	OR[a] adjusted	*P* adjusted
c.1167+1G>A	p.Ala75Val	1.40	$7.1x10^{-4}$
p.Ala75Val	c.1167+AG>A	1.20	$1.6x10^{-5}$

[a]OR = odds ratio

Appendix 7

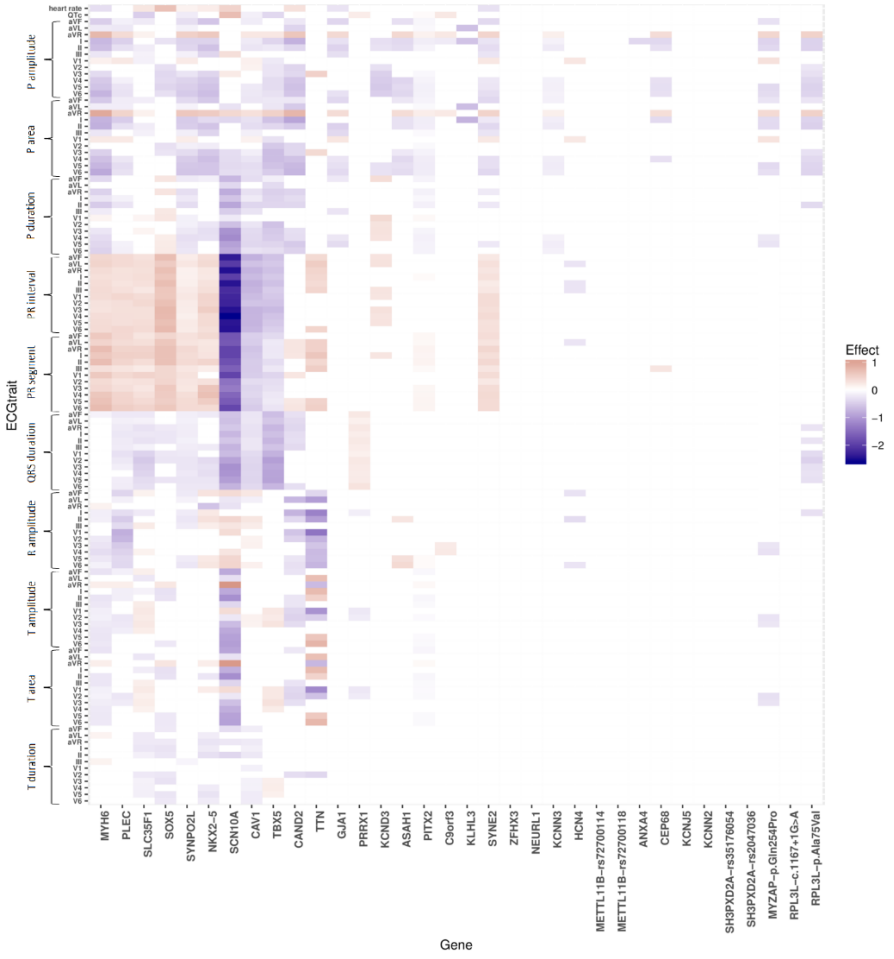
Associations of low frequency (<5%) imputed moderate and high impact variants in *RPL3L* with AF in Iceland (N = 14,710) and results of conditional analysis with p.Ala75Val and c.1167+1G>A as covariates (paper II)

Rs name	Position (hg38)	Amaj/ Amin[a]	MAF[b] (%)	Coding change	Coding effect	Info	OR[c]	*P*	OR[c] adjusted	*P* adjusted	Covariate
rs140185678	chr16:1953015	G/A	3.6	p.Ala75Val	missense	0.99	1.18	6.4×10^{-5}	1.00	1.00	p.Ala75Val, c.1167+1G>A
rs140192228	chr16:1945498	C/T	0.6	c.1167+1G>A	Splice	0.99	1.37	8.7×10^{-4}	1.00	1.00	p.Ala75Val, c.1167+1G>A
rs146294352	chr16:1954118	C/T	0.8	p.Gly12Arg	missense	0.99	1.29	0.0029	1.32	0.0018	p.Ala75Val, c.1167+1G>A
rs113956264	chr16:1947003	C/T	4.4	p.Val262Met	missense	0.99	1.06	0.16	1.06	0.12	p.Ala75Val, c.1167+1G>A
rs147948209	chr16:1945869	G/C	1.1	p.Ala338Gly	missense	0.98	0.90	0.20	0.92	0.30	p.Ala75Val, c.1167+1G>A
rs147972626	chr16:1947063	G/A	0.7	p.Arg242Trp	missense	0.99	1.13	0.20	1.16	0.13	p.Ala75Val, c.1167+1G>A
rs146749305	chr16:1954028	G/A	0.004	p.His42Tyr	missense	0.99	0.02	0.22	0.02	0.23	p.Ala75Val, c.1167+1G>A
rs75401081	chr16:1947205	C/T	0.002	p.Arg226Gln	missense	0.95	0.02	0.36	0.02	0.37	p.Ala75Val, c.1167+1G>A
rs140116056	chr16:1945588	C/T	0.4	p.Val360Met	missense	0.99	0.91	0.42	0.90	0.41	p.Ala75Val, c.1167+1G>A
rs141796888	chr16:1947283	C/T	1.1	p.Arg200Gln	missense	0.98	1.05	0.52	1.08	0.35	p.Ala75Val, c.1167+1G>A
rs79075024	chr16:1947384	C/G	1.8	c.502-4G>C	splice regio	0.99	1.03	0.59	1.06	0.36	p.Ala75Val, c.1167+1G>A
rs201602086	chr16:1952995	G/C	0.002	p.Pro82Ala	missense	1.00	0.02	0.61	0.018	0.62	p.Ala75Val, c.1167+1G>A
rs201864074	chr16:1954141	C/T	0.001	p.Arg4Gln	missense	0.96	1.97	0.63	2.35	0.56	p.Ala75Val, c.1167+1G>A
rs34265469	chr16:1946704	G/A	0.03	p.Pro291Leu	missense	0.99	0.84	0.73	0.81	0.68	p.Ala75Val, c.1167+1G>A
rs118144581	chr16:1952959	C/T	0.02	p.Ala94Thr	missense	0.97	0.88	0.79	0.92	0.86	p.Ala75Val, c.1167+1G>A

[a]Amaj/Amin = major allele/minor allele; [b]MAF = minor allele frequency; [c]OR = odds ratio

Appendix 8

Heatmap showing the effects of AF variants on ECG traits of all ECGs (434,000 ECGs from 88,217 individuals), not excluding AF cases (papers I and II)



Appendix 9

Identified known (n = 31) and novel (n = 80) AF risk loci in a GWAS meta-analysis of 60,620 AF cases and 970,216 controls (paper III)

Rs name	Position (hg19)	Risk/ref[a] allele	RAF[b] (%)	OR (CI)[c]	*P* value	*P* het[d]	Novelty	Annotation	Prioritized genes
rs284277	chr1:10790797	C/A	38.3	1.04 (1.03-1.06)	1.2×10^{-9}	0.73	novel	intronic/*CASZ1*	*CASZ1*
rs7529220	chr1:22282619	C/T	84.7	1.06 (1.04-1.08)	2.0×10^{-10}	0.52	novel	intergenic/.	*HSPG2*
rs2885697	chr1:41544279	G/T	35.2	1.04 (1.03-1.06)	2.9×10^{-10}	0.91	novel	intronic/*SCMH1*	*SCMH1*
rs11590635	chr1:49309764	A/G	2.4	1.16 (1.10-1.21)	4.1×10^{-9}	0.040	novel	intronic/*AGBL4*	*AGBL4*
rs146518726	chr1:51535039	A/G	3.3	1.17 (1.13-1.22)	8.3×10^{-15}	0.095	known	intergenic/.	*MIR6500*
rs1545300	chr1:112464004	C/T	69.1	1.06 (1.04-1.07)	1.5×10^{-14}	0.17	known	intronic/*KCND3*	*KCND3*
rs4073778	chr1:116297758	A/C	56.4	1.05 (1.04-1.06)	5.0×10^{-13}	0.021	novel	intronic/*CASQ2*	*CASQ2*
rs79187193	chr1:147255831	G/A	94.3	1.12 (1.09-1.16)	3.2×10^{-14}	0.94	novel	intergenic/.	*GJA5*
rs11264280	chr1:154862952	T/C	33.3	1.14 (1.13-1.16)	3.1×10^{-79}	0.21	known	intergenic/.	*KCNN3*
rs72700114	chr1:170193825	C/G	7.6	1.22 (1.19-1.26)	3.3×10^{-54}	7.38×10^{-5}	known	intergenic/.	*LINC01142*
rs10753933	chr1:203026214	T/G	4.8	1.06 (1.05-1.08)	9.8×10^{-20}	0.16	known	intronic/*PPFIA4*	*PPFIA4*
rs4951258	chr1:205691316	A/G	41.6	1.04 (1.02-1.05)	2.1×10^{-8}	0.65	novel	intronic/*NUCKS1*	*NUCKS1,SLC41A1*
rs7578393	chr2:26165528	T/C	79.6	1.06 (1.05-1.08)	2.4×10^{-12}	0.73	novel	intronic/*KIF3C*	*KIF3C*
rs11125871	chr2:61470126	C/T	60.5	1.04 (1.03-1.05)	6.4×10^{-9}	0.62	novel	intronic/*USP34*	*USP34*
rs2540949	chr2:65284231	A/T	61.5	1.07 (1.05-1.08)	3.0×10^{-22}	0.23	known	intronic/*CEP68*	*CEP68*
rs6747542	chr2:70106832	T/C	53.6	1.06 (1.04-1.07)	1.1×10^{-16}	0.042	known	downstream/*GMCL1*	*GMCL1,ANXA4*
rs72926475	chr2:86594487	G/A	87.7	1.07 (1.05-1.09)	2.4×10^{-11}	5.93×10^{-4}	novel	intergenic/.	*REEP1*
rs28387148	chr2:127433465	T/C	10.5	1.08 (1.05-1.10)	6.3×10^{-11}	0.23	novel	intronic/*GYPC*	*GYPC*
rs67969609	chr2:145760353	G/C	7.1	1.07 (1.05-1.10)	1.7×10^{-8}	0.22	novel	ncRNA_intronic/ *TEX41*	*TEX41*
rs56181519	chr2:175555714	C/T	73.2	1.07 (1.05-1.08)	6.5×10^{-18}	0.65	novel	intergenic/.	*WIPF1*
rs2288327	chr2:179411665	G/A	15.6	1.10 (1.08-1.12)	7.3×10^{-25}	0.025	known	intronic/*TTN*	*TTN,MIR548N,FKBP7,TTN-AS1*
rs3820888	chr2:201180023	C/T	39.2	1.07 (1.06-1.09)	5.8×10^{-24}	0.27	novel	intronic/*SPATS2L*	*SPATS2L*
rs35544454	chr2:213266003	A/T	80.8	1.06 (1.04-1.08)	1.1×10^{-11}	0.21	novel	intronic/*ERBB4*	*ERBB4*
rs7650482	chr3:12841804	G/A	64.0	1.07 (1.06-1.09)	1.8×10^{-24}	0.33	known	intronic/*CAND2*	*CAND2*
rs73041705	chr3:24463235	T/C	70.2	1.05 (1.03-1.06)	1.5×10^{-9}	0.16	novel	intronic/*THRB*	*THRB*
rs6790396	chr3:38771925	G/C	59.6	1.06 (1.05-1.08)	2.4×10^{-20}	0.93	known	intronic/*SCN10A*	*SCN10A,SCN5A*
rs34080181	chr3:66454191	G/A	62.1	1.05 (1.03-1.06)	1.3×10^{-10}	0.47	novel	intronic/*LRIG1*	*LRIG1,SLC25A26*
rs17005647	chr3:69406181	T/C	36.4	1.04 (1.03-1.06)	2.7×10^{09}	0.31	novel	intronic/*FRMD4B*	*FRMD4B*
rs6771054	chr3:89489529	T/C	59.6	1.05 (1.03-1.06)	2.4×10^{-11}	8.19×10^{-4}	novel	intronic/*EPHA3*	*EPHA3*
rs10804493	chr3:111554426	A/G	65.1	1.06 (1.04-1.07)	1.6×10^{-15}	0.36	novel	intronic/*PHLDB2*	*PHLDB2,PLCXD2*
rs1278493	chr3:135814009	G/A	43.6	1.04 (1.03-1.05)	8.8×10^{-9}	0.58	novel	intronic/*PPP2R3A*	*PPP2R3A*
rs7612445	chr3:179172979	T/G	18.8	1.05 (1.03-1.07)	4.8×10^{-9}	0.70	novel	upstream/*GNB4*	*GNB4*
rs60902112	chr3:194800853	T/C	22.6	1.05 (1.03-1.06)	1.7×10^{-8}	0.70	novel	intronic/*XXYLT1*	*XXYLT1*
rs1458038	chr4:81164723	T/C	30.9	1.04 (1.03-1.06)	1.7×10^{-9}	0.82	novel	intergenic/.	*FGF5*
rs10006327	chr4:103890980	C/T	49.0	1.04 (1.02-1.05)	4.4×10^{-8}	0.80	novel	intronic/*SLC9B1*	*SLC9B1*
rs67249485	chr4:111699685	T/A	19.9	1.44 (1.42-1.46)	7.3×10^{-443}	8.24×10^{-10}	known	intergenic/.	*PITX2*
rs6829664	chr4:114448656	G/A	26.2	1.06 (1.04-1.07)	1.9×10^{-13}	0.15	novel	intronic/*CAMK2D*	*CAMK2D*
rs10213171	chr4:148937537	G/C	6.1	1.10 (1.07-1.12)	1.3×10^{-11}	0.51	novel	intronic/*ARHGAP10*	*ARHGAP10*
rs12648245	chr4:174641184	T/C	92.4	1.10 (1.07-1.12)	3.5×10^{-13}	0.61	known	intergenic/.	*HAND2,HAND2-AS1*
rs6596717	chr5:106427609	C/A	39.5	1.04 (1.03-1.06)	3.0×10^{-9}	0.63	novel	intergenic/.	*LOC102467213*
rs337705	chr5:113737062	G/T	37.5	1.06 (1.04-1.07)	1.6×10^{-16}	0.84	known	intronic/*KCNN2*	*KCNN2*
rs2012809	chr5:128190363	G/A	79.0	1.06 (1.04-1.08)	4.9×10^{-10}	0.45	novel	intergenic/.	*SLC27A6*
rs2040862	chr5:137419989	T/C	17.8	1.11 (1.10-1.13)	1.1×10^{-35}	0.12	known	intronic/*WNT8A*	*WNT8A,NPY6R,MYOT,FAM13B*
rs6580277	chr5:142818123	G/A	23.7	1.07 (1.05-1.09)	1.6×10^{-17}	0.33	novel	upstream/*NR3C1*	*NR3C1*
rs12188351	chr5:168386089	A/G	5.6	1.09 (1.06-1.12)	2.5×10^{-9}	0.83	novel	intronic/*SLIT3*	*SLIT3*
rs6891790	chr5:172670745	G/T	71.7	1.08 (1.06-1.09)	4.5×10^{-22}	0.11	novel	intergenic/.	*NKX2-5*
rs73366713	chr6:16415751	G/A	86.0	1.11 (1.09-1.13)	1.5×10^{-25}	0.21	novel	intronic/*ATXN1*	*ATXN1*
rs34969716	chr6:18210109	A/G	30.5	1.07 (1.06-1.09)	1.6×10^{-19}	0.18	novel	intronic/*KDM1B*	*KDM1B,DEK*
rs3176326	chr6:36647289	G/A	80.2	1.06 (1.05-1.08)	1.4×10^{-13}	0.40	novel	intronic/*CDKN1A*	*CDKN1A,PANDAR, PI16*
rs2031522	chr6:87821501	A/G	62.4	1.04 (1.03-1.06)	1.5×10^{-10}	0.91	novel	intergenic/.	*CGA*
rs3951016	chr6:118559658	A/T	45.9	1.07 (1.05-1.08)	2.1×10^{-22}	0.21	known	intronic/*SLC35F1*	*SLC35F1,PLN*
rs13195459	chr6:122403559	G/A	63.8	1.06 (1.05-1.08)	4.2×10^{-19}	0.063	known	intergenic/.	*HSF2*
rs117984853	chr6:149399100	T/G	10.1	1.13 (1.10-1.16)	1.3×10^{-24}	0.22	novel	downstream/*UST*	*UST*
rs55734480	chr7:14372009	A/G	24.9	1.06 (1.04-1.07)	2.2×10^{-12}	0.11	novel	intronic/*DGKB*	*DGKB*

[a]ref allele = reference allele, [b]RAF= Risk allele frequency; [c]OR (CI) = Odds ratio (95 % confidence interval); [d]*P* het = *P* heterogeneity

Continued on the next page

Appendix 9 - continued

Rs name	Position (hg19)	Risk/ref[a] allele	RAF[b] (%)	OR (CI)[c]	*P* value	*P* het[d]	Novelty	Annotation	Prioritized genes
rs6462079	chr7:28415827	A/G	72.1	1.05 (1.03-1.06)	8.8×10^{-10}	0.34	novel	intronic/*CREB5*	*CREB5*
rs35005436	chr7:74134911	C/T	15.5	1.06 (1.04-1.08)	3.3×10^{-10}	0.93	novel	intronic/*GTF2I*	*GTF2I,LOC101926943,GTF2IRD2*
rs56201652	chr7:92278116	G/A	73.3	1.05 (1.04-1.07)	1.7×10^{-12}	0.21	novel	intronic/*CDK6*	*CDK6*
rs11773845	chr7:116191301	A/C	58.6	1.11 (1.10-1.13)	2.4×10^{-55}	0.064	known	intronic/*CAV1*	*CAV1,CAV2*
rs55985730	chr7:128417044	G/T	6.0	1.09 (1.06-1.12)	5.2×10^{-9}	0.95	novel	upstream/*OPN1SW*	*OPN1SW,CALU*
rs7789146	chr7:150661409	G/A	82.1	1.06 (1.04-1.08)	2.1×10^{-11}	0.84	novel	intronic/*KCNH2*	*KCNH2*
rs35620480	chr8:11499908	C/A	15.7	1.06 (1.04-1.07)	5.2×10^{-9}	0.41	novel	intergenic/.	*GATA4*
rs7508	chr8:17913970	A/G	71.1	1.07 (1.06-1.09)	1.7×10^{-21}	0.39	known	UTR3/*ASAH1*	*ASAH1*
rs7834729	chr8:21821778	G/T	88.5	1.07 (1.05-1.09)	3.6×10^{-10}	0.24	novel	intronic/*XPO7*	*XPO7*
rs62521286	chr8:124551975	G/A	6.6	1.13 (1.10-1.16)	4.5×10^{-19}	0.58	novel	intronic/*FBXO32*	*FBXO32*
rs6994744	chr8:141740868	C/A	49.5	1.04 (1.03-1.05)	1.1×10^{-9}	0.19	novel	intronic/*PTK2*	*PTK2*
rs10821415	chr9:97713459	A/C	41.3	1.09 (1.07-1.10)	2.9×10^{-34}	0.26	known	intronic/*C9orf3*	*C9orf3*
rs2274115	chr9:139094773	G/A	70.0	1.05 (1.03-1.07)	1.7×10^{-10}	0.76	novel	intronic/*LHX3*	*LHX3*
rs12245149	chr10:65321147	C/A	52.6	1.05 (1.03-1.06)	1.7×10^{-12}	0.35	novel	intronic/*REEP3*	*REEP3,NRBF2*
rs7096385	chr10:69664881	T/C	9.2	1.07 (1.05-1.10)	4.9×10^{-08}	0.24	novel	intronic/*SIRT1*	*SIRT1,MYPN*
rs60212594	chr10:75414344	G/C	85.6	1.12 (1.10-1.15)	9.2×10^{-35}	0.14	known	intronic/*SYNPO2L*	*SYNPO2L,NUDT13, MYOZ1,AGAP5*
rs10458660	chr10:77936576	G/A	17.3	1.06 (1.04-1.07)	6.8×10^{-10}	0.57	novel	intronic/C10orf11	*C10orf11*
rs11598047	chr10:105342672	G/A	16.2	1.17 (1.15-1.19)	9.0×10^{-66}	0.18	known	intronic/*NEURL1*	*NEURL1*
rs10749053	chr10:112576695	T/C	15.8	1.06 (1.04-1.08)	1.0×10^{-8}	0.98	novel	intronic/*RBM20*	*RBM20*
rs10741807	chr11:20011445	T/C	24.5	1.08 (1.06-1.09)	1.6×10^{-20}	0.60	novel	intronic/*NAV2*	*NAV2*
rs4935786	chr11:121661507	T/A	26.7	1.05 (1.03-1.06)	4.9×10^{-9}	0.25	novel	intergenic/.	*SORL1*
rs76097649	chr11:128764570	A/G	9.3	1.12 (1.10-1.15)	1.3×10^{-20}	0.042	known	intronic/*KCNJ5*	*KCNJ5*
rs4963776	chr12:24779491	G/T	81.8	1.10 (1.08-1.11)	1.8×10^{-25}	0.20	known	intergenic/.	*LINC00477*
rs17380837	chr12:26345526	C/T	69.3	1.05 (1.04-1.07)	4.8×10^{-12}	0.17	novel	upstream/*SSPN*	*SSPN*
rs12809354	chr12:32978437	C/T	14.4	1.07 (1.05-1.09)	2.9×10^{-14}	0.0063	novel	intronic/*PKP2*	*PKP2*
rs2860482	chr12:57105938	A/C	27.4	1.06 (1.04-1.07)	1.2×10^{-12}	0.70	novel	downstream/*NACA*	*NACA*
rs71454237	chr12:70013415	G/A	79.1	1.06 (1.05-1.08)	1.8×10^{-13}	0.97	novel	intergenic/.	*LRRC10*
rs12426679	chr12:76237987	C/T	47.2	1.04 (1.03-1.05)	4.9×10^{-9}	0.025	novel	intergenic/.	*PHLDA1*
rs883079	chr12:114793240	T/C	70.7	1.10 (1.09-1.12)	2.8×10^{-40}	0.049	known	UTR3/*TBX5*	*TBX5*
rs10773657	chr12:123327900	C/A	13.8	1.06 (1.04-1.08)	2.5×10^{-8}	0.36	novel	intronic/*HIP1R*	*HIP1R*
rs6560886	chr12:133150210	C/T	78.8	1.05 (1.03-1.07)	1.5×10^{-8}	0.59	novel	intronic/*FBRSL1*	*FBRSL1*
rs9506925	chr13:23368943	T/C	26.7	1.05 (1.03-1.06)	2.7×10^{-9}	0.34	novel	intergenic/.	*LINC00540, LINC00621,SGCG*
rs35569628	chr13:113872712	T/C	77.7	1.05 (1.03-1.06)	1.4×10^{-8}	0.027	novel	intronic/*CUL4A*	*CUL4A*
rs422068	chr14:23864804	C/T	34.9	1.04 (1.03-1.06)	3.9×10^{-10}	0.0018	known	intronic/*MYH6*	*MYH6,MYH7*
rs11156751	chr14:32990437	C/T	28.5	1.07 (1.06-1.09)	6.9×10^{-21}	0.37	novel	intronic/*AKAP6*	*AKAP6*
rs73241997	chr14:35173775	T/C	14.2	1.08 (1.06-1.10)	2.9×10^{-15}	0.45	novel	intergenic/.	*CFL2*
rs2738413	chr14:64679960	A/G	49.5	1.08 (1.07-1.10)	2.5×10^{-31}	0.32	known	intronic/*SYNE2*	*SYNE2,MIR548AZ, ESR2,MTHFD1*
rs74884082	chr14:73249419	C/T	75.0	1.05 (1.03-1.07)	3.5×10^{-10}	0.28	novel	intronic/*DPF3*	*DPF3*
rs10873298	chr14:77426525	C/T	36.6	1.04 (1.03-1.06)	7.1×10^{-9}	0.080	novel	intergenic/.	*IRF2BPL*
rs147301839	chr15:57924714	C/A	0.7	1.39 (1.26-1.55)	1.9×10^{-10}	0.85	known	nonsynonymous/ *GCOM1*	*GCOM1/MYZAP*
rs7170477	chr15:64103777	A/G	30.4	1.04 (1.03-1.05)	5.0×10^{-8}	0.47	novel	intronic/*HERC1*	*HERC1*
rs74022964	chr15:73677264	T/C	15.7	1.12 (1.10-1.14)	3.5×10^{-36}	0.38	known	intergenic/.	*HCN4*
rs12908004	chr15:80676925	G/A	16.4	1.08 (1.06-1.10)	4.1×10^{-16}	0.08	novel	intergenic/.	*ARNT2*
rs4965430	chr15:99268850	C/G	38.6	1.05 (1.03-1.06)	1.3×10^{-10}	0.25	novel	intronic/*IGF1R*	*IGF1R*
rs140185678	chr16:2003016	A/G	3.5	1.18 (1.13-1.23)	2.4×10^{-14}	0.21	known	nonsynonymous/ *RPL3L*	*RPL3L*
rs2359171	chr16:73053022	A/T	17.6	1.19 (1.17-1.21)	4.7×10^{-91}	8.58×10^{-4}	known	intronic/*ZFHX3*	*ZFHX3*
rs7225165	chr17:1309850	G/A	88.7	1.07 (1.04-1.09)	3.2×10^{-9}	0.75	novel	intergenic/.	*YWHAE,CRK, MYO1C*
rs9899183	chr17:7452977	T/C	71.4	1.05 (1.03-1.06)	2.02×10^{-9}	0.49	novel	intronic/*TNFSF12*	*TNFSF12,TNFSF12TNFSF13, SOX15,FXR2*
rs72811294	chr17:12618680	G/C	88.7	1.07 (1.05-1.10)	9.7×10^{-12}	0.17	novel	intronic/*MYOCD*	*MYOCD*
rs11658278	chr17:38031164	T/C	47.9	1.05 (1.03-1.06)	3.5×10^{-11}	0.81	novel	intronic/*ZPBP2*	*ZPBP2,GSDMB, ORMDL3*
rs1563304	chr17:44874453	T/C	17.8	1.07 (1.05-1.09)	2.6×10^{-12}	0.43	known	intronic/*WNT3*	*WNT3*
rs12604076	chr17:76773638	T/C	47.8	1.04 (1.02-1.05)	3.6×10^{-8}	0.57	novel	intronic/*CYTH1*	*CYTH1,USP36*
rs9953366	chr18:46474192	C/T	66.3	1.05 (1.04-1.07)	1.8×10^{-11}	0.87	novel	intronic/*SMAD7*	*SMAD7*
rs8088085	chr18:48708548	A/C	53.5	1.04 (1.02-1.05)	4.8×10^{-8}	0.56	novel	intronic/*MEX3C*	*MEX3C*
rs2834618	chr21:36119111	T/G	89.4	1.10 (1.08-1.12)	3.4×10^{-17}	0.20	novel	ncRNA_intronic/ *LINC01426*	*LINC01426*
rs464901	chr22:18597502	T/C	66.5	1.05 (1.04-1.07)	1.5×10^{-12}	0.08	novel	intronic/*TUBA8*	*TUBA8*
rs133902	chr22:26164079	T/C	42.7	1.04 (1.03-1.06)	9.1×10^{-10}	0.68	novel	intronic/*MYO18B*	*MYO18B*

[a]ref allele = reference allele, [b]RAF= Risk allele frequency; [c]OR (CI) = Odds ratio (95 % confidence interval); [d]*P* het = *P* heterogeneity

Appendix 10

Additional, independent AF risk variants (n = 31) identified through approximate, stepwise conditional analyses (paper III)

Rs name	Position (hg19)	Risk/ref[a] allele	RAF[b] (%)	OR (CI)[c]	*P* value	Novelty	Annotation	Prioritized genes
rs10465885	chr1:147232740	C/T	47.2	1.04 (1.03-1.06)	1.4×10^{-10}	novel	intergenic/.	GJA5
rs6689306	chr1:154395946	A/G	41.3	1.06 (1.05-1.08)	1.2×10^{-18}	known	intergenic/.	*KCNN3*
rs4999127	chr1:154714006	A/G	83.9	1.10 (1.08-1.12)	1.8×10^{-23}	known	intergenic/.	*KCNN3*
rs577676	chr1:170587340	C/T	56.2	1.07 (1.05-1.08)	2.4×10^{-20}	known	intergenic/.	*LINC01142*
rs7374540	chr3:38634142	A/C	60.7	1.05 (1.03-1.06)	7.4×10^{-12}	known	intronic/*SCN10A*	*SCN10A, SCN5A*
rs7373065	chr3:38710315	T/C	4.9	1.23 (1.17-1.29)	2.7×10^{-16}	known	intronic/*SCN10A*	*SCN10A, SCN5A*
rs244017	chr4:111255917	T/G	77.6	1.06 (1.04-1.08)	7.7×10^{-11}	known	intergenic/.	*PITX2*
rs61501369	chr4:111524629	T/C	23.4	1.10 (1.09-1.12)	1.0×10^{-28}	known	intergenic/.	*PITX2*
rs6850025	chr4:111596360	A/G	4.9	1.12 (1.08-1.16)	9.5×10^{-11}	known	intergenic/.	*PITX2*
rs79399769	chr4:111925656	C/T	96.9	1.24 (1.18-1.30)	2.0×10^{-18}	known	intergenic/.	*PITX2*
rs1532170	chr4:112165212	G/A	45.1	1.07 (1.05-1.08)	8.0×10^{-18}	known	intergenic/.	*PITX2*
rs138311480	chr4:112454295	C/T	97.8	1.18 (1.12-1.25)	3.0×10^{-9}	known	intergenic/.	*PITX2*
rs114904067	chr4:112604821	G/A	97.2	1.13 (1.09-1.18)	8.6×10^{-9}	known	intergenic/.	*PITX2*
rs7687819	chr4:113329345	A/G	77.1	1.05 (1.03-1.06)	8.7×10^{-9}	known	intergenic/.	*PITX2*
rs10520260	chr4:174447349	A/G	67.9	1.05 (1.03-1.06)	5.9×10^{-10}	known	intergenic/.	*HAND2, HAND2-AS1*
rs28439930	chr5:173393111	G/C	51.7	1.05 (1.03-1.06)	1.9×10^{-11}	novel	intergenic/.	*NKX2-5*
rs9401451	chr6:122099152	G/A	90.0	1.08 (1.05-1.10)	3.4×10^{-11}	known	intergenic/.	*HSF2*
rs4871397	chr8:124635197	G/C	8.9	1.09 (1.06-1.12)	1.3×10^{-9}	novel	intronic/*FBXO32*	*FBXO32*
rs55693294	chr10:105277474	T/C	6.1	1.09 (1.06-1.12)	4.2×10^{-9}	known	intronic/*NEURL1*	*NEURL1*
rs2291437	chr12:24715048	G/T	10.5	1.09 (1.07-1.12)	2.0×10^{-17}	known	intergenic/.	*LINC00477*
rs11614818	chr12:56055815	C/T	36.3	1.04 (1.03-1.05)	1.9×10^{-8}	novel	downstream/ *NACA*	*NACA*
rs775498	chr12:70071513	G/A	28.0	1.04 (1.03-1.06)	9.4×10^{-9}	novel	intergenic/.	*LRRC10*
rs1957021	chr14:32924505	C/T	22.2	1.06 (1.05-1.08)	4.8×10^{-15}	novel	intronic/*AKAP6*	*AKAP6*
rs2759301	chr15:80994288	A/G	45.4	1.04 (1.03-1.05)	9.2×10^{-9}	novel	intergenic/.	*ARNT2*
rs118159104	chr16:1676804	G/T	1.4	1.20 (1.13-1.28)	1.6×10^{-8}	known	nonsynonymous/ *RPL3L*	*RPL3L*
rs77316573	chr16:2265271	T/C	19.9	1.05 (1.03-1.07)	2.4×10^{-8}	known	nonsynonymous/ *RPL3L*	*RPL3L*
rs876727	chr16:73067761	T/G	79.1	1.05 (1.03-1.07)	6.8×10^{-9}	known	intronic/*ZFHX3*	*ZFHX3*
rs9963878	chr18:48679522	C/T	8.5	1.07 (1.04-1.09)	2.5×10^{-8}	novel	intronic/*MEX3C*	*MEX3C*
rs72700118[d]	chr1:170194823	A/C	12.7	1.14 (1.12-1.16)	2.0×10^{-36}	known	intergenic/.	*LINC01142*
rs3853445[d]	chr4:111761487	T/C	73.4	1.16 (1.13-1.18)	3.7×10^{-52}	known	intergenic/.	*PITX2*
rs35176054[d]	chr10:105480387	A/T	13.0	1.15 (1.12-1.17)	8.2×10^{-41}	known	intronic/*NEURL1*	*NEURL1*

[a]ref allele = reference allele, [b]RAF= Risk allele frequency; [c]OR (CI) = Odds ratio (95 % confidence interval)
[d]Independent of locus lead variant but not from previously reported lead variants