Abstract

Atrial fibrillation (AF) is the most prevalent cardiac arrhythmia in clinical practice. Over the past two decades, we have come to appreciate that AF has a significant heritable component. The recent advent of next-generation sequencing technology has spawned a new era of research into the genetic basis of AF. Genome-wide association studies (GWAS) have identified multiple common variants underlying AF. Further, exome sequencing has emerged as a potentially powerful technique for the identification of mutations underlying familial forms of AF. In this thesis, we sought to further elucidate the genetic basis of AF though two specific aims. Firstly, we investigated the mechanistic link between KCNN3, a potassium channel gene which was identified in a GWAS for lone AF, and arrhythmia pathogenesis. Secondly, we performed exome sequencing and classical linkage analysis in two AF pedigrees to identify novel mutations for the arrhythmia. We demonstrate that overexpression of Kcnn3 in a murine model results in an increased susceptibility to AF. Interestingly, these mice also display a high incidence of sudden death due to heart block. Exome sequencing in an AF pedigree identified a potentially causative mutation in the transcription factor gene GATA6. In a second AF pedigree, we identified a novel locus for the arrhythmia on chromosome 1. However the causative mutation at this locus remains elusive. Ultimately, the identification of the genetic substrate underlying AF is likely to uncover novel therapeutic targets as well as enhancing risk stratification for this common and morbid arrhythmia.
ACKNOWLEDGEMENTS

I cannot overstate my gratitude to Dr. Patrick Ellinor for his excellent mentorship and support throughout my time in his laboratory. Dr. Ellinor selflessly dedicated time and effort to ensure that I gained the most out of the experience. He has been an inspirational role model and has played a large part in my decision to pursue a career in academia. I also thank my co-supervisors, Drs. David Milan and Saumya Das for their encouragement and insightful criticisms.

I am extremely grateful to Professor Nilesh Samani for his invaluable support in acting as my mentor at the University of Leicester. Professor Samani has always been available for help and advice and has been a constant source of motivation. I have been very fortunate to have one of the leaders in the field as a mentor.

I thank all the members of the Ellinor, Milan and Das labs, past and present. Without their support and assistance, it would not have been possible for me to accomplish my research goals. I am particularly indebted to Dr. Robert Mills for setting-up the optical mapping system, assisting me with the experiments, and performing optical mapping data analysis. I am also very grateful to Dr. Nathan Tucker for his help with analysis of ambulatory cardiac monitoring data.

I am grateful to Marisa Shea for recruiting patients for the pedigree-based exome sequencing studies and to Dr. Honghuang Lin for performing bioinformatic analysis of the exome sequencing data. Finally, I would like to thank Dr. Bridget Simonson for performing histological analysis of heart slides and the members of the surgical core lab at Beth Israel Deaconess Medical Center for implanting the ambulatory cardiac monitors.

Last but by no means least, a special thank you to my parents, Yashwantkaka, and to my sister for their untiring encouragement and to my wife, Rocio, for travelling halfway around the world and making so many sacrifices to support me.
I hereby declare that the work described in this thesis is original unless otherwise acknowledged in the text or noted by citations. None of the work has previously been accepted for a degree or diploma in this or any other University/Institution. All the work described in the thesis is my own, unaided except in the areas specified below.

1. Optical mapping experiments in $SK3^{T/T}$ mice – Dr. Robert Mills from the Milan laboratory at MGH initially built the Langendorff perfusion apparatus and set-up the optical mapping system. Further, during each experiment, Dr. Mills was responsible for setting-up Langendorff perfusion of the isolated mouse hearts. He also performed analysis of all optical mapping data. My role during these experiments was to explant the mouse hearts and to undertake electrophysiological stimulation protocols during optical mapping.

2. Telemetry recording of the cardiac rhythm in $SK3^{T/T}$ mice – Telemetry devices were implanted in $SK3^{T/T}$ mice by members of the Small Animal Surgical Core Facility at Beth Israel Deaconess Medical Center. My role in these experiments was initially to set-up the monitoring system at the MGH laboratory and subsequently to perform analysis of telemetry data. During later stages of the experiment, Dr. Nathan Tucker from the Ellinor laboratory performed analysis of the telemetry data.

3. Histological analysis of $SK3^{T/T}$ mouse hearts – All histological experiments (Hematoxylin and eosin staining, Masson's trichrome staining and staining for connexin-43) were performed by Dr. Bridget Simmonson from the Das laboratory at Beth Israel Deaconess Medical Center.

4. Patient recruitment – Recruitment of pedigree AF-325 was performed in its entirety by Dr. Ellinor, Dr. Milan and Marisa Shea (Research Nurse). Recruitment of pedigree AF-435 was performed by myself, Dr. Ellinor and Marisa Shea. I was responsible for performing all echocardiograms, Dr. Ellinor interviewed patients while Marisa Shea performed ECGs and bled patients.

5. Exome sequencing in pedigrees AF-325 and AF-435 – Exome sequencing was performed commercially by Perkin Elmer DNA Sequencing and Analysis services. Bioinformatic analysis of exome sequencing data was performed initially by Perkin Elmer and subsequently by Dr. Honghuang Lin at the Framingham Heart Study.
AUTHOR’S STATEMENT OF ORIGINALITY

6. Linkage analysis in pedigree AF-325 – Genotyping of SNPs using the Affymetrix chip was performed by members of the genotyping core at the Broad Institute. Genotyping of microsatellite markers was performed by members of the MGH sequencing core facility.

Some of the work described in this thesis has been presented at a scientific meeting and has also been published in *Cardiovascular Research* and the *Journal of Cardiovascular Electrophysiology*. The full references are included below:

1. Mahida S, Mills RW, Macri VS, Lemoine MD, Cooper R, Milan DJ, Ellinor PT. Overexpression of KCNN3 Results In Sudden Cardiac Death. In: Heart Rhythm Society congress; 2011 May 9-12; Boston.


LIST OF CONTENTS

List of tables x
List of figures xi
List of abbreviations xiii

Chapter 1 – Introduction
1.1 Epidemiology of AF 3
1.2 Pathophysiology and mechanisms of AF 5
1.2.1 Normal cardiac physiology 5
1.2.1.1 The cardiac action potential 8
1.2.1.1.1 The non-pacemaker action potential 9
1.2.1.1.2 The pacemaker action potential 13
1.2.1.2 Relationship between the action potential and excitability 15
1.2.2 Basic mechanisms of cardiac arrhythmia 18
1.2.2.1 Abnormal automaticity 19
1.2.2.2 Triggered activity 19
1.2.2.3 Reentry 20
1.2.3 Basic mechanisms of AF 24
1.2.3.1 Pulmonary vein triggers and AF 25
1.2.3.2 The susceptible atrial substrate and re-entry during AF 26
1.2.3.2.1 Structural remodelling in AF 27
1.2.3.2.2 Electrical remodelling in AF 28
1.3 Genetics of AF 29
1.3.1 Structure of the genome 30
1.3.2 Mechanisms of gene expression 31
1.3.2.1 Regulation of gene expression by cis acting regulatory elements 32
1.3.2.2 Regulation of gene expression by trans acting regulatory elements 33
1.3.2.3 Regulation of gene expression through chromatin structure 35
1.3.3 Genetic variation 36
1.3.3.1 Single nucleotide variations 37
1.3.3.2 Structural variants 38
1.3.4 Genetics of human disease 41
1.3.4.1 Mendelian diseases 41
1.3.4.1.1 Mapping of disease genes in Mendelian pedigrees 42
1.3.4.2 Complex trait diseases 46
1.3.4.2.1 Candidate-gene association studies 48
1.3.4.2.2 Genome wide association studies 49
1.3.5 Monogenic mutations in AF 52
1.3.5.1 Ion channel mutations 53
1.3.5.1.1 Potassium channel mutations 55
1.3.5.1.1.1 $I_{ks}$ channel mutations 55
1.3.5.1.1.2 $I_{K1}$ channel mutations 57
1.3.5.1.1.3 $I_{kur}$ channel mutations 57
1.3.5.1.1.4 Potassium channel mutations in AF cohorts 58
1.3.5.1.2 Sodium channel mutations
1.3.5.2 Non-ion channel mutations
1.3.5.2.1 Nucleoporin gene (NUP155) mutation
1.3.5.2.2 Connexin-40 gene (GJAS) mutations
1.3.5.2.3 Atrial natriuretic peptide gene (NPPA) mutation
1.3.6 AF in the general population
1.3.6.1 Heritability of AF
1.3.6.1.1 Candidate gene studies in AF
1.3.6.1.2 Genome wide association studies in AF
1.3.6.1.2.1 The association between the 4q25 locus and AF
1.3.6.1.2.2 The association between the 16q22 locus and AF
1.3.6.1.2.3 The association between the 1q21 locus and AF
1.3.6.1.2.4 Six novel loci for AF
1.4 Hypothesis and Aims
1.4.1 Aim 1: Characterisation of the mechanistic link between the KCNN3 gene at 1q21 locus and AF
1.4.1.1 Calcium activated potassium channels
1.4.1.2 SK channel pharmacology
1.4.1.3 SK channel expression and function
1.4.1.3.1 SK channels in neural tissue
1.4.1.3.2 SK channels in the heart
1.4.1.3.3 Characterising the role of the SK3 channel in cardiac electrophysiology
1.4.2 Aim 2: Identify novel mutations for AF using a combination of linkage analysis and exome sequencing
1.4.2.1 Pedigree-based exome sequencing
1.4.2.2 Exome sequencing in AF pedigrees
1.4.3 Summary of hypothesis and aims

Chapter 2 – Methods
2.1 Characterisation of the cardiac electrophysiological phenotype of a mouse line with variable modulation of Kcnn3 expression
2.1.1 Generation of SK3<sup>T</sup> mice
2.1.2 Ear tagging
2.1.3 Mouse tail biopsies
2.1.4 Isolation of genomic DNA from mouse tails and genotyping
2.1.5 Administration of doxycycline
2.1.6 Euthanasia for tissue collection
2.1.7 Quantitative PCR
2.1.8 Histological staining and immunohistochemistry
2.1.8.1 Haematoxylin Eosin staining
2.1.8.2 Staining for fibrosis
2.1.8.3 Immunohistochemistry
2.1.9 Echocardiography
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1.10 <em>In vivo</em> electrophysiology studies</td>
<td>104</td>
</tr>
<tr>
<td>2.1.11 High resolution optical mapping</td>
<td>109</td>
</tr>
<tr>
<td>2.1.12 Implantable loop recording</td>
<td>114</td>
</tr>
<tr>
<td>2.1.13 Statistical analysis</td>
<td>116</td>
</tr>
<tr>
<td>2.2 Identify rare mutations in families with AF using pedigree-based</td>
<td>117</td>
</tr>
<tr>
<td>exome sequencing</td>
<td></td>
</tr>
<tr>
<td>2.2.1 Clinical evaluation</td>
<td>117</td>
</tr>
<tr>
<td>2.2.2 12-lead electrocardiograms</td>
<td>118</td>
</tr>
<tr>
<td>2.2.3 Echocardiography</td>
<td>118</td>
</tr>
<tr>
<td>2.2.4 DNA extraction</td>
<td>119</td>
</tr>
<tr>
<td>2.2.5 Candidate gene screening in pedigree AF-435</td>
<td>121</td>
</tr>
<tr>
<td>2.2.6 Exome sequencing</td>
<td>123</td>
</tr>
<tr>
<td>2.2.7 Target enrichment and sequencing</td>
<td>129</td>
</tr>
<tr>
<td>2.2.8 Analysis of exome sequencing data</td>
<td>132</td>
</tr>
<tr>
<td>2.2.9 Bioinformatic analysis</td>
<td>134</td>
</tr>
<tr>
<td>2.2.10 Confirmation of variants</td>
<td>137</td>
</tr>
<tr>
<td>2.2.11 Mutagenesis and cell culture</td>
<td>137</td>
</tr>
<tr>
<td>2.2.12 Functional analysis</td>
<td>141</td>
</tr>
<tr>
<td>2.2.12.1 Plasmids</td>
<td>141</td>
</tr>
<tr>
<td>2.2.12.2 Cell culture and transfection</td>
<td>142</td>
</tr>
<tr>
<td>2.2.12.3 Luciferase assays</td>
<td>143</td>
</tr>
<tr>
<td>2.2.12.4 Statistical analysis</td>
<td>144</td>
</tr>
<tr>
<td>2.2.13 Linkage analysis in pedigree AF-325</td>
<td>145</td>
</tr>
<tr>
<td>2.2.13.1 The logarithm of odds (LOD) score</td>
<td>147</td>
</tr>
<tr>
<td>2.2.13.2 Two-point versus multipoint linkage analysis</td>
<td>148</td>
</tr>
<tr>
<td>2.2.13.3 Computation of LOD scores</td>
<td>149</td>
</tr>
<tr>
<td>2.2.13.4 SNP genotyping</td>
<td>150</td>
</tr>
<tr>
<td>2.2.13.5 Microsatellite marker genotyping</td>
<td>150</td>
</tr>
<tr>
<td>2.2.13.6 Computation of LOD scores</td>
<td>151</td>
</tr>
<tr>
<td>2.2.14 Analysis of candidate variants in family AF-325</td>
<td>152</td>
</tr>
</tbody>
</table>

Chapter 3 – The role of the SK3 channel in cardiac electrophysiology

3.1 Introduction                                                      154
3.2 Results                                                          156
3.2.1 Generation of *SK3*<sup>T</sup> mutant mice                   156
3.2.2 Sudden death in *SK3*<sup>T/T</sup> mice                       160
3.2.3 Ambulatory monitoring in *SK3*<sup>T</sup> mice                162
3.2.4 Ventricular conduction velocity and action potential duration   | 170 |
| *SK3*<sup>T/T</sup> mice                                           |      |
| 3.2.5 Connexin-43 expression in *SK3*<sup>T/T</sup> mice            | 172  |
| 3.2.6 Programmed stimulation in *SK3*<sup>T/T</sup> and *SK3*<sup>+/T</sup> mice | 173 |
3.3 Discussion                                                       175
3.3.1 Summary of main findings                                       | 175  |
| 3.3.2 *SK3*<sup>T/T</sup> mice display marked ventricular overexpression of *Kcnn3* | 176 |
List of Contents

3.3.3 $SK_3^{T/T}$ mice have a high incidence of sudden death 177
3.3.4 $SK_3^{T/T}$ mice have slowed ventricular conduction 180
3.3.5 $SK_3^{T/T}$ mice have increased repolarisation heterogeneity 181
3.3.6 $SK_3^{T/T}$ mice have an increased susceptibility to AF and atrial arrhythmias 182
3.3.7 Limitations 185
3.3.7.1 Mouse model of AF 185
3.3.8 Next steps 186
3.3.8.1 Measurement of APD and $I_{KCa}$ current in $SK_3^{T/T}$ mice 186
3.3.8.2 Characterisation of the mechanistic link between SNPs at the 1q21 locus and $KCNN3$ expression 187

Chapter 4 – Identification of a novel locus for AF
4.1 Introduction 189
4.2 Results 191
4.2.1 Clinical characteristics 191
4.2.2 Linkage analysis 196
4.2.3 SNP linkage analysis 197
4.2.4 Microsatellite linkage analysis 201
4.2.5 Exome sequencing 204
4.3 Discussion 206
4.3.1 Summary of main findings 206
4.3.2 Potential explanations for failure to identify causative mutation in AF-325 207
4.3.2.1 A non-coding mutation as a potential cause of AF 207
4.3.2.2 Potential reasons for failure to identify a causative coding mutation 212
4.3.2.2.1 Incorrect assumptions during analysis of exome sequencing data 212
4.3.2.2.2 Incorrect assumptions during linkage analysis 213
4.3.2.2.3 Technical limitations of exome sequencing 215
4.3.3 Limitations 216
4.3.3.1 Mis-phenotyping in pedigree-based studies 216
4.3.3.2 Large genetic interval in pedigree AF-325 217
4.3.4 Next steps 218
4.3.4.1 Genotyping for potential non-coding mutations 218
4.3.4.2 Screening for potential structural variants at the chromosome 1 locus 221

Chapter 5 – Identification of a novel mutation for AF
5.1 Introduction 223
5.2 Results 225
5.2.1 Clinical characteristics of pedigree AF-435 225
5.2.2 Mutation detection 229
5.2.2.1 Candidate gene screening 229
5.2.2.2 Exome sequencing 230
## List of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2.3 Segregation analysis of the GATA6 R585L mutation</td>
<td>235</td>
</tr>
<tr>
<td>5.2.4 Functional analysis of the GATA6 R585L mutation</td>
<td>238</td>
</tr>
<tr>
<td>5.3 Discussion</td>
<td>240</td>
</tr>
<tr>
<td>5.3.1 Summary of main findings</td>
<td>240</td>
</tr>
<tr>
<td>5.3.2 Cardiac transcription factors</td>
<td>241</td>
</tr>
<tr>
<td>5.3.2.1 The GATA family of transcription factors</td>
<td>241</td>
</tr>
<tr>
<td>5.3.3 Cardiac development and congenital heart disease</td>
<td>246</td>
</tr>
<tr>
<td>5.3.3.1 Transcription factor mutations and congenital heart disease</td>
<td>248</td>
</tr>
<tr>
<td>5.3.3.2 Transcription factor variants and AF</td>
<td>250</td>
</tr>
<tr>
<td>5.3.4 R585L expands the spectrum of GATA6 mutations causing AF and</td>
<td>251</td>
</tr>
<tr>
<td>structural cardiac defects</td>
<td></td>
</tr>
<tr>
<td>5.3.5 Gain-of-function mutation of GATA6 causes AF and structural</td>
<td>253</td>
</tr>
<tr>
<td>cardiac defects</td>
<td></td>
</tr>
<tr>
<td>5.3.6 Mechanistic link between the R585L GATA6 mutation and AF</td>
<td>255</td>
</tr>
<tr>
<td>5.3.7 Limitations</td>
<td>257</td>
</tr>
<tr>
<td>5.3.7.1 Complex discordant phenotypes</td>
<td>257</td>
</tr>
<tr>
<td>5.3.7.2 Incomplete segregation analysis</td>
<td>258</td>
</tr>
<tr>
<td>5.3.8 Next steps</td>
<td>259</td>
</tr>
<tr>
<td>5.3.8.1 Screening AF cohorts for GATA6 variants</td>
<td>259</td>
</tr>
<tr>
<td>5.3.8.2 Analysis of interaction of mutant GATA6 with other transcription factors</td>
<td>259</td>
</tr>
</tbody>
</table>

**Chapter 6 – Discussion**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Moving from association to mechanism in the post-GWAS era</td>
<td>262</td>
</tr>
<tr>
<td>6.2 The question of missing heritability</td>
<td>267</td>
</tr>
<tr>
<td>6.3 Clinical relevance and future therapeutic perspectives</td>
<td>270</td>
</tr>
<tr>
<td>6.3.1 Novel therapeutic targets for AF</td>
<td>270</td>
</tr>
<tr>
<td>6.3.2 Risk prediction of AF</td>
<td>272</td>
</tr>
<tr>
<td>6.3.3 Risk prediction of congenital heart disease</td>
<td>273</td>
</tr>
<tr>
<td>6.4 Conclusions</td>
<td>273</td>
</tr>
</tbody>
</table>

**Appendix**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix</td>
<td>276</td>
</tr>
<tr>
<td>Bibliography</td>
<td>283</td>
</tr>
</tbody>
</table>
LIST OF TABLES

| Table 1: | Summary of monogenic mutations associated with AF | 54 |
| Table 2: | Summary of results from association studies in AF cohorts | 66 |
| Table 3: | Summary of susceptibility loci identified by GWAS | 74 |
| Table 4: | Nomenclature of calcium activated potassium channels | 77 |
| Table 5: | Ratios from SK3+/T x SK3+/T cross | 157 |
| Table 6: | Echocardiographic measurements in 3-month-old mice | 160 |
| Table 7: | Baseline conduction intervals in 3-month-old mice | 169 |
| Table 8: | Baseline Intracardiac Conduction Intervals in 1-month-old mice | 174 |
| Table 9: | Baseline Intracardiac Conduction Intervals in 3-month-old mice | 174 |
| Table 10: | Clinical characteristics of AF cases in family AF-325 | 196 |
| Table 11: | Two-point linkage analysis using 7 microsatellite markers | 202 |
| Table 12: | Multipoint linkage analysis using 4 microsatellite markers | 203 |
| Table 13: | Candidate missense mutations at 1p21.1-q22 locus | 205 |
| Table 14: | Transmission of variants at 1p21.1-q22 locus in individual IV-1 | 206 |
| Table 15: | Variants identified in NKX-2.5 and TBX20 in family AF-435 | 230 |
| Table 16: | Number of paired-end reads aligned to the reference genome | 231 |
| Table 17: | Exome sequencing coverage data | 231 |
| Table 18: | Number of candidate variants indentified by exome sequencing in AF-435 | 232 |
| Table 19: | List of potentially causative genetic mutations in family AF-435 | 233 |
| Table 20: | Clinical characteristics of family AF-435 | 237 |
| Table 21: | Cardiac proteins regulated by GATA transcription factors | 245 |
| Appendix Table A.1: | Kcnn3 genotyping primer sequences | 277 |
| Appendix Table A.2: | Kcnn3 genotyping PCR reaction components | 277 |
| Appendix Table A.3: | Kcnn3 genotyping PCR cycling protocol | 277 |
| Appendix Table A.4: | First step of cDNA synthesis reaction mixture | 278 |
| Appendix Table A.5: | Second step cDNA synthesis reaction mixture | 278 |
| Appendix Table A.6: | RT-PCR - Thermal cycling parameters cDNA synthesis | 278 |
| Appendix Table A.7: | RT-PCR - Primer sequences | 278 |
| Appendix Table A.8: | RT-PCR - Two-step mastermix | 279 |
| Appendix Table A.9: | RT-PCR - Thermal Cycling Parameters | 279 |
| Appendix Table A.10: | Primer sequences for transcription factor candidate gene screening | 280 |
| Appendix Table A.11: | PCR cycling protocol for candidate gene screening | 282 |
| Appendix Table A.12: | Primers for confirmation of GATA6 R585L variant | 282 |
| Appendix Table A.13: | Comparison of wild type and mutagenic primers for site directed mutagenesis | 282 |
| Appendix Table A.14: | Reaction mixture for GATA6 mutant strand synthesis | 282 |
| Appendix Table A.15: | GATA6 mutant strand synthesis thermal cycling | 283 |
| Appendix Table A.16: | Sequencing primers for GATA6 clone | 283 |
| Appendix Table A.17: | Primers for confirmation of mutations identified by exome sequencing AF-325 | 283 |
| Figure 1: | Surface electrocardiogram (lead I) illustrating AF | 2 |
| Figure 2: | The cardiac conduction system | 6 |
| Figure 3: | Surface electrogram during sinus rhythm | 7 |
| Figure 4: | Representative atrial and ventricular action potentials | 12 |
| Figure 5: | Currents and channels involved in the generation of non-pacemaker and pacemaker action potentials | 14 |
| Figure 6: | Absolute and relative refractory periods | 16 |
| Figure 7: | Basic mechanisms of arrhythmia | 22 |
| Figure 8: | Conceptual model of reentry and implications for AF | 23 |
| Figure 9: | Classes of known regulatory elements for gene expression | 34 |
| Figure 10: | A heterozygous cross illustrating calculation of recombination fraction | 44 |
| Figure 11: | Linkage and linkage disequilibrium | 51 |
| Figure 12: | Pictorial image of genes implicated in Mendelian forms of AF | 53 |
| Figure 13: | Manhattan plot of meta-analysis results for GWAS of lone AF | 70 |
| Figure 14: | Manhattan plot of meta-analysis results for GWAS of AF | 74 |
| Figure 15: | Representative illustration protein products of genes implicated in GWAS for AF | 75 |
| Figure 16: | Schematic representation of BK and SK channels | 79 |
| Figure 17: | Neuronal action potential illustrating afterhyperpolarisation following an action potential | 81 |
| Figure 18: | Action potential recordings from SK2 knockout mice | 83 |
| Figure 19: | The “missing heritability” of complex traits | 85 |
| Figure 20: | Illustration of the SK3 regulatory cassette | 92 |
| Figure 21: | Photographs demonstrating the electrophysiology catheter and surgical field | 106 |
| Figure 22: | Schematic representation of perfusion system for isolated mouse hearts using a Langendorff apparatus | 110 |
| Figure 23: | Components of an optical mapping setup | 112 |
| Figure 24: | Schematic representation of implantation of telemetry device | 115 |
| Figure 25: | Schematic representation of data acquisition system telemetry devices | 116 |
| Figure 26: | One cycle dye terminator sequencing | 123 |
| Figure 27: | Pictorial representation of liquid phase exome capture | 125 |
| Figure 28: | Illumina HiSeq 2000 sequencing steps | 127 |
| Figure 29: | SureSelect target enrichment system capture process | 131 |
| Figure 30: | Pathway for processing of exome sequencing data | 134 |
| Figure 31: | A stepwise approach to filtration and prioritisation of variants identified by exome sequencing | 136 |
| Figure 32: | pCMV6-XL6 vector design | 138 |
LIST OF FIGURES

Figure 33: Overview of QuikChange site-directed mutagenesis method 140
Figure 34: Schematic representation of a luciferase reporter assay 144
Figure 35: Relative expression levels of KCNN3 in SK3\textsuperscript{T} mice as measured by RT-PCR 158
Figure 36: Histological analysis of WT and SK3\textsuperscript{T/T} mouse hearts 159
Figure 37: Kaplan-Meier survival plot at 90 days 161
Figure 38: Representative lead II electrogram recordings from SK3\textsuperscript{T/T} mice at the time of sudden death 165
Figure 39: Spontaneous atrioventricular dissociation in SK3\textsuperscript{T/T} mice 166
Figure 40: Plot graphs of raw data from continuous heart rate monitoring in SK3\textsuperscript{T/T} and WT mice 167
Figure 41: PR interval in conscious SK3\textsuperscript{T/T} and WT mice 168
Figure 42: RR intervals in conscious SK3\textsuperscript{T/T} and WT mice 169
Figure 43: High resolution optical mapping in SK3\textsuperscript{T/T}, SK3\textsuperscript{+/T} and WT mice 171
Figure 44: Immunostaining of SK3\textsuperscript{T/T} and WT mouse hearts for connexin-43 172
Figure 45: Programmed electrical stimulation in SK3\textsuperscript{T/T}, SK3\textsuperscript{+/T} and WT mice 175
Figure 46: Pedigree of family AF-325 195
Figure 47: Pedigree of family AF-325 demonstrating model for linkage analysis 198
Figure 48A: LOD score graph of two-point linkage analysis on chromosome 1 199
Figure 48B: Graphs of two-point LOD score for each autosome from a genome-wide scan 200
Figure 49: LOD score graph from multipoint linkage analysis locus on chromosome 1 201
Figure 50: LOD score graph from multipoint analysis using 4 microsatellite markers 203
Figure 51: Pedigree of family AF-435. 228
Figure 52: Schematic representation of GATA4 and GATA6 proteins. 234
Figure 53: Multi-species alignment of GATA6 amino acid sequence 235
Figure 54: Electrophoretogram of GATA6 R585L mutation. 236
Figure 55: Luciferase reporter assays of mutant GATA6 239
Figure 56: Pictorial representation of the structural organisation of GATA4, GATA5 and GATA6 243
Figure 57: Normal cardiac development 247
Figure 58: Origin and genetic mutations underlying congenital heart diseases 248
LIST OF ABBREVIATIONS

αMHC – gene encoding the α-myosin heavy chain
αMHC promoter-luc - α-myosin heavy chain promoter luciferase reporter construct
AERP - atrial effective refractory period
AF - atrial fibrillation
ANP - atrial natriuretic peptide
APD – action potential duration
APD_{80} - action potential duration at 80% repolarisation
ASD – atrial septal defect
AV node - atrioventricular node
AVERP - AV node effective refractory period
AVWB – atrioventricular Wenckebach cycle length
BK channel – large conductance calcium activated potassium channel.
BSA – bovine serum albumin
BNP – gene encoding the brain natriuretic peptide
BNP promoter-luc - brain natriuretic peptide promoter luciferase reporter construct
BWA - Burrows-Wheeler Aligner
C9orf3 - chromosome 9 open reading frame 3
C1orf62 - chromosome 1 open reading frame 62
CACNA1C – gene encoding the Cav1.2 channel
CAV1 – gene encoding caveolin-1, a transmembrane scaffolding protein
Cav1.2 - voltage-gated L-type calcium channel
cDNA – complementary DNA
CELF3 – gene coding CUGBP, Elav-like family member 3
CGH - microarray-based comparative genome hybridisation array.
cM - centimorgan
CMV - cytomegalovirus
CNV – copy number variant
cSNRP – corrected sinus node recovery period
C_{T} – cycle threshold
CV – conduction velocity
DNA - deoxyribonucleic acid
ECG - cardiac electrocardiogram
EF – ejection fraction
ERP - effective refractory period
FBP1 and FBP2 – Fat body protein 1 and 2 genes
FLG – gene encoding filaggrin
FS – fractional shortening
FBS - fetal bovine serum
GATK - Genome Analysis Toolkit
GJA5 – gene that encodes connexin-40
GNB3 – gene encoding the β3-subunit of heterotrimeric G protein
GWAS – genome-wide association study
HCN4 channel – predominant channel subtype underlying the I_{f} current
HEK293 - Human Embryonic Kidney 293 cells
HERG channel – voltage-gated potassium channel that underlies the $I_{Kr}$ current.
HGNC - HUGO Gene Nomenclature Committee
$I_{CaL}$ and $I_{CaT}$ - inward calcium currents that underlie phase 2, or the plateau phase of the action potential
ICD - Implantable cardiac defibrillator
$I_f$ - hyperpolarisation activated funny current
ICD – implantable cardiac defibrillator
IHD – ischaemic heart disease
IK channel – intermediate conductance calcium activated potassium channel.
$I_{K(ACh)}$ - repolarising potassium current activated in response to vagal stimulation
$I_{K1}$ - inward rectifier repolarising potassium current.
$I_{KATP}$ – repolarising potassium current that is activated during conditions of metabolic stress
$I_{KCa}$ - calcium-activated potassium current
$I_{KCa1}$ – intermediate conductance calcium-activated potassium channel
$I_{Kr}$ - delayed rectifier repolarising potassium current
$I_{Ks}$ - delayed rectifier repolarising potassium
$I_{Kur}$ - ultrarapid delayed rectifier potassium current that plays a prominent role in atrial repolarisation.
$I_{Na}$ - depolarising inward sodium current
$I_{NCX}$ - the $Na^+\text{-}Ca^{2+}$ exchanger current
Indel – insertion-deletion variant
$I_{to}$ - transient repolarising potassium current
IUPHAR - International Union of Pharmacology
IVDd – interventricular septum in diastole
Kb - kilobase
$K_{Ca1.1}$ - large conductance calcium activated potassium channel.
$K_{Ca3.1}$ - intermediate conductance calcium activated potassium channel.
$K_{Ca2.1}$ - small conductance calcium activated potassium channel subtype 1
$K_{Ca2.2}$ - small conductance calcium activated potassium channel subtype 2
$K_{Ca2.3}$ - small conductance calcium activated potassium channel subtype 3
$K_{Ca3.1}$ - intermediate conductance calcium activated potassium channel
$KCNA5$ encodes the Kv1.5 channel which underlies the ultrarapid delayed rectifier
$KCND2$ – gene encoding the Kv4.2 channel
$KCND3$ – gene encoding the Kv4.3 channel
$KCNE1-KCNE5$ – genes that encode the accessory β subunits for the HERG channel.
$KCNH2$ - gene that encodes the pore-forming α-subunit if the HERG channel
$KCNJ2$ – gene encoding the Kir2.1 channel
$KCNMA1$ – gene encoding large conductance calcium-activated potassium
$KCNN1$ - gene encoding small conductance calcium-activated potassium
$KCNN2$ - gene encoding small conductance calcium-activated potassium
$KCNN3$ - gene encoding small conductance calcium-activated potassium
$KCNN4$ - gene encoding intermediate conductance calcium-activated potassium
$KCNQ1$ - gene encoding pore-forming α-subunit of the KCNQ1 channel.
List of abbreviations

KCNQ1 channel - voltage-gated potassium channel that underlies the $I_{Ks}$ current.
kHz - kilohertz
Kir2.1 - non-voltage gated channel underlying the inward rectifier
Kv1.5 – voltage gated potassium channel that underlies the ultrarapid delayed rectifier ($I_{Kur}$) current.
Kv4.2 and Kv4.3 - potassium channels conducting the transient potassium current ($I_{K1}$)
LA – left atrium
LOD - logarithm of odds
LQTS - long QT syndrome
LV – left ventricle
LVH – left ventricular hypertrophy
LVIDd – left ventricular end diastolic dimension in diastole
LVIDs – left ventricular end diastolic dimension in diastole
LVPWd – left ventricular posterior wall diastole
MAF – minor allele frequency
mGATA6 - mutant GATA6 expression plasmid
MGH – Massachusetts General Hospital
mRNA – messenger RNA
miRNA – microRNA
MRPS21 – gene encoding 28S ribosomal protein S21, mitochondrial
MYOZ1 – gene encoding Myozenin-1 protein which localizes to the sarcomeric Z disc
Nav1.5 – cardiac voltage gated sodium channel α subunit
Navβ – cardiac voltage gated sodium channel β subunit
NKX-2.5 – gene encoding NK2 homeobox 5
NPC - nuclear pore complex
NPPA - gene encoding atrial natriuretic peptide
NPPA promoter-luc - natriuretic peptide precursor A promoter luciferase reporter construct
NUP155 – gene encoding nucleoporin, a component of the nuclear pore complexes
NPC – nuclear pore complex
OR – odds ratio
PBS - phosphate buffered saline
PCR – polymerase chain reaction
PITX2 – gene encoding the paired-like homeodomain transcription factor 2 potassium current ($I_{K1}$)
PMVK – gene encoding phosphomevalonate kinase
PPM – permanent pacemaker
PRRX1 – gene encoding paired mesoderm homebox protein 1
PRRX2 – gene encoding paired mesoderm homebox protein 2
PVC – premature ventricular complex
PVI – pulmonary vein isolation
PWD – posterior wall in diastole
RA – right atrium
RBBB – right bundle branch block
List of abbreviations

RNA – ribonucleic acid
RRP - relative refractory period
RT-PCR – reverse transcriptase polymerase chain reaction
RV – right ventricle
SA node - sino-atrial node
SAM - Sequence Alignment/Map
SEM – standard error of the mean
SCN1B – gene encoding the β1 subunit of the Nav1.5 channel.
SCN2B – gene encoding the β2-subunit of the Nav1.5 channel.
SCN3B – gene encoding the β3-subunit of the Nav1.5 channel.
SCN4B – gene encoding the β4-subunit of the Nav1.5 channel.
SCN5A – gene encoding the pore-forming α subunit of the Nav1.5 channel.
SK channel - small conductance calcium activated potassium channel
SK3T – mouse line designed to overexpress KCNN3 or to induce knockout of the gene
SK3+/T - heterozygous mouse with one SK3T allele and one wild type allele
SK3T/T - homozygous mouse with two SK3T alleles
SK3T/T+ DOX - homozygous mouse with two SK3T alleles treated with doxycycline to eliminate KCNN3 expression
SKCa1 - small conductance calcium activated potassium channel subtype 1
SKCa2 - small conductance calcium activated potassium channel subtype 2
SKCa3 - small conductance calcium activated potassium channel subtype 3
SNP - single nucleotide polymorphism
SNRP - sinus node recovery period
STR - short tandem repeats
SYNE2 – gene encoding nesprin-2, a nuclear membrane protein that anchors the nucleus to the cytoskeleton
SYNPO2L – gene encoding a cytoskeletal heart-enriched actin-associated protein
TBX5 – gene encoding the T-box transcription factor TBX5
TBX20 – gene encoding the T-box transcription factor TBX20
teto5 CMV - tet operator fused to a cytomegalovirus minimal promoter
tTA - tetracycline transactivator protein
VERP - ventricular effective refractory period
VSD – ventricular septal defect
VT – ventricular tachycardia
WT - control mouse with two wild type alleles
ZFHX3 - gene encoding the homeodomain zinc-finger transcription factor
Θ – theta, denotes recombination fraction
CHAPTER 1.
INTRODUCTION

Nearly a century ago Sir Thomas Lewis, a preeminent cardiologist of his time, made the following statement with reference to atrial fibrillation (AF)\(^1\):

*The history of the recognition of fibrillation of the auricles will impress you with the dimness of our eyes and the opacity of the obstacles which embarrass our vision.*

While significant progress has been made in understanding the pathophysiology of this common and morbid arrhythmia, important questions remain, and in many ways, Sir Lewis’ statement still rings true today.\(^2\)

AF is a condition characterised by chaotic electrical activity in the atria which results in loss of mechanical atrial function and irregularly irregular ventricular contractions. On the surface cardiac electrocardiogram (ECG), this disorganised atrial electrical activity manifests as rapid and irregular oscillations of the baseline, referred to as fibrillatory waves.\(^3\) Figure 1 illustrates a surface ECG of AF.

Typically, the disorganised electrical activity during AF results in atrial rates as high as 400-600 beats per minute. In order to protect the ventricles from these excessively high rates, the atrioventricular (AV) node acts as a filter and only conducts a proportion of atrial impulses. Thus, during AF the ventricular rate is determined by how effectively the AV node can filter the atrial impulses rather than the metabolic
demands of the body.\textsuperscript{4} Not uncommonly therefore, the heart rates are disproportionately high relative to metabolic demands of the body.

\textbf{Figure 1. Surface electrocardiogram (lead I) illustrating AF.} The top rhythm strip demonstrates AF with fibrillatory waves (red arrow). The bottom strip demonstrates normal sinus rhythm. The purple arrow points at a P wave, which is a reflection of organised atrial electrical activity during sinus rhythm. The organised electrical activity is lost during AF. From: Wikimedia commons. https://en.wikipedia.org/wiki/Atrial_fibrillation.\textsuperscript{5}

In addition to rapid heart rates, AF results in a loss of synchrony between the atria and the ventricles due to ineffective atrial contraction. As a result, there is a reduction in diastolic ventricular filling which in turn reduces cardiac output. The impairment of the cardiac output is compounded further by the irregularity of ventricular contraction, as the diastolic filling interval of the ventricles is constantly changing.\textsuperscript{6} Overall, the uncontrolled heart rates and loss of atrioventricular synchrony during AF is associated with a variety of symptoms and adverse physiological effects. Palpitations, which are defined as an awareness of an irregular or rapid heartbeat, are a common symptom of AF.\textsuperscript{7} Dyspnoea and impaired exercise tolerance are also prominent symptoms associated with the arrhythmia. It has been proposed that these symptoms arise due to abnormal cardiac haemodynamics, although the exact mechanisms remain unclear.\textsuperscript{7} Other symptoms associated with AF include chest
discomfort, presyncope and syncope. In a substantial proportion of patients, AF is asymptomatic and is discovered as an incidental finding. In AF patients who have prolonged periods of persistently elevated heart rates, left ventricular function may become impaired, a condition referred to as tachycardia mediated cardiomyopathy.\(^8\)

The pattern and natural history of AF is variable. In the majority of patients, the initial episodes of AF are self-terminating. As the disease progresses, the episodes are more prolonged and eventually the arrhythmia becomes sustained.\(^9\) In a proportion of patients however, the arrhythmia may be permanent from the initial episode whilst in others, AF may remain paroxysmal without evidence of progression. The pattern of AF commonly reflects the underlying arrhythmogenic substrate.

### 1.1 Epidemiology of AF

AF is by far the most prevalent cardiac rhythm abnormality and is a major cause of morbidity and mortality. The estimated global prevalence of AF is 596 per 100,000 in males and 359 per 100,000 amongst females.\(^{10}\) The prevalence is significantly higher in developed countries. With advancing age, there is a sharp rise in the prevalence of AF. For instance, amongst male patients in their 70’s, the prevalence of AF is five times higher than those in their 50’s.\(^{10}\) Not surprisingly therefore, it is estimated that as the population ages, the prevalence of AF will double over the next 50 years.\(^{11}\)
Chapter 1. Introduction

AF is associated with a significant increase in morbidity and mortality. Cerebrovascular complications associated with AF include stroke and dementia. The risk of stroke is four to five-fold higher amongst AF patients. AF is associated with more rapid cognitive decline and dementia. AF is also associated with a significantly increased risk of heart failure and exists as a comorbid disease in a large proportion of heart failure patients. Further, in patients with pre-existing heart failure, AF is a strong predictor of mortality. Of note, even in the absence of coexisting cardiovascular disease, AF is an independent predictor of mortality. Overall, AF represents a major burden to healthcare systems.

AF is associated with multiple risk factors. Pre-existing cardiac diseases such as congestive heart failure, ischaemic heart disease and valvular heart disease are important predictors for the development of AF. Between a quarter and a third of patients with AF have an underlying history of either coronary heart disease or valvular disease. Heart failure exists as a comorbid condition in a significant proportion of AF patients. Further, as discussed above, prolonged episodes of tachycardia during AF may cause heart failure. Therefore, AF and heart failure co-exist and both conditions promote one another. While valvular heart disease remains an important cause of AF, the incidence of valve-related AF has fallen in developed countries in parallel with the decline in the incidence of rheumatic valve disease.
Chapter 1. Introduction

In addition to the aforementioned structural cardiac diseases, risk factors for cardiovascular disease such as hypertension, diabetes and obesity are associated with an increased risk of AF. Between 60 and 80% of patients with AF have a pre-existing history of hypertension. Diabetes exists as a co-morbid condition in a fifth of AF patients. A number of non-cardiac conditions have also been reported to cause AF. Examples include pulmonary diseases causing hypoxia and infections.

In a proportion of cases, AF occurs in the absence of demonstrable risk factors, referred to as ‘lone AF’. Previous studies have reported anything from a 1.6 to 30% prevalence of lone AF. While age is not a diagnostic criterion for lone AF, the majority of patients with this subtype of the arrhythmia are under 60-years-old.

1.2 Pathophysiology and mechanisms of AF

In order to understand the pathophysiological mechanisms underlying AF, one must gain an appreciation of normal cardiac electrophysiology and the basic mechanisms of cardiac arrhythmia. The following section provides an overview of these basic concepts followed by a description of the mechanisms underlying AF.

1.2.1 Normal cardiac electrophysiology

Under physiological circumstances, the mechanical function of the heart is finely tuned to meet the metabolic demands of the body. Normal functioning of the heart is critically dependent upon coordinated electrical activity.
During the normal cardiac rhythm, referred to as sinus rhythm, electrical activity originates in the sino-atrial node (SA node) which is the natural pacemaker of the heart. In mammals, the SA node is located in the roof of the right atrium at the junction between the atrium and the superior vena cava (Figure 2). At rest, the SA node spontaneously fires at approximately 60 beats per minute while during peak exercise, the heart rate could rise to 200 beats per minute. Regulation of the heart rate is achieved through modulation of sympathetic and parasympathetic inputs into the SA node.

![Figure 2. The cardiac conduction system viewed though a longitudinal section of the heart. Abbreviations; SA node, sinoatrial node; AV node, atrioventricular node. From: Lab 8: Vertebrate Circulation and Respiration. http://openware.org/wiki/Lab_8:_Vertebrate_Circulation_and_Respiration.](image)

The electrical impulses from the SA node are conducted through the atria to the atrioventricular node (AV node). Whilst there is no specialised conduction tissue between the SA node and the AV node, the orientation of atrial muscle fibres facilitates propagation of electrical activity in an anatomically and temporally
organised fashion. The result is coordinated atrial contraction. On the surface electrogram, activation of the atria manifests as a P wave (Figure 3).

**Figure 3.** Schematic diagram of the surface electrogram during sinus rhythm. From: Nerbonne et al. Physiol Rev 2005;85(4):1205-1253.

The AV node is a highly specialised part of the cardiac conduction system which is located at the atrioventricular junction (Figure 2). The AV node connects the atria and the ventricles and fulfils three main functions; 1) delaying conduction in a rate-dependent manner thereby allowing sufficient time for optimal ventricular filling, 2) protecting the ventricles from rapid atrial rates during atrial arrhythmias, and 3) acting as an escape pacemaker during episodes of suppressed SA node activity or supraventricular conduction block. The AV node is also innervated by sympathetic and parasympathetic inputs which influence the degree of conduction delay between the atria and the ventricles. On the surface ECG, the conduction delay in the AV node is represented by the PR interval (Figure 3).

Following an appropriate delay at the AV node, the cardiac impulses are conducted to the ventricles via the specialised ventricular conduction system which consists of the bundle of His, the bundle branches, the fascicles, and the Purkinje fibres. The
specialised conduction fibres run in the interventricular septum and subsequently spread to both ventricles. The coordinated propagation of electrical activity in the ventricles results in simultaneous ventricular contraction. On the surface ECG, the spread of electrical activity through the ventricles is represented by the QRS complex while recovery of ventricular excitability is represented by the T wave (Figure 3). The terminal segment of the ECG is discussed in more detail in Section 1.2.1.2.

### 1.2.1.1 The cardiac action potential

At a cellular level, the propagation of electrical activity in the myocardium is dependent upon the generation of action potentials. The cardiac action potential is a representation of the myocardial transmembrane potential as a function of time. During sinus rhythm, action potentials originate from specialised cells within the SA node which possess the ability to depolarise spontaneously. These action potentials are subsequently conducted from cell-to-cell through intracellular gap junctions resulting in waves of excitation that activate different regions of the heart in a synchronised fashion.

At a molecular level, generation of the action potential involves a complex interplay between inward currents that depolarise the cell and outward currents that repolarise it. The specific characteristics of the action potential vary according to the cell type. Pacemaker cells, such as SA nodal cells, display spontaneous depolarisation with slower rates of depolarisation. Non-pacemaker cells on the other hand have action
Chapter 1. Introduction

potentials with faster rates of depolarisation. These cells do not display spontaneous depolarisation under normal circumstances. The specific characteristics of non-pacemaker and pacemaker action potentials are described below.

1.2.1.1  The non-pacemaker action potential

The non-pacemaker action potential is divided into 5 phases, designated phase 0-4. During the resting phase, which corresponds to diastole, the membrane potential is maintained at a stable negative potential. The negative potential is determined primarily by an inward rectifier potassium current ($I_{K1}$). Upon stimulation by a propagating impulse, the membrane potential reaches a specific threshold and subsequently rises rapidly (depolarisation). This rapid upstroke (phase 0) reflects an inward sodium current ($I_{Na}$) which is the result of activation of voltage-gated sodium (Nav1.5) channels. The Nav1.5 channel is encoded by the SCN5A gene. A number of ancillary beta subunits (Navβ1, Navβ2 and Navβ3), which modulate the function of Nav1.5, have also been identified.

The depolarisation phase is followed by an early and rapid repolarisation phase (phase 1). Phase 1 is characterised by an inactivation of the Nav channels and an activation of potassium channels conducting the transient potassium current ($I_{to}$). These channels are designated Kv4.2 and Kv4.3 and are encoded by the KCND2 and KCND3 genes respectively. The activation of the $I_{to}$ conducting channels is short-lived and results in a transient efflux of potassium ions.
Chapter 1. Introduction

Depolarisation of the membrane also activates voltage-gated L-type calcium channels (Cav1.2). Opening of Cav1.2 channels causes an influx of calcium ions resulting in phase 2, or the plateau phase of the action potential. The Cav1.2 channel is encoded by the CACNA1C gene. During the plateau phase, $I_{Na}$ and $I_{to}$ currents decrease and the inward calcium current ($I_{CaL}$) counteracts any repolarising forces. The influx of calcium from the cell membrane also results in calcium release from intracellular stores. These calcium fluxes are central to the mechanism of excitation-contraction coupling.

During the latter stages of the plateau phase, $I_{CaL}$ slowly declines as the voltage- and calcium-dependent Cav channels are inactivated. At the same time, a number of repolarising potassium currents are activated. The three predominant repolarising currents are $I_{Kr}$, $I_{Ks}$ and $I_{K1}$. The point at which the balance of currents shifts in favour of the repolarising forces marks the start of phase 3, or the repolarisation phase of the action potential. Phase 3 returns the membrane potential to baseline. $I_{Kr}$ and $I_{Ks}$ are referred to as delayed rectifier currents and are conducted by voltage-gated potassium channels. The $I_{Kr}$ current is conducted by the HERG channel, which is encoded for by the KCNH2 gene. Two accessory β subunits have been identified for this channel, designated KCNE1 and KCNE2. The $I_{Ks}$ current is conducted by a channel composed of the KCNQ1 α subunit and the KCNE1 β subunit. The $I_{K1}$ current, also known as the inward rectifier current, plays an important role in repolarisation as well.
as stabilisation of the resting membrane potential. \( \textit{l}_{\text{K1}} \) is conducted by the Kir2.1 channel, a non-voltage gated channel encoded for by the KCNJ2 gene.\(^{22}\)

The potassium currents outlined above are the major determinants of cardiac repolarisation. However, a range of other potassium currents are also recognised. Some of these currents have been demonstrated to contribute to repolarisation while the roles of others are less clearly defined. The \( \textit{l}_{\text{Kur}} \) current is a voltage-dependent current that is not present in ventricular cells but plays a prominent role in atrial repolarisation.\(^{22}\) The \( \textit{l}_{\text{KATP}} \) current is activated during conditions of metabolic stress and thus provides a link between metabolic stress and cardiac repolarisation.\(^{31}\) \( \textit{l}_{\text{K(Ach)}} \) channels are activated in response to vagal stimulation through a G protein coupled mechanism.\(^{32}\) These channels are also predominantly expressed in the atria and contribute to atrial repolarisation.\(^{30}\)

In recent years, another family of potassium channels, the calcium-activated potassium channels have emerged as potentially important regulators of cardiac repolarisation. It has been postulated that these channels provide a link between intracellular calcium concentration and membrane conductance to potassium ions.\(^{33}\) Based on potassium conductance, these channels are sub-classified into large (BK), intermediate (IK) and small-conductance (SK) channels. The BK channel, also known as \( \textit{K}_{\text{Ca}1.1} \), is encoded for by the KCNMA1 gene. The IK channel, also known as \( \textit{K}_{\text{Ca}3.1} \), is encoded for by the KCNN4 gene. The SK channels constitute three subtypes; SK1
Chapter 1. Introduction

(K\(_{\text{Ca}2.1}\)), SK2 (K\(_{\text{Ca}2.2}\)), and SK3 (K\(_{\text{Ca}2.3}\)). These channels are encoded for by the KCNN1, KCNN2 and KCNN3 genes respectively.\(^\text{34}\)

Overall, the non-pacemaker action potential is a summation of a range of depolarising and repolarising currents. The waveform of the action potential varies according to the region of the heart, an effect that is determined by the relative contribution of the repolarising currents. For instance, the atrial action potential is shorter than the ventricular action potential. There are also observed variations in the action potential between the ventricular epicardium and endocardium, albeit more subtle.\(^\text{26}\) Typical action potential waveforms in the atrium and ventricle are illustrated in Figure 4.

**Figure 4.** Representative atrial and ventricular (right) action potentials with the underlying currents contributing to each phase of the action potential. From: Ravens U, et al. Europace. 2008;10(10):1133-7.\(^\text{35}\)
1.2.1.1.2 The pacemaker action potential

In addition to the SA node, the AV node and His-Purkinje tissue harbour specialised pacemaker cells. However, as discussed in the previous section, the SA node is the sole pacemaker under normal physiological circumstances. In certain disease states, where either generation or conduction of electrical activity from the SA node is compromised, the AV node or His-Purkinje cells may act as ancillary pacemakers.

The unique property of pacemaker cells is their inherent ability to depolarise spontaneously. In contrast to non-pacemaker cells, pacemaker cells have a more positive phase 4 resting potential, which may be attributable to the lack of a background $I_{K1}$ current. As phase 4 progresses, pacemaker cells display further time-dependent depolarisation, referred to as the pacemaker potential. The funny current ($I_{f}$) is an important contributor to the pacemaker potential. This current is described as ‘funny’ based on the fact that it is activated upon membrane hyperpolarisation and it conducts a non-selective cationic current. In the mammalian heart, HCN4 is the predominant channel subtype underlying the $I_{f}$ current.

Upon reaching the threshold, the upstroke of the pacemaker action potential is significantly slower than that of the non-pacemaker action potential. The difference in the slope is explained by the fact that the upstroke is mediated by slow inward $I_{CaL}$ and $I_{Cat}$ currents and that these cells have minimal $I_{Na}$. Following the depolarisation phase, calcium ions are extruded from the pacemaker cells through a sodium-calcium exchanger.
Depolarisation of the membrane potential results in activation of the delayed rectifier potassium currents ($I_{Kr}$ and $I_{Ks}$) which return the cell to the resting membrane potential. In contrast to non-pacemaker cells, the repolarising potassium currents are not counteracted by an inward calcium current. Therefore there is no prolonged plateau phase. The delayed rectifier potassium currents deactivate gradually upon reaching the resting membrane potential. Coupled with an activation of the $I_{If}$ and $I_{Ca}$ currents, this deactivation commences the next phase 4 pacemaker potential. The differences between the pacemaker and non-pacemaker action potentials are highlighted in Figure 5.

**Figure 5.** Currents and channels involved in the generation of the non-pacemaker and pacemaker (right) action potentials. The various pumps and channels that contribute to the phases of the action potential are shown above and below. In these diagrams, IK refers to a combination of $I_{Kr}$ and $I_{Ks}$. From: Shih H, et al. Tex Heart Inst J. 1994;21:30-41.
1.2.1.2 Relationship between the cardiac action potential and excitability

As mentioned in Section 1.2.1.1.1, when a cardiomyocyte is stimulated by a propagating impulse, the activation of voltage-gated sodium channels (Nav1.5) results in the rapid upstroke (phase 0) of the action potential. Following phase 0, the Nav1.5 channels are rapidly inactivated and remain in an inactivated state until the cell repolarises to a transmembrane potential of approximately -60mV. During the inactivation phase of the Nav1.5 channels, an action potential cannot be evoked by a stimulus and the cell is said to be refractory. This period of inactivation is referred to as the absolute refractory period (ARP).

During the early phase of Nav1.5 channel recovery, some channels may recover faster than others. When sufficient numbers of Nav1.5 channels have recovered, the cell can be stimulated to generate an action potential. However, a larger amplitude stimulating impulse than normal is required to generate an action potential as not all sodium channels have recovered from inactivation. This period is referred to as the relative refractory period (RRP). The end of the RRP is marked by full recovery of the sodium channels which coincides with the late phase of the action potential. In summary therefore, the recovery of excitability of cardiomyocytes following depolarisation is dependent upon membrane repolarisation. The relationship between the ARP and RRP and the action potential are illustrated in Figure 6.
Chapter 1. Introduction

Figure 6. Absolute and relative refractory periods illustrated relative to the phases of the cardiac action potential.

The intimate relationship between the ARP and the action potential duration means that the two parameters are commonly used as estimates of one another. A number of different techniques can be used to measure the repolarisation time and the refractory periods including intracellular recordings in isolated cardiomyocytes, optical mapping, intracardiac electrophysiology techniques in intact human or animal hearts, and measurement of the QT interval on the surface ECG. These techniques often provide complementary information regarding the action potential and refractory periods. A more detailed description of the techniques is included below.
Chapter 1. Introduction

Intracellular recording in isolated cardiomyocytes is the traditional method of recording the action potential duration. The technique involves direct measurement of the action potential using microelectrodes. Optical mapping involves the use of fluorescent, voltage-sensitive dyes to measure the action potential duration on the epicardial surface of intact hearts. This technique plays an important role in bridging the gap between cellular electrophysiology and macroscopic activation of the myocardium.

Intracardiac electrophysiology studies involve the introduction of programmed extrastimuli during either a paced rhythm or sinus rhythm to measure refractory periods of cardiac tissue. The refractory periods are defined by introducing progressively more premature programmed extrastimuli and measuring the response of the tissue. The effective refractory period (ERP) is defined as the longest coupling interval between the drive train and the premature impulse that fails to conduct through the tissue of interest. The RRP on the other hand represents the longest premature coupling interval that results in delayed conduction in that tissue relative to baseline conduction. Refractory period measurements can be applied to atrial tissue, ventricular tissue and the atrioventricular conduction system.

The surface ECG is a commonly used tool for the evaluation of ventricular repolarisation as it represents a summation of action potentials in individual cardiomyocytes. Conventionally, the QT interval is measured from the beginning of
the QRS complex, which marks ventricular depolarisation, to the end of the T wave, which marks the end of repolarisation. Therefore, the QT interval is a measure of both ventricular depolarisation and ventricular repolarisation.\textsuperscript{44} The QT interval can thus be prolonged either due to a delay in depolarisation, which is also associated with a prolonged QRS duration, or a prolongation of the time to repolarisation, which manifests as an isolated prolongation of the ST interval.

### 1.2.2 Basic mechanisms of cardiac arrhythmia

As discussed in Section 1.2.1, normal sinus rhythm is characterised by coordinated electrical activation of the heart with heart rates that are optimised to meet the metabolic demands of the body. Cardiac arrhythmias on the other hand are abnormal cardiac rhythms that arise due to aberrant initiation and/or propagation of the action potential wavefront.\textsuperscript{45} Cardiac arrhythmias can be classified according to heart rate into tachyarrhythmias, which are characterised by abnormally rapid heart rates, and bradyarrhythmias, which are characterised by abnormally slow heart rates. The following section will focus the basic mechanisms of tachyarrhythmia.

All cardiac tachyarrhythmias, whether arising from the atrium, the ventricle, or specialised cardiac conduction tissue, share the same basic mechanisms.\textsuperscript{45} The three main mechanisms underlying cardiac tachyarrhythmias are abnormal automaticity, triggered activity and reentry.
1.2.2.1 Abnormal automaticity

In the normal heart, three structures display automaticity and an ability to drive the cardiac rhythm; the SA node, the AV node, and the His-Purkinje fibres. Abnormal automaticity refers to the ability of non-pacemaker cells to depolarise in the absence of external stimuli or when the aforementioned pacemaker cells display non-physiological depolarisation (Figure 7A). As discussed in Section 1.2.1.1.1, during the resting phase of the non-pacemaker action potential, the inward rectifier potassium current ($I_{K1}$) maintains a negative membrane potential, thus preventing the generation of spontaneous action potentials. Non-pacemaker cells with abnormal automaticity display a time-dependent depolarisation of the resting membrane potential which eventually reaches threshold resulting in the generation of an action potential. Increased automaticity maybe the result of an attenuated $I_{K1}$ or enhanced inward currents such as the $I_I$ current.

1.2.2.2 Triggered activity

Triggered rhythms are generated by afterdepolarisations, which can occur during the repolarisation phase (phase 2 or 3) or the resting phase (phase 4) of the action potential, termed early and late afterdepolarisations respectively (Figure 7B). Early afterdepolarisations commonly occur in the context of a prolonged action potential duration. During the repolarisation phase of the action potential, the L-type calcium current ($I_{CaL}$) is inactivated and calcium is removed from the cytoplasm by a combination of transmembrane Na$^+$-Ca$^{2+}$ exchange and uptake into the sarcoplasmic
Chapter 1. Introduction

reticulum. Prolongation of the action potential can result in a time-dependent reactivation of $I_{CaL}$ which can in turn reach the threshold potential resulting in an early afterdepolarisation.\textsuperscript{46}

Delayed afterdepolarisations also result from abnormal calcium homeostasis. During the systolic phase of the action potential, calcium is released from the sarcoplasmic reticulum. Under normal circumstances, ryanodine receptors in the sarcoplasmic reticulum release calcium during systole in response to transmembrane calcium influx. In cases where the sarcoplasmic reticulum is overloaded with calcium, or if ryanodine receptors have abnormal function, calcium release can occur from the sarcoplasmic reticulum during diastole. In response to the calcium release, membrane Na$^+$-Ca$^{2+}$ exchangers are activated resulting in a net influx of sodium relative to the efflux of calcium ions.\textsuperscript{47} The overall result is an inward movement of positive ions and delayed afterdepolarisations. In contrast to increased automaticity therefore, triggered activity, either due to early or late afterdepolarisations, is dependent on the preceding action potential.

1.2.2.3 Reentry

Reentry is the most common mechanism of cardiac arrhythmia. During reentry, the action potential wavefront circulates continuously around a zone of conduction block.\textsuperscript{46} The zone of conduction block may arise in structurally normal tissue with functional alterations in conduction properties or, more commonly, in structurally
abnormal tissue. Conceptually, the reentry circuit consists of two pathways surrounding the region of conduction block. In order to sustain reentry, the pathways must fulfil a number of criteria; 1) they must be connected proximally and distally, 2) they must possess different electrophysiological properties, and 3) they must be able to conduct an action potential both antegrade and retrograde.

As illustrated in Figure 7C, during initiation of reentry, a stimulus is blocked in one pathway of the reentry circuit which is refractory (zone 1) and conducts down the alternate pathway which has recovered excitability more rapidly (zone 2). This situation is commonly referred to as ‘unidirectional block’. As the impulse propagates through zone 2, zone 1 recovers excitability and the impulse is able to conduct retrogradely, thereby initiating reentry.

The maintenance of reentry is dependent upon the impulse propagating around the circuit slowly enough for all regions to regain excitability. In other words, the time taken to traverse the circuit, or the conduction time, should be longer than the longest refractory period within the circuit. The conduction time is determined by the length of the circuit and the conduction velocity. Therefore, slower conduction and a longer path length increases the likelihood of a propagating wavefront consistently having an excitable gap before it, because the tissue has had sufficient time to recover. Another important factor for the maintenance of reentry is the
refractory period of the tissue. The shorter the refractory period, the greater the speed of recovery and hence the higher the chance of maintaining reentry.

**Figure 7.** Basic mechanisms of arrhythmia

A. Normal ‘automaticity’ occurs when a spontaneously depolarising cell reaches threshold potential and fires (1). When the rate of depolarisation increases, abnormally rapid firing will occur (2).

B. Afterdepolarisations are depolarisations caused by excessively large inward currents carried by the Na⁺/Ca²⁺ exchanger (2). If afterdepolarisations are large enough to reach threshold, premature ectopic action potentials result (3) before the next expected normal action potential (1).

C. Reentry occurring between two tissue zones, I and II, which are connected as shown on the right. A premature activation (2) in zone II, which fails to initiate firing in zone I because zone I is still refractory, may conduct back (red dashed line) to zone I at a time when it can respond with an action potential (3). This action potential may propagate to initiate (4) in zone II, and the process can continue indefinitely. RP, refractory period; TP, threshold potential.

A useful concept for understanding reentry is the leading circle model. This model is based on the wavelength of the circulating impulse, which is a product of the conduction velocity and the refractory period of a circulating impulse. In effect, the wavelength is the distance travelled by a propagating impulse in a single refractory period. According to the leading circle model, in order to maintain reentry, the available pathlength of the reentry circuit must be longer than the wavelength of the circulating impulse. The leading circle model predicts that the shorter the wavelength, the greater the number of simultaneous reentry circuits that can be maintained for a given mass of atrial tissue (Figure 8). Therefore, factors that shorten the refractory period or reduce the conduction velocity would be predicted to promote AF.

**Figure 8.** Conceptual model of reentry and implications for AF. Role of wavelength (WL) in AF maintenance. A. In normal atria, the number of reentrant waves that can be accommodated is small, and reentry easily terminates. B. When wavelength is reduced, by decreasing the refractory period (RP) or conduction velocity, reentrant circuits are smaller and more can be accommodated; AF becomes unlikely to self-terminate. C, Drugs that increase wavelength reduce the number of circuits, favouring AF termination. From: Iwasaki YK et al. Circulation. 2011;124(20):2264-74.
Chapter 1. Introduction

At a cellular level, the conduction velocity and refractory period, and hence the probability of maintaining reentry, is dependent upon the balance between the depolarising and repolarising currents that constitute the cardiac action potential. As discussed in Section 1.2.1.2, the refractory period is determined by the period of inactivation of Nav1.5 channels. Following phase 0 of the action potential, Nav1.5 channels cannot be activated again until the membrane is close to full repolarisation. Therefore, membrane repolarisation is a major determinant of the refractory period.

Conduction velocity on the other hand is influenced by the magnitude of the $I_{Na}$ current and the rate of rise of the action potential upstroke. Effectively therefore, alterations in both the depolarising $I_{Na}$ current and repolarising potassium currents can influence susceptibility to AF by altering the conduction velocity and/or the refractory period.

1.2.3 Basic mechanisms of AF

All three of the mechanisms of arrhythmia discussed in the previous section are thought to contribute to the initiation and maintenance of AF. The contemporary paradigm for AF acknowledges the contribution of focal triggers as well as a susceptible atrial substrate, which may promote reentry. In the majority of AF cases, the focal triggers and susceptible atrial substrate are predicted to interact. These concepts are discussed in more detail below.
1.2.3.1 Pulmonary vein triggers and AF

The most frequent site of origin of the focal triggers that provoke and potentially drive AF is the pulmonary veins. Four pulmonary veins return oxygenated blood from the pulmonary circulation to the left atrium. The ectopic electrical impulses from the pulmonary veins are conducted to the left atrial tissue through surrounding muscular sleeves. In recent years, the importance of pulmonary vein triggers in the pathogenesis of AF has been underscored by the observation that electrical isolation of these veins by radiofrequency catheter ablation successfully terminates AF.

The basic mechanisms underlying pulmonary vein triggers have yet to be fully characterised. Based on cellular electrophysiology studies, it has been proposed that reentry, enhanced automaticity and triggered activity could all potentially contribute to abnormal impulse generation. In comparison to atrial myocytes, pulmonary venous cells have a shorter action potential duration, a slower upstroke velocity and a more depolarised resting membrane potential, factors that are predicted to promote micro-reentry in the myocardial sleeves. Isolated pulmonary venous myocytes also display increased automaticity and early and late afterdepolarisations. Further evidence to support the contribution of abnormal automaticity comes from the identification of specialised node-like cells in the pulmonary veins.
1.2.3.2 The susceptible atrial substrate and reentry during AF

The mechanisms of reentry during AF are complex and are not clearly understood. Over the years, a number of different theories have emerged regarding the spatiotemporal organisation of reentry during AF. In 1921, Lewis et al proposed that the maintenance of AF is dependent upon a single reentry circuit within the atrium.\textsuperscript{58} This theory later evolved into the ‘mother rotor’ hypothesis. A competing theory was proposed by Moe et al in 1964.\textsuperscript{59} They proposed that multiple high-frequency reentrant circuits, which propagate randomly within the atrium, are responsible for sustained AF. According to this model, labelled the ‘multiple wavelet hypothesis’, AF is a self-sustaining arrhythmia which is not dependent upon initiating triggers. Supporting evidence for both these hypotheses has emerged from studies in animal models and it is likely that different patterns of reentry predominate in different clinical situations.\textsuperscript{60}

Irrespective of the spatiotemporal organisation of AF, the stabilisation of atrial reentry circuits, and hence the persistence of AF, is dependent upon a susceptible atrial substrate with altered conduction velocity and refractory periods. The development of a susceptible atrial substrate is dependent upon atrial remodelling, a process which results in a persistent change in the structure and/or function of the atrium.\textsuperscript{48} Remodelling is the result of a number of adaptive mechanisms in atrial myocytes designed to maintain homeostasis in the face of external stressors. Prolonged tachycardia associated with AF is one of the most common ‘stressors’.\textsuperscript{61} Other
important causes of atrial remodelling include neurohormonal upregulation and
myocyte stretch, which maybe the result of conditions such as hypertension, valvular
disease, and heart failure.\textsuperscript{61}

Atrial remodelling can result in alterations in the cellular/extracellular matrix, referred
to as structural remodelling, and/or ion channel expression, referred to as
electrophysiological remodelling.\textsuperscript{61} The predominant type of remodelling commonly
depends on the external stressor. For instance, electrical remodelling is more
prominent in the context of atrial tachycardia while structural remodelling
predominates in patients with congestive heart failure. Structural and electrical
remodelling is discussed in more detail below.

\subsection{Structural remodelling in AF}

Structural remodelling of the atrium occurs with age and is accelerated by cardiac
conditions such as congestive cardiac failure, hypertension and mitral valve disease.
At a macroscopic level, structural remodelling is characterised by atrial dilatation and
impaired atrial contractile function.\textsuperscript{61} At a cellular level, structural remodelling is
associated with multiple changes including progressive interstitial fibrosis,
hypertrophy, dedifferentiation and apoptosis of atrial myocytes.\textsuperscript{61} A number of
studies in animal models have implicated atrial fibrosis as an important contributor to
the development of a susceptible atrial substrate.\textsuperscript{48} The deposition of interstitial
collagen leads to non-uniform anisotropy, associated with disruption of cell-cell
communication and electrical uncoupling. In addition, fibrosis is associated with a reduction in conduction velocity and areas of conduction block. The overall effect is a reduction in the wavelength and unidirectional block, both conditions that promote reentry circuits.

### 1.2.3.2.2 Electrical remodelling in AF

Electrical remodelling is the predominant form of remodelling associated with prolonged atrial tachycardia and AF. Over the years, a number of elegant studies in animal models have demonstrated the concept of electrical remodelling in response to AF or atrial tachycardia pacing. In these studies, prolonged rapid atrial rates were associated with a shortening of the atrial ERP and action potential duration. As discussed in the previous section, an abbreviated ERP is predicted to promote reentry by shortening the wavelength, which in effect allows for a greater number of reentry wavelets within the atrium. Therefore, an important consequence of atrial remodelling is a progressively increased susceptibility to AF. Based on these observations, Allessie *et al* coined the phrase ‘*Atrial Fibrillation Begets Atrial Fibrillation*’.

The changes in atrial ERP and action potential duration in response to AF or atrial tachycardia are thought to be the consequence of alterations in ion channel expression. One of the most prominent changes is a marked downregulation in the expression of L-type calcium channels, which underlie the $I_{\text{Cal}}$ current. Another
important change is an increase in inward-rectifier potassium currents, including \( I_{K1} \) and \( I_{K(Ach)} \). Interestingly however, the transient outward potassium current (\( I_{to} \)) is attenuated in tachycardia-mediated electrophysiological remodelling. The physiological relevance of this observation is unclear.

As mentioned in the previous section, structural remodelling is the predominant form of remodelling in the context of heart failure. However, it is important to note that congestive heart failure is also associated with alterations in a number of ionic currents. In contrast to atrial tachycardia mediated remodelling, heart failure induced electrical remodelling is not associated with an abbreviated action potential duration or ERP.\(^{61}\) In addition to an attenuated \( I_{Ca} \) current, heart failure is associated with a downregulation of the repolarising potassium currents \( I_{to} \) and \( I_{Ks} \). On the other hand, the \( Na^+ - Ca^{2+} \) exchanger current (\( I_{NCX} \)) is upregulated.\(^{64}\) These changes are also predicted to create a proarrhythmic substrate within the atrium albeit through a different mechanism to that described for tachycardia mediated remodelling; an enhanced \( I_{NCX} \) current can induce triggered activity and delayed afterdepolarisations by carrying depolarising currents.\(^{64}\)

### 1.3 Genetics of AF

In order to understand the genetic mechanisms underlying AF, it is important to gain an appreciation of basic principles of genetics and genomics. The following section
Chapter 1. Introduction

provides an overview of basic genetics followed by a discussion on the genetic basis of AF.

1.3.1 Structure of the genome

The genetic information that determines the physical and physiological traits of an individual is encoded in sequences of deoxyribonucleic acid (DNA). The DNA sequences consist of four base molecules; adenine (A), thymine (T), guanine (G), and cytosine (C). The human genome is made up of 3,000 million base pairs of DNA. The vast majority of the DNA is located in the nucleus and is organised into 23 pairs of chromosomes; 22 pairs of homologous chromosomes and one pair of sex chromosomes. Each haploid set of chromosomes is derived from one parent. During fertilisation, a zygote receives a haploid set of chromosome from the oocyte and a complementary haploid set of chromosomes from the sperm.

A chromosome consists of a very long DNA molecule surrounded by a protein matrix. The main function of a chromosome is to carry genes. Genes are the basic functional units of the genome that encode for proteins or structural RNA molecules. The human genome contains approximately 22,000 genes. The coding sequences of genes, referred to as exons, are interrupted by intervening non-coding sequences known as introns. In addition to the exons and introns, genes consist of a number of elements; 1) a promoter sequence which binds to an RNA polymerase enzyme, 2) a transcription initiation site, 3) a translation initiation site, 4) a 5′ untranslated region,
5) a translation termination codon and 6) a 3’ untranslated region. These elements are essential for gene expression and are discussed in more detail in the next section.

Overall, the coding region of the genome, or the exome, constitutes approximately 1 % of the entire DNA sequence. The vast majority of the remainder of the genome consists of non-coding inter- or intragenic sequences of DNA. These non-coding sequences may harbour regulatory elements that influence gene expression.

1.3.2 Mechanisms of gene expression

The expression of genes into proteins is divided into two stages, transcription and translation. During transcription, DNA specifies the synthesis of messenger RNA (mRNA) using an RNA polymerase enzyme. In turn, transcription is divided into three steps; transcription initiation, elongation and transcription termination. The second stage of gene expression, referred to as translation, involves synthesis of polypeptides which are specified by the mRNA. Translation is also divided into three distinct steps; translation initiation, elongation and translation termination.

Gene expression is a remarkably intricate process which is tightly regulated at a number of different levels. One of the most important regulatory points is transcription initiation. During transcription initiation, RNA polymerase binds to DNA at the transcription initiation site of a gene. This process is dependent upon a number of regulatory elements which can broadly be divided into cis- and trans-acting
regulatory elements. In addition to these regulatory elements, the conformation of DNA also plays a role in the regulation of gene transcription. Each of these regulatory mechanisms is discussed in more detail below.

1.3.2.1 Regulation of gene expression by cis-acting regulatory elements

Cis-acting regulatory elements are located on the same DNA molecule as their target genes. Just upstream of the coding sequence of a gene lies the promoter sequence, which is an essential cis-acting regulatory element. The minimal (or core) promoter element, which is contiguous with the transcription start site, can be thought of as the fulcrum around which transcription pivots. The basal transcriptional machinery, which includes transcription factor and co-factor proteins and the RNA polymerase enzyme, assembles at the core promoter sequence.

In addition to the promoter sequence, most genes interact with cis-acting regulatory elements which may have positive or negative effects on gene transcription. Examples include enhancers, repressors and insulators (Figure 9). These non-coding regulatory elements may be located upstream, downstream or within the introns of the genes that they regulate. Cis-acting regulatory elements can be located up to 1 Mb away from their target genes. Due to the distance, these regulatory elements are dependent on appropriate chromatin structure for their normal function.

Enhancers and repressors mediate positive and negative regulation of gene
transcription respectively. Enhancers are thought to play an important role in recruitment of transcription factor proteins, which bind to promoter sequences of genes. Insulators on the other hand function by establishing discrete regulatory domains for individual genes. The main function of these regulatory elements is to protect genes from inappropriate signals originating from their surrounding environment. As illustrated in Figure 9, insulators may function by blocking the activating effects of enhancers or by acting as ‘barriers’ which set a boundary between domains of repressive and active chromatin.

1.3.2.2 Regulation of gene expression by trans-acting regulatory elements

Trans-acting regulatory elements regulate the expression of remotely located target genes. The most common trans-acting regulatory elements in the human genome are transcription factors and microRNAs (miRNA). Transcription factors are adaptor proteins which bind to specific DNA sequences in the promoter region of genes and target the assembly of complexes of proteins which are necessary for the formation of functional DNA-RNA polymerase complexes. Transcription factor proteins consistent of multiple distinct domains; 1) a DNA binding domain which binds to specific promoter sequences on genes, 2) a transcription activation domain which binds to co-activator proteins to form a transcription complex and 3) a multimerisation domain which allows binding to other transcription factors.
Figure 9. Classes of known regulatory elements for gene expression. a) Promoter sequence which binds to the general transcriptional machinery and mediates basal transcription of a gene. b) and c) depict enhancer and repressor sequences respectively. These sequences interact with the promoter sequence and can have positive or negative effects on gene transcription. d) and e) depict enhancer-blocking and barrier insulator elements respectively. These regulatory elements are responsible for blocking promiscuous activation of genes by regulatory elements of adjacent genes. From: Noonan JP et al. Annu Rev Genomics Hum Genet, 2010;11: 1-23.\textsuperscript{73}

Transcription factors typically regulate the expression of numerous target genes. In turn, individual genes are regulated by multiple different transcription factors.\textsuperscript{76} These observations reflect the complexity of the interactions between transcription factors and target genes. The human genome encodes more than 500 transcription factors.\textsuperscript{78}
These proteins may be ubiquitous or tissue-specific. Based on their structure and DNA binding domains, transcription factors are classified into multiple distinct families.

From a functional perspective, transcription factors may act to enhance or, less commonly, to repress gene expression.\(^7^6\)

The regulatory elements discussed thus far are responsible for regulation of gene transcription. miRNAs are trans-acting regulatory elements that are involved in post-transcriptional regulation of gene expression. miRNAs influence gene expression through imperfect pairing with target mRNA. In contrast to transcription factors, the vast majority of miRNAs act to repress gene expression through either mRNA destabilisation or translational inhibition.\(^7^6\) More than 700 miRNAs have been identified in the human and mouse genomes.\(^7^9\) It is predicted that each of these miRNAs regulate the expression of between 100 and 200 target genes.\(^8^0\)

### 1.3.2.3 Regulation of gene expression through chromatin structure

In the nucleus, chromosomes are wrapped around histone proteins to form chromatin. Transcriptionally active regions of DNA are more than 1000-fold extended as compared to transcriptionally quiescent regions.\(^7^1\) The specific pattern of packaging regulates accessibility to specific DNA sequences.\(^7^2\) Therefore, the packaging of DNA into chromatin can repress gene expression by inhibiting access of transcription factors to DNA.\(^7^1\) This nucleosome mediated repression can be alleviated by binding of activator proteins which alter the conformation of chromatin.
1.3.3 Genetic variation

For any given gene or regulatory element in the genome, there may exist multiple DNA sequence variants within a population. The most common or ‘normal’ DNA sequence is referred to as the wild-type allele. Rarer variants of the sequence are referred to as variant alleles. An allele is defined as an alternative form of a gene or regulatory element.\(^8^1\) A single gene or regulatory element may have multiple alleles. The position of the genome that harbours a specific allele is referred to as a locus. If the alleles at a locus are identical on both chromosomes that form a pair, they are referred to as homozygous. If they differ, they are referred to as heterozygous.

Most of the physical and physiological differences amongst individuals may be explained by variations in the DNA sequence of genes and/or regulatory elements. In any given population, genetic variation is generated continuously by the process of mutation. According to population genetic theory, the frequency of a variant in the population is determined to a large degree by its effect on reproductive fitness.\(^8^2\) The majority of genetic variants are predicted to have either no effect or very subtle effects on reproductive fitness, and are therefore likely to persist in the population. A small but important proportion of genetic variants have large functional effects, which are often deleterious. Given the pathogenicity of these variants, they tend to be selected against and therefore have a very low frequency in the general population (<<1%).\(^8^2\)
Chapter 1. Introduction

Genetic variants can arbitrarily be classified as being common or rare based on their minor allele frequency (MAF) in the general population. Common genetic variants, which are also referred to as polymorphisms, have a MAF of \( \geq 1\% \). Rare variants on the other hand have a MAF of \(< 1\% \). The majority of human traits, for instance height or susceptibility to disease, are influenced by multiple genetic variants with small effect sizes as well as environmental factors. A small proportion of traits are determined by rare monogenic variants or mutations with large effect sizes.

Genetic variants can also broadly be classified based on nucleotide composition into single nucleotide variants or structural variants. The following section provides a brief description of the different types of genetic variation.

1.3.3.1 Single nucleotide variants

Single nucleotide variants, which constitute single base-pair alterations in the DNA sequence, represent the simplest type of variation in the genome. Single nucleotide variants maybe located in the coding or non-coding region of the genome. Coding single nucleotide variants maybe nonsynonymous, i.e. resulting in a change of amino acid sequence of the gene product, or synonymous i.e. not altering the amino acid sequence. Nonsynonymous variants can be further classified into missense and nonsense variants. Missense variants are characterised by replacement of one amino acid for another. Nonsense variants on the other hand are characterised by the introduction of a premature stop codon resulting in an incomplete protein product.
The terminology used to describe a single nucleotide variant depends on the frequency of the allele in the general population. Variants that occur with a MAF of ≥ 1% are referred to as single nucleotide polymorphisms (SNP). Traditionally, single nucleotide variants with a MAF < 1% have been referred to as mutations. However, in recent years, the use of this terminology has changed. These variants are more commonly referred to as low-frequency variants. The term ‘mutation’ is generally reserved for rare variants that cause a disease phenotype.

SNPs account for approximately 90% of the sequence variation in the genome. The total number of SNPs in the human genome is estimated at 38 million. A minority of these SNPs (approximately 10,000–50,000) are located in the coding region of the genome. In addition to SNPs, numerous low frequency single nucleotide variants (MAF < 1%) are found in the general population. In contrast, mutations that cause monogenic diseases are extremely rare (MAF << 1%) and are typically restricted to isolated individuals or pedigrees.

1.3.3.2 Structural variants

The term structural variant broadly refers to any genetic variant that is not a single nucleotide variant. A range of structural variants have been identified in the human genome, including tandem repeat sequences, insertions and deletions, copy number variants (CNV), translocations, and inversions.
Chapter 1. Introduction

Tandem repeat sequences are a class of genetic variation characterised by repeats of nucleotide motifs consisting of anything from a single base pair to thousands of base pairs. Although definitions vary, repeats of less than nine nucleotides are generally referred to as microsatellites or short tandem repeats (STR) while those consisting of longer repeats are referred to as minisatellites. Tandem repeat sequences represent the most variable type of DNA sequence in the human genome. The majority of these sequences are located in non-coding regions of the genome and do not exert functional effects. However, a fraction of these variants are located in the coding region of the genome and may alter protein function by causing frameshift mutations or alterations in the length of amino acid sequences.

Indels refer to insertions or deletions of DNA segments ranging from 100 to 1000 base pairs. Indels represent the second most common type of genetic variation in the human genome. Close to two million indels have been identified to date. Subtypes of indels include repeat expansions, transposon insertions and random sequences. Coding indels may result in alterations in protein structure through either precise insertion or deletion of amino acids, or by causing frame shift mutations. Non-coding indels also have the potential to affect gene function by altering important regulatory regions.

CNVs are defined as segments of DNA that are larger than 1 kb in size and that display variable numbers of copies in comparison to the reference genome. CNVs can be
Chapter 1. Introduction

classified into deletions, insertions and duplications. The distinction between indels and CNVs is based on size and is largely arbitrary. CNVs have the potential to alter gene dosage directly by altering the number of copies of a gene, or indirectly by altering the number of regulatory elements that enhance or repress gene expression.

Inversions are relatively rare genetic variants characterised by segments of DNA with reversed orientation relative to the reference genome. Inversions arise as a result of breakage of DNA segments and reinsertion in reversed orientation. Inversions are classified as pericentric if they include the centromere, or paracentric if the centromere is not involved. Inversions are not associated with a change in the total DNA content. Therefore, these variants are less likely to result in abnormal phenotypes. However, if the chromosome breaks occur within genes, inversions may be disease causing.

Translocations refer to alterations in the position of chromosomal segments within the genome. These changes may be inter- or intra-chromosomal. Balanced translocations involve the reciprocal exchange of segments of DNA between chromosomes. Translocations are not associated with a change in the total DNA content. However, these variants have the potential to alter the regulatory landscape of the genome by altering the spatial organisation of DNA.
1.3.4 Genetics of human disease

The majority of human diseases have some underlying genetic component. The extent to which the genetic component influences disease varies widely. The simplest forms of genetic disease are Mendelian diseases. These diseases are typically caused by single, highly penetrant mutations which result in discrete phenotypes. More complex genetic diseases, or ‘complex traits’ are caused by multiple genetic variants which may act in concert with environmental factors. These variants are individually predicted to exert more subtle effects on the phenotype. Mendelian diseases and complex traits are discussed in more detail below.

1.3.4.1 Mendelian diseases

Mendelian diseases are determined by the genotype at a single locus. If only a single allele is sufficient to express the disease phenotype, the allele is said to have a dominant effect. On the other hand, if two alleles are necessary to express the phenotype, the allele is said to be recessive. Depending on the effect of a variant and its location (i.e. on an autosomal or sex chromosome), five different patterns of inheritance of Mendelian diseases are recognised; autosomal dominant, autosomal recessive, X-linked dominant, X-linked recessive and Y-linked.

An important phenomenon when considering Mendelian diseases, particularly those with a dominant pattern of inheritance, is the penetrance of the causative variant. Penetrance is defined as the probability that an individual harbouring a variant will
express a trait. In most Mendelian pedigrees, the assumption is that the causative variant is both necessary and sufficient to cause the trait.\textsuperscript{93} In other words, the variant is fully penetrant. However, in a proportion of dominant traits, non-penetrance or incomplete penetrance of the causative variant is an important complication.\textsuperscript{93} Therefore, individuals who harbour a causative variant may either be phenotypically normal or may express a subtle form of the phenotype.

Traditional methods for the identification of disease genes in Mendelian families include linkage analysis, karyotyping, homozygosity mapping, and CNV analysis.\textsuperscript{94} Linkage analysis is a particularly effective technique for mapping disease genes in large Mendelian pedigrees. The following section contains a more detailed discussion on the principles of linkage analysis.

### 1.3.4.1.1 Mapping of disease genes in Mendelian pedigrees

Linkage analysis is the most widely used technique for mapping of disease genes in Mendelian pedigrees. Linkage analysis is based on two important and related concepts, genetic linkage and genetic recombination. When two genes are located on different chromosomes, transmission of the genes from one generation to the next is predicted to be random. Therefore, the genes are said to be unlinked.\textsuperscript{95} On the other hand, when two genes are located on the same chromosome, one might predict that they are transmitted together 100\% of the time. In other words, they are completely
Chapter 1. Introduction

linked. However, the latter assumption does not take into account genetic recombination.

During the process of reproduction, gametes are produced by germ cells through the process of meiosis. In order to generate genetic diversity in gametes, meiosis involves an exchange of genetic material between maternal and paternal chromosomes through genetic recombination. Recombination refers to the crossover of fragments of DNA between adjacent maternal and paternal chromosomes.74

Due to recombination events, most genes on the same chromosome are not completely linked but rather exist in a state of incomplete linkage.96 Only genes that lie in very close proximity to one another are predicted to be completely linked. As the distance between two loci increases, the number of crossovers is expected to increase. Therefore, the number of crossovers between two loci can be used to infer the genetic distance.

The recombination fraction, which is denoted by the symbol \( \Theta \), is a reflection of how often two genes or loci are separated during meiosis. The value of \( \Theta \) varies between 0 and 0.5. At one extreme, genes that are located on different chromosomes, and therefore display independent assortment, have a \( \Theta \) of 0.5. At the other extreme, adjacent genes on the same chromosome that have no recombination between them...
have a $\Theta$ of 0. Those genes that are located on the same chromosome, but have a degree of recombination between them, have a $\Theta$ of between 0 and 0.5.95

In practical terms, the recombination fraction can be determined by counting the number of recombinant offspring (or haplotypes) from a double heterozygous cross.97 An example of calculation of the recombination fraction is illustrated in Figure 10.

---

**Figure 10:** A heterozygous cross (generation 2) illustrating calculation of recombination fraction. The segregation of two sets of alleles (A1 and A2 and B1 and B2) through a pedigree is illustrated. The individuals in generation III who received the A1B2 or A2B1 combination from their father have had a recombination event. Therefore, the individuals highlighted in blue have had a recombination event. The recombination fraction is 0.29. From: Strachan T, Read AP. Human Molecular Genetics. 2nd edition. New York: Wiley-Liss; 1999.97

The basic principle of linkage analysis involves investigating the co-segregation of a disease locus with a series of markers of known position distributed throughout the...
genome. Each typed marker is tested for linkage with the disease phenotype, and by implication, the disease locus. Markers used for linkage analysis include SNPs and microsatellite repeats.

In its simplest form, linkage analysis involves counting the number recombinants and non-recombinants between individual markers and the disease phenotype and calculating the recombination fraction.\(^9\) Markers that segregate with the disease phenotype, and hence lie in close proximity to the disease locus, have a low recombination fraction. Therefore, the identification of a series of markers with a low recombination fraction can be used to define the boundaries of a disease locus. Of note however, such an approach is only possible if precise information on the inheritance patterns of markers is available. Specifically, individuals included in the analysis must fulfil two criteria; firstly, they must be informative for linkage i.e. genotype known and doubly heterozygote and secondly, they have to be phase known i.e. the ancestral origin of the observed alleles must be known in order to allow reconstruction of the haplotypes.\(^9\)

In most pedigrees, the available information is incomplete. For instance some family members may not have genetic data while in others, the phase of the alleles may not be known. In other words, it is not possible to determine which alleles are received as haplotypes from which parent.\(^9\) As a result precise inheritance patterns cannot be defined and multiple inheritance patterns can generate the observed data. In order to
overcome this limitation, statistical methods have been developed to calculate the likelihood of linkage in Mendelian pedigrees. The statistical analyses involve calculation of the probability of each individual inheritance pattern against the inheritance pattern of the trait.\textsuperscript{100}

The most efficient statistical test to assess the likelihood of linkage between a disease locus and a marker locus is the logarithm of the odds (LOD) score. The LOD score compares the likelihood of two alternative hypotheses; the first is that the marker and disease locus co-segregate due to true linkage, while the second is that co-segregation occurs purely by chance. A LOD score of 3 is conventionally considered as statistical evidence of linkage.

The identification of a series of markers with a high LOD score can be used to define the boundaries of a disease locus.\textsuperscript{99} Once the boundaries of the locus have been demarcated, further analysis is necessary to identify the causative gene. Potential candidate genes at the locus can be identified using online genome databases and pathophysiological considerations. Ultimately, the identification of a causative mutation requires systematic analysis of putative genes at the locus by mutation screening and sequencing.

\textbf{1.3.4.2 Complex trait diseases}

Complex traits are caused by multiple genetic variants which may act in concert with
Chapter 1. Introduction

Environmental factors. The majority of common diseases in the general population are complex traits. A popular hypothesis regarding the genetic basis of complex diseases posits that common genetic variants underlie common diseases. On the basis of this hypothesis, referred to as the ‘common disease common variant hypothesis’, the majority of studies investigating complex traits have focused on very common variants (MAF > ~5%).

Complex traits are not amenable to traditional methods of gene discovery such as linkage analysis. This is primarily because classic linkage studies require the assumption of a precise genetic model. Association studies are considered more effective for the investigation of the genetic basis complex traits. In contrast to linkage studies, which investigate the cosegregation of genetic markers and disease traits, association studies investigate the co-occurrence of genetic markers and disease traits. In practical terms, association studies are case-control studies which involve a comparison of the frequency of an allele between cohorts of affected and unaffected individuals. Alleles that occur with a higher frequency in affected individuals are said to be associated with disease.

The identification of genetic variants underlying complex traits using association studies is challenging for two main reasons. Firstly, as discussed above, variants associated with complex traits commonly have modest effect sizes and therefore very large sample sizes may be required for their detection. Secondly, complex traits are
Chapter 1. Introduction

associated with gene-gene and gene-environment interactions. It is often very difficult to take account of these interactions when performing data analysis.

There are two main research paradigms for genetic association studies; candidate-gene association studies and genome-wide association studies (GWAS). Both types of association study are based on genotyping of SNPs. Much of the early research in complex trait genetics involved candidate-gene association studies. However, since the completion of the human genome project\textsuperscript{103} and HapMap project\textsuperscript{104} more than a decade ago, research into the genetic basis of complex traits has undergone a revolution. These sequencing projects have led to the identification of millions of SNPs in the human genome and have also led to the realisation that SNPs are organised into linkage disequilibrium blocks.\textsuperscript{102} These findings have led to the emergence of GWAS. The following section discusses candidate-gene association studies and GWAS in more detail.

1.3.4.2.1 Candidate-gene association studies

Candidate-gene association studies are hypothesis-based studies that focus on SNPs occurring within specific genes or loci.\textsuperscript{105} Candidate genes may be selected on the basis of their functional role in relevant biological pathways or their position in the genome. Positional information may be based on previous linkage studies in Mendelian pedigrees or studies that have identified structural variation in the genome. Within these candidate genes or loci, SNPs that are predicted to have
Chapter 1. Introduction

functional effects, for instance nonsynonymous and splice junction SNPs, are favoured over non-functional SNPs.\textsuperscript{106}

Candidate gene association studies are associated with relatively high statistical power.\textsuperscript{107} However, they are also associated with a number of important limitations. Firstly, they depend on pre-existing knowledge about disease pathophysiology, which may be limited or indeed absent. Secondly, candidate-gene studies have a limited ability to include all possible causative variants. Thirdly, the majority of candidate-gene studies are associated with a low replication of results.\textsuperscript{108}

1.3.4.2.2 Genome wide association studies

The linkage disequilibrium structure of the genome forms the basis of GWAS. Linkage disequilibrium refers to the interdependence of multiple alleles which arises due to the proximity of the alleles on a chromosome.\textsuperscript{109} If an allele at one locus is found together with an allele at a second locus more frequently than would be predicted if they were segregating independently in the population, the loci are said to be in linkage disequilibrium.\textsuperscript{110} As illustrated in Figure 11, the concept of linkage disequilibrium in a population is related to chromosomal linkage in pedigrees. When two markers on a chromosome remain physically joined through generations of a pedigree, they are said to be linked. Figure 11 illustrates two founder chromosomes, one depicted in orange, and the other in blue. As the chromosomes are inherited in a pedigree from one generation to the next, segments of the chromosome break apart
Chapter 1. Introduction

due to recombination events. On a population scale, this effect is amplified and repeated recombination events break segments containing linked alleles. However, markers that are in close proximity to one another may continue to be linked and therefore have a higher probability of being inherited together in the population. In other words, they are in linkage disequilibrium. Over time, as recombination events occur between every possible combination of chromosomes, pairs of markers may go from linkage disequilibrium to linkage equilibrium.

If a series of SNPs are in strong linkage disequilibrium, they are predicted to be inherited together. Combinations of adjacent alleles in strong linkage disequilibrium are commonly referred to as linkage disequilibrium blocks. Specifically selected ‘tag SNPs’ can be used to capture all the genetic variation in a linkage disequilibrium block. It is estimated that between 500,000 and 1,000,000 SNPs can be used to capture close to 80% of the genetic variation in the genome.

The basic approach in GWAS involves assaying 500,000 to 1,000,000 tag SNPs distributed across the entire genome and testing for association with disease in hundreds, or in many studies, thousands of people. There are two possible positive outcomes from GWAS. The first is that the SNPs associated with disease in GWAS are functional SNPs and directly influence the phenotype by affecting biological pathways. The second is that the SNPs associated with disease do not influence the phenotype directly but rather are in linkage disequilibrium with the disease causing variants. The
latter outcome is by far the most common.

**Figure 11. Linkage and linkage disequilibrium.** Linkage occurs when genetic markers are not broken apart by genetic recombination (illustrated as red lines). From a population perspective, contiguous stretches of founder chromosomes from the initial generation are progressively reduced in size by recombination. Over time, as recombination events occur between every possible combination of chromosomes, pairs of markers go from linkage disequilibrium to linkage equilibrium. From: Bush WS, Moore JH (2012) Chapter 11: Genome-Wide Association Studies. PLoS Comput Biol 8(12): e1002822.111

GWAS are powerful tools to investigate how common genetic variants influence disease susceptibility.113 In contrast to candidate-gene association studies, GWAS are an unbiased genetic mapping approach that do not rely on *a priori* assumptions about
biological pathways underlying a given disease. GWAS therefore have the potential to identify novel disease susceptibility loci. However, refinement of an association signal following a GWAS may be necessary in order to identify causal variants involved in disease pathogenesis.

1.3.5 Monogenic mutations in AF

Inherited forms of AF, characterised by Mendelian segregation of the trait, have been recognised for many decades. The first report of a familial AF came from Wolff et al in 1943.\textsuperscript{114} They identified a pedigree with three brothers who developed AF at a young age. Since this discovery, multiple monogenic AF families have been identified. While these are rare and isolated pedigrees, they have received much attention and have been the focus of most of the early studies in AF genetics.

Over the past two decades, classical linkage studies have identified multiple susceptibility loci for AF. In the majority of these studies the causative mutations at the loci have been identified, although some mutations remain elusive.\textsuperscript{115, 116} The identification of these mutations has provided important insights into the molecular pathogenesis of AF and has also facilitated the interrogation of genetic variants that may predispose to the more common form of AF. The following section outlines the genetic discoveries in Mendelian families with AF.
1.3.5.1 Ion channel mutations

The majority of mutations identified in monogenic AF families are located in genes that encode ion channel subunits (summarised in Table 1 and Figure 12). Functional analysis of these mutations has revealed either gain-of-function effects or loss-of-function effects.

**Figure 12** - Pictorial image of adjacent cardiomyocytes illustrating the genes implicated in Mendelian forms of AF and the presumed mechanism of action of the mutation. From: Mahida *et al*. Cardiovasc Res. 2011;89(4):692-700.117
### Table 1 – Summary of monogenic mutations associated with AF

<table>
<thead>
<tr>
<th>Gene</th>
<th>Gene product</th>
<th>Family/proband characteristics</th>
<th>Familial segregation yes/no</th>
<th>Functional effect of mutation</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>KCNQ1</td>
<td>α-subunit of I&lt;sub&gt;Ks&lt;/sub&gt; channel</td>
<td>Chinese family with autosomal dominant AF</td>
<td>Yes</td>
<td>Increased I&lt;sub&gt;Ks&lt;/sub&gt;</td>
<td>118</td>
</tr>
<tr>
<td>KCNQ1</td>
<td>α-subunit of I&lt;sub&gt;Ks&lt;/sub&gt; channel</td>
<td>Isolated case of AF detected in utero (Caucasian)</td>
<td>No</td>
<td>Increased I&lt;sub&gt;Ks&lt;/sub&gt;</td>
<td>119</td>
</tr>
<tr>
<td>KCNQ1</td>
<td>α-subunit of I&lt;sub&gt;Ks&lt;/sub&gt; channel</td>
<td>Caucasian family with autosomal dominant AF</td>
<td>Yes</td>
<td>Increased I&lt;sub&gt;Ks&lt;/sub&gt;</td>
<td>120</td>
</tr>
<tr>
<td>KCNE2</td>
<td>β-subunit of I&lt;sub&gt;Ks&lt;/sub&gt; channel</td>
<td>Two Chinese AF kindreds</td>
<td>Yes</td>
<td>Increased I&lt;sub&gt;Ks&lt;/sub&gt;</td>
<td>121</td>
</tr>
<tr>
<td>KCNE5</td>
<td>β-subunit of I&lt;sub&gt;Ks&lt;/sub&gt; channel</td>
<td>Isolated non-familial case of AF (Caucasian)</td>
<td>No</td>
<td>Increased I&lt;sub&gt;Ks&lt;/sub&gt;</td>
<td>122</td>
</tr>
<tr>
<td>KCNJ2</td>
<td>K&lt;sub&gt;ir&lt;/sub&gt; 2.1 channel</td>
<td>Chinese AF kindred</td>
<td>Yes</td>
<td>Increased I&lt;sub&gt;Kf&lt;/sub&gt;</td>
<td>123</td>
</tr>
<tr>
<td>KCNA5</td>
<td>K&lt;sub&gt;v&lt;/sub&gt; 1.5 channel</td>
<td>Caucasian proband with refractory AF</td>
<td>Yes</td>
<td>Reduced I&lt;sub&gt;Kvur&lt;/sub&gt;</td>
<td>124</td>
</tr>
<tr>
<td>SCN5A</td>
<td>Sodium channel α-subunit</td>
<td>Caucasian family with AF, DCM and impaired conduction</td>
<td>Yes</td>
<td>Predicted to have reduced I&lt;sub&gt;Na&lt;/sub&gt;</td>
<td>125</td>
</tr>
<tr>
<td>SCN5A</td>
<td>Sodium channel α-subunit</td>
<td>Caucasian proband with familial AF</td>
<td>Yes</td>
<td>Reduced I&lt;sub&gt;Na&lt;/sub&gt;</td>
<td>126</td>
</tr>
<tr>
<td>SCN5A</td>
<td>Sodium channel α-subunit</td>
<td>Japanese family with autosomal dominant AF</td>
<td>Yes</td>
<td>Increased I&lt;sub&gt;Na&lt;/sub&gt;</td>
<td>127</td>
</tr>
<tr>
<td>SCN1B</td>
<td>Sodium channel β-subunit</td>
<td>2 isolated non-familial cases of AF</td>
<td>No</td>
<td>Reduced I&lt;sub&gt;Na&lt;/sub&gt;</td>
<td>128</td>
</tr>
<tr>
<td>SCN2B</td>
<td>Sodium channel β-subunit</td>
<td>2 isolated non-familial cases of AF</td>
<td>Yes</td>
<td>Reduced I&lt;sub&gt;Na&lt;/sub&gt;</td>
<td>128</td>
</tr>
<tr>
<td>NUP155</td>
<td>Nucleoporin</td>
<td>Consanguineous family with early onset AF</td>
<td>Yes</td>
<td>Reduced nuclear membrane permeability.</td>
<td>129</td>
</tr>
<tr>
<td>GJA5</td>
<td>Connexin-40</td>
<td>4 isolated non-familial cases</td>
<td>No</td>
<td>Impaired intracellular coupling</td>
<td>130</td>
</tr>
<tr>
<td>NPPA</td>
<td>Atrial natriuretic peptide (ANP)</td>
<td>Caucasian family with autosomal dominant AF</td>
<td>Yes</td>
<td>Elevated mutant ANP</td>
<td>131</td>
</tr>
</tbody>
</table>
Chapter 1. Introduction

1.3.5.1.1 Potassium channel mutations

1.3.5.1.1.1 $I_{Ks}$ channel mutations

Chen et al \textsuperscript{118} provided the first link between ion channel mutations and AF. In a Chinese family with an autosomal dominant pattern of AF inheritance, they reported a missense mutation (S140G) located in the first transmembrane spanning domain of $KCNQ1$. The KCNQ1 gene encodes a pore-forming $\alpha$-subunit which associates with ancillary $\beta$ subunits to form a channel which carries the $I_{Ks}$ current. Functionally, the S140G mutant channel was associated with a marked increase in current density suggesting a gain-of-function effect. Further, in a subsequent study, the S140G mutation was demonstrated to cause marked slowing of $I_{Ks}$ channel deactivation.\textsuperscript{132}

Since Chen and colleagues original discovery, two further mutations in the KCNQ1 gene have been described. In an unusual case of AF detected \textit{in utero}, a valine-to-methionine mutation (V141M) adjacent to the aforementioned S140G mutation has been identified.\textsuperscript{119} More recently, a serine-to-proline mutation (S209P) was reported in a family with an autosomal dominant pattern of inheritance of AF.\textsuperscript{120} Both $KCNQ1$ mutations displayed a gain-of-function effect with enhanced $I_{Ks}$ current density and altered gating kinetics.

Mutations in the $I_{Ks}$ channel $\beta$-subunit genes have also been described in familial as well as isolated AF cases. $I_{Ks}$ channel $\beta$-subunits are encoded by five genes, $KCNE1$-
Chapter 1. Introduction

*KCNE5.* In a study evaluating 28 unrelated Chinese families with lone AF, Yang *et al* identified a mutation in the KCNE2 gene which resulted in an arginine to cysteine substitution (R27C). More recently, an isolated non-familial case of AF with a missense (L65F) mutation has been identified after a cohort of 158 patients were screened for KCNE5 gene mutations. Interaction of both mutant β subunits (KCNE2 and KCNE5) with the KCNQ1 channel produced a gain-of-function effect with an increased $I_{KS}$ current.

The KCNQ1 α-subunit of the $I_{KS}$ channel can associate with any one of the five accessory β-subunits (KCNE1-5). Previous studies have demonstrated that each of the β-subunits causes a specific alteration in the $I_{KS}$ current. Based on these observations, it has been proposed that alterations in the patterns of association between KCNQ1 and the β-subunits may allow modulation of the $I_{KS}$ current.

From a mechanistic perspective, the gain-of-function mutations in α- and β-subunits of the $I_{KS}$ channel are associated with increased repolarising potassium currents which in effect would abbreviate the action potential duration as well as the ERP in cardiomyocytes. These effects are likely to create a profibrillatory substrate within the atrium.
Chapter 1. Introduction

1.3.5.1.2 \( I_{K_1} \) channel mutations

In 2005, Xia et al.\(^{123}\) reported a novel missense mutation in the KCNJ2 gene in a Chinese AF kindred. \( KCNJ2 \) encodes the Kir2.1 channel which underlies the inward rectifier potassium current, \( I_{K_1} \).\(^{22}\) A valine-to-isoleucine mutation (V93I) was identified which resulted in a gain-of-function effect with increased potassium current amplitudes, both in the inward and outward directions. Of note, enhanced inward-rectifier currents have been demonstrated to promote AF by accelerating and stabilising atrial rotors that maintain the arrhythmia.\(^{135}\)

1.3.5.1.3 \( I_{Kur} \) channel mutations

Olson and colleagues\(^{124}\) identified a loss-of-function potassium channel gene mutation associated with AF. In a proband with lone AF which was refractory to conventional therapy, they reported a heterozygous nonsense mutation (E375X) in the KCNA5 gene. \( KCNA5 \) encodes the Kv1.5 channel which underlies the ultrarapid delayed rectifier (\( I_{Kur} \)) current. \( I_{Kur} \) is an important repolarising current specific to the atrium.\(^{22}\) Functional analysis of the mutant Kv1.5 channel revealed prolongation of the atrial action potential and triggered activity with stress, effects that would be predicted to promote initiation of AF. More recently, three further loss-of-function \( KCNA5 \) mutations (T527M, A576V and E610K) have been reported in 4 families after screening a total of 120 families.\(^{136}\)
1.3.5.1.4 Potassium channel mutations in AF cohorts

Following on from early linkage studies in monogenic AF kindred, investigators broadened their studies to screen cohorts of unrelated AF patients for mutations using a candidate-gene approach. In 2004, Ellinor et al screened a cohort of 141 patients with lone AF for KCNQ1 mutations and failed to identify any mutations. In a subsequent study, the same group screened 96 unrelated probands with familial AF for mutations in the KCNJ2 and KCNE1-5 genes and once again found no evidence of causal mutations. In a study by Otway et al, four potassium channel genes (KCNQ1, KCNE1, KCNE2, and KCNE3) were screened for mutations in 50 AF families. Only one missense mutation in the KCNQ1 gene was identified. Functional analysis of the mutant gene product did not demonstrate altered channel kinetics suggesting that the KCNQ1 mutation might not be causative. Taken together, these data suggest that potassium channel mutations are not a major cause of AF in the general population.

1.3.5.1.2 Sodium channel mutations

Mutations in the genes encoding both the α- and β-subunits of the voltage gated sodium channel (Nav1.5) have been reported in patients with AF. Mutations in the SCN5A gene - which encodes the major pore-forming α-subunit of the sodium channel - are also associated with a spectrum of other arrhythmias. Further, mutations in the function-modifying sodium channel β-subunits (SCN1B and SCN3B) have been associated with Brugada syndrome and cardiac conduction disease. As a
result, patients with AF associated with sodium channel mutations often have complex overlapping phenotypes.

In 2005, Olson et al\textsuperscript{125} reported an SCN5A mutation (D1275N) in a large multigenerational family. The mutation was associated with variable clinical manifestations which included AF, dilated cardiomyopathy and abnormal cardiac conduction. Based on reports from other studies, the D1275N mutation is expected to cause a loss of Nav1.5 channel function.\textsuperscript{145} In a more recent study, a cohort of 189 AF patients was screened for SCN5A mutations and a single missense mutation (N1986K) was identified in one AF kindred. Functional analysis of the mutation revealed a loss-of-function effect with a hyperpolarised shift of steady-state inactivation of the channel.\textsuperscript{126}

The role of the function-modifying sodium channel β-subunits in arrhythmic cardiac diseases is less clearly defined. In a recent study of 480 AF patients, Watanabe et al\textsuperscript{128} screened the four beta-subunit genes (SCN1B-SCN4B) for mutations and reported 2 mutations in SCN1B (R85H, D153N) and 2 mutations in SCN2B (R28Q, R28W). Functional analysis of the mutant β1- and β2-subunits demonstrated altered channel gating and a reduction in $I_{Na}$, indicating a loss-of-function effect.

Makiyama et al\textsuperscript{127} recently reported a gain-of-function SCN5A mutation associated with AF. They identified a novel missense mutation (M1875T) in a Japanese family
with autosomal dominant hereditary AF. Analysis of the mutant Nav1.5 channel demonstrated that the voltage dependence of steady-state inactivation was shifted in the depolarising direction, suggesting a gain-of-function.

The electrophysiological mechanisms by which sodium channel mutations cause AF are not clearly understood. Increased inward sodium currents induce triggered activity and stabilise high frequency rotors. However, they also make reentry less likely. Conversely, reduced sodium current density promotes reentry by shortening action potential duration and shortening the atrial reentry wavelength. However, attenuation of the sodium current also destabilises high frequency rotors. Overall, multiple effects in various experimental models make it difficult to predict a priori what the effects of sodium channel dysfunction will be.

Consistent with the data reported for potassium channel gene mutations, mutations in genes encoding sodium channel subunits do not appear to be a common cause of AF. Chen and colleagues screened a cohort of 157 lone AF patients and did not identify any SCN5A mutations. Similarly, Ellinor et al identified SCN5A mutations in only one kindred out of a cohort of 189 AF patients and Watanabe et al identified only four patients with SCN1B mutations in a cohort of 480 patients. Darbar and colleagues sequenced the SCN5A gene in a cohort of 375 AF patients and discovered 8 novel variants. However segregation analysis suggested that only 6 of the novel SCN5A variants are associated with AF.
1.3.5.2 Non-ion channel mutations

1.3.5.2.1 Nucleoporin gene (NUP155) mutation

In 2004, Oberti et al.\textsuperscript{129} identified a large consanguineous family from Uruguay with autosomal recessive inheritance of AF. The pattern of disease was characterised by an early onset of AF at the foetal or infantile stage with severe associated complications including cardiomyopathy, ventricular arrhythmias and sudden death. The locus was mapped on chromosome 5p13 and a homozygous mutation (R391H) in a nucleoporin gene (NUP155) was identified.\textsuperscript{149}

\textit{NUP155} encodes a nucleoporin which is an essential molecular component of the nuclear pore complex (NPC).\textsuperscript{150} NPCs mediate exchange of macromolecules between the nucleus and the cytoplasm.\textsuperscript{151} The mechanistic link between the \textit{NUP155} mutation and AF remains unclear. It has been proposed that a reduction in nucleocytoplasmic transport due to NUP155 deficiency may alter expression of atrial genes, which in turn may influence cellular processes such as maturation of calcium handling proteins and ion channels. These effects may ultimately alter the action potential duration and promote AF. An alternative hypothesis is that altered function of the nuclear envelope due to \textit{NUP155} deficiency may reduce myocyte survival by blocking mitosis. Myocyte apoptosis may promote cardiac fibrosis and conduction heterogeneity which may in turn create a substrate for arrhythmia.\textsuperscript{48}
1.3.5.2.2 Connexin-40 gene (GJA5) mutations

In a study by Gollob et al \(^{130}\) involving a small cohort of unrelated patients with lone AF, four novel mutations were identified in the GJA5 gene. GJA5 encodes connexin-40, a gap junction protein in the atrium which plays a critical role in mediating coordinated conduction of the action potential through cell-to-cell electrical coupling.\(^{152}\) Interestingly, only one of the patients had a germ-line sequence variant. The three remaining patients had tissue-specific mutations suggesting that somatic mutations could also be involved in AF predisposition. Functional analysis of the mutant gene product revealed abnormal intracellular transport in addition to a reduction in electrical coupling between cells. It has been proposed that impaired cell-cell electrical coupling results in conduction heterogeneity, microreentrant circuits and hence AF.\(^{130}\)

1.3.5.2.3 Atrial natriuretic peptide gene (NPPA) mutation

Hodgson-Zingman et al \(^{131}\) reported on a family with an autosomal dominant pattern of AF which cosegregated with a frameshift mutation in the gene encoding atrial natriuretic peptide (NPPA). The mutation was associated with markedly elevated levels of mutant atrial natriuretic peptide (ANP). ANP is involved in the regulation of sodium and water homeostasis and arterial blood pressure. In response to volume expansion and atrial stretch, ANP release causes natriuresis, diuresis and vasodilator effects.\(^{153}\)
At this stage, the role of ANP in the pathogenesis of AF is not clear. Previous studies have demonstrated electrophysiological derangements on exposure of atrial myocytes to pathophysiological doses of ANP. Further, the mutant peptide in the above AF kindred was demonstrated to shorten atrial monophasic action potentials in an isolated whole-heart animal model. These effects may provide a potential electrophysiological substrate for arrhythmia. An alternative plausible hypothesis is that excessive ANP causes structural atrial remodelling due to its pro-apoptotic effect.

1.3.6 AF in the general population

For many years, it was assumed that the heritability of AF is restricted to the rare familial forms of the arrhythmia discussed in the previous section. However, over the past several years, this paradigm has been challenged. A number of population-based studies have demonstrated that the heritability of AF is not restricted to rare families but rather extends to AF in the general population. These discoveries have been followed by a rapid expansion in the number of studies aimed at elucidating the genetic substrate underlying AF in the general population. The following section discusses studies in population-based AF genetics.

1.3.6.1 Heritability of AF

In 2003, investigators from the Framingham Heart Study demonstrated that one-third of AF patients have a first-degree relative with the arrhythmia. Further, they
reported that among the offspring of AF patients, the odds ratio (OR) for AF was 1.85 (p = 0.02). In a subsequent study from Iceland, Arnar and colleagues reported that in comparison to controls, first-degree relatives of AF patients had a relative risk of 1.77 (p < 0.001) for the arrhythmia. \(^{157}\) Ellinor et al also demonstrated a familial aggregation of AF. Specifically, they demonstrated that 38% of subjects with lone AF had at least one relative with the arrhythmia. \(^{158}\)

In contrast to rare monogenic forms of AF, the common form of AF encountered in the general population is likely to have multifactorial or complex inheritance. In other words, disease risk is likely to be influenced by multiple genetic variants which interact with environmental factors. According to the common disease-common variant hypothesis, common genetic variants underlie complex traits and these variants individually exert a relatively subtle effect.

**1.3.6.1.1 Candidate gene studies in AF**

In recent years, case-control association studies in AF cohorts have identified a variety of SNPs that influence susceptibility to the arrhythmia. Examples include SNPs in cardiac potassium channel genes, sodium channel genes, genes that regulate ion channel function, gap junction protein genes, genes encoding circulating hormones, and genes encoding inflammatory mediators. \(^{117}\)
Interestingly, some of the association studies have identified variants that are predicted to cause functional alterations in the same ion channels as those implicated in monogenic forms of AF. Examples include variants in \textit{KCNE1} and \textit{KCNE5}, which encode the $\alpha$- and $\beta$-subunits of the $I_{Ks}$ channel respectively and \textit{SCN5A}, which encodes the $\alpha$-subunit of the $I_{Na}$ channel.\textsuperscript{117} In addition, one of the reported variants is located in the gene that encodes the $\beta_3$-subunit of the heterotrimeric G protein (\textit{GNB3}) which has been linked with an increased inward rectifier current ($I_{K1}$).\textsuperscript{159,160} These results suggest that the same molecular mechanisms may underlie familial and sporadic forms of AF. Of note however, the majority of case-control association studies in AF cohorts have been limited by relatively small sample sizes, inconsistent replication, and a low pre-test probability of the polymorphism actually causing AF. The results from these studies are summarised in Table 2.
Table 2 – Summary of results from association studies in AF cohorts

<table>
<thead>
<tr>
<th>Gene</th>
<th>Polymorphism</th>
<th>Cases</th>
<th>Controls</th>
<th>Comment</th>
<th>P-value</th>
<th>Odds ratio</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>KCNE1 minK</td>
<td>38G</td>
<td>331</td>
<td>441</td>
<td></td>
<td>0.004</td>
<td>1.73</td>
<td>161</td>
</tr>
<tr>
<td>KCNE1 minK</td>
<td>38G</td>
<td>108</td>
<td>108</td>
<td></td>
<td>0.024</td>
<td>1.80</td>
<td>162</td>
</tr>
<tr>
<td>KCNE5</td>
<td>97T</td>
<td>158</td>
<td>96</td>
<td></td>
<td>0.007</td>
<td>0.52</td>
<td>163</td>
</tr>
<tr>
<td>KCNH2</td>
<td>K897T</td>
<td>1207</td>
<td>2475</td>
<td></td>
<td>0.0003</td>
<td>1.25</td>
<td>164</td>
</tr>
<tr>
<td>GNB3</td>
<td>C825T</td>
<td>291</td>
<td>292</td>
<td></td>
<td>0.02</td>
<td>0.46</td>
<td>159</td>
</tr>
<tr>
<td>eNOS</td>
<td>2786C</td>
<td>331</td>
<td>441</td>
<td></td>
<td>0.01</td>
<td>1.50</td>
<td>161</td>
</tr>
<tr>
<td>eNOS</td>
<td>G894T</td>
<td>51</td>
<td>289</td>
<td>HF patients</td>
<td>0.001</td>
<td>3.2</td>
<td>165</td>
</tr>
<tr>
<td>SCN5A</td>
<td>H558R</td>
<td>157</td>
<td>314</td>
<td></td>
<td>0.002</td>
<td>1.6</td>
<td>148</td>
</tr>
<tr>
<td>GJA5</td>
<td>~44AA/+71GG</td>
<td>173</td>
<td>232</td>
<td></td>
<td>&lt;0.006</td>
<td>1.51</td>
<td>166</td>
</tr>
<tr>
<td>AGT</td>
<td>M235T</td>
<td>250</td>
<td>250</td>
<td></td>
<td>&lt;0.001</td>
<td>2.5</td>
<td>167</td>
</tr>
<tr>
<td>AGT</td>
<td>G-6A</td>
<td>250</td>
<td>250</td>
<td></td>
<td>0.005</td>
<td>3.3</td>
<td>167</td>
</tr>
<tr>
<td>AGT</td>
<td>G-217A</td>
<td>250</td>
<td>250</td>
<td></td>
<td>0.002</td>
<td>2.0</td>
<td>167</td>
</tr>
<tr>
<td>AGT</td>
<td>T174M</td>
<td>968</td>
<td>8267</td>
<td></td>
<td>0.05</td>
<td>1.2</td>
<td>168</td>
</tr>
<tr>
<td>AGT</td>
<td>20C/C</td>
<td>968</td>
<td>8267</td>
<td></td>
<td>0.01</td>
<td>1.5</td>
<td>168</td>
</tr>
<tr>
<td>ACE</td>
<td>D/D</td>
<td>51</td>
<td>289</td>
<td>HF patients</td>
<td>0.016</td>
<td>1.5</td>
<td>165</td>
</tr>
<tr>
<td>ACE</td>
<td>D/D</td>
<td>404</td>
<td>520</td>
<td></td>
<td>&lt;0.001</td>
<td>1.89</td>
<td>169</td>
</tr>
<tr>
<td>MMP2</td>
<td>C1306T</td>
<td>196</td>
<td>873</td>
<td></td>
<td>1.2x10^-2</td>
<td>8.1</td>
<td>170</td>
</tr>
<tr>
<td>IL10</td>
<td>A-592C</td>
<td>196</td>
<td>873</td>
<td></td>
<td>3.7x10^-3</td>
<td>0.32</td>
<td>170</td>
</tr>
<tr>
<td>IL6</td>
<td>G-174C</td>
<td>26</td>
<td>84</td>
<td>Post-operative AF</td>
<td>&lt;0.001</td>
<td>3.25</td>
<td>171</td>
</tr>
<tr>
<td>SLN</td>
<td>C-65G</td>
<td>147</td>
<td>92</td>
<td></td>
<td>0.011</td>
<td>1.98</td>
<td>172</td>
</tr>
</tbody>
</table>

Abbreviations: ACE, Angiotensin converting enzyme; ATG, Angiotensinogen; CABG, coronary artery bypass graft surgery; eNOS, endothelial nitric oxide synthase 3; GNB3, guanine nucleotide binding protein; GJA5, connexin-40; HF, heart failure; IL6, interleukin 6; IL10, interleukin 10; MMP, matrix metalloproteinase; SLN, sarcolipin gene.
Chapter 1. Introduction

1.3.6.1.2 Genome-wide association studies in AF

1.3.6.1.2.1 The association between the 4q25 locus and AF

In 2007, the first GWAS in AF reported a susceptibility locus on the long arm of chromosome 4 (4q25). Two polymorphisms at the locus (rs2200733 and rs10033464) were associated with AF.\(^{173}\) rs2200733 was associated with an OR of 1.71 (\(p = 6.1\times 10^{-41}\)) while rs10033464 was associated with an OR of 1.42 (\(p =3.1\times 10^{-11}\)). Replication analyses were performed in two different ethnic populations. Amongst European cohorts, both SNPs were consistently associated with AF. In a Chinese cohort on the other hand, only one of the SNPs, rs2200733 demonstrated a significant association. Overall, the European cohorts consisted of 3,580 patients with AF and 19,256 control subjects.\(^{174}\)

Multiple subsequent studies have independently demonstrated the association between the rs2200733 SNP at the 4q25 locus and AF. Once again, the association between rs10033464 and AF was not consistent across all cohorts. More recently, further compelling evidence of the association between rs2200733 and AF came from a large meta-analysis which included 10,115 patients with AF and 65,229 control subjects. The OR for the association in this study was 1.68 (95% confidence interval (CI) 1.50–1.87, \(p = 7.0 \times 10^{-20}\)).\(^{174}\) Of note, the magnitude of risk conferred by variants at this locus is comparable to other commonly accepted risk factors for AF, and appears to be independent of such risk factors.\(^{175}\)
Chapter 1. Introduction

The association signal at the 4q25 locus lies within a genomic region that has not been reported to encode any genes or transcripts. It is therefore challenging to establish a mechanistic link between this locus and AF. The closest gene, which is located 150,000 bases away, is designated paired-like homeodomain transcription factor 2 or PITX2. PITX2 encodes a transcription factor which has been reported to be involved in embryonic cardiac development. Specifically, PITX2 is involved in determination of right-left asymmetry, suppresses the default formation of a SA node in the left atrium, and is also necessary for development of the pulmonary venous myocardium. These observations make it an intriguing candidate gene for AF. Studies are currently ongoing to further characterise the role of PITX2 in AF and will be discussed in subsequent sections.

1.3.6.1.2.2 The association between the 16q22 locus and AF

The second susceptibility locus for AF was identified by investigators from the CHARGE-AF consortium. A meta-analysis of GWAS from different cohorts identified a novel susceptibility signal on chromosome 16q22. Each copy of the T allele of SNP rs2106261 carried a modest estimated 1.19-fold increase in the relative risk of AF. This was consistent in both prevalent and incident AF cases. Corroborating evidence for this finding was provided by a GWAS in an Icelandic cohort which reported an adjacent SNP at the same locus, rs2106261, as being associated with AF. These findings were initially not replicated in cohorts of Chinese descent. However, in a
more recent study, a significant association has been demonstrated between rs2106261 and AF in a Chinese Han population.\textsuperscript{180}

The top SNP at the 16q22 locus is located within one of the introns of the ZFHX3 gene.\textsuperscript{178} ZFHX3 encodes a transcription factor, the homeodomain zinc-finger protein. The potential role of this gene in cardiac development or function has yet to be determined. ZFHX3 has however previously been demonstrated to mediate neural and myogenic differentiation.\textsuperscript{181, 182} Further, in a number of neoplastic diseases, ZFHX3 has been demonstrated to function as a tumour suppressor gene.\textsuperscript{183}

1.3.6.1.2.3 The association between the 1q21 locus and AF

In an attempt to elucidate more specifically the genetic mechanisms underlying lone AF, or AF in the absence of structural heart disease, Dr.Ellinor’s group and colleagues from the CHARGE-AF consortium performed a meta-analysis of GWAS in lone AF.\textsuperscript{184} The rationale for this approach was that subjects with lone AF have a higher familial aggregation of the arrhythmia. The study involved 1,335 lone AF cases and 12,844 control subjects and the findings were replicated in two additional studies. The meta-analysis identified a novel locus for AF on chromosome 1q21. Overall, taking the primary and replication cohorts into account, the most significant SNP at this locus (rs13376333) had an OR of 1.52 ($p=1.8\times10^{-21}$). rs13376333 lies within the KCNN3 gene, which encodes the SK3 channel (also known as KCNN3 or $K_{Ca}2.3$). This small
Chapter 1. Introduction

conductance, calcium-activated potassium channel is a strong candidate gene for AF and is discussed in more detail in the next section.

In addition to the association between the 1q21 locus and AF, the study also confirmed the association with the previously reported 4q25 locus. The results are illustrated in the Manhattan plot in Figure 13 where the physical coordinates on the 22 autosomal chromosomes are plotted against the -log_{10}(P value).

Figure 13. Manhattan plot of meta-analysis results for a genome-wide association of lone AF. The -log_{10}(p value) is plotted against the physical positions of each SNP on each chromosome. The threshold for genome-wide significance, p <5x10^{-8}, is indicated by the dashed line (Figure provided by Dr. Patrick Ellinor).
1.3.6.1.2.4 Six novel loci for AF

The latest GWAS meta-analysis, also conducted by investigators from the CHARGE-AF consortium, was reported in 2011. The study, also in subjects of European descent, involved a large-scale meta-analysis of 6,707 subjects with AF and 52,426 control subjects. In the discovery stage of the study, the investigators identified 10 susceptibility loci. Three of the loci have been reported previously (as discussed above) while seven were novel. Of the novel loci, six were replicated in 5,381 AF cases and 10,030 referent subjects of European descent. The results are summarised in Table 3 and Figures 14 and 15. The replication analysis was also extended to a Japanese cohort where four of the loci were replicated.

During the discovery stage of the GWAS, the strongest observed association was on chromosome 1q24 in the PRRX1 gene. The most significant SNP at this locus, rs3903239, had an overall p value of $8.4 \times 10^{-14}$. PRRX1 encodes a homobox transcription factor which displays high levels of expression during cardiac development. Interestingly, in a murine model, ablation of PRRX1 expression results in abnormal development of the pulmonary vasculature. Further, abnormalities of great vessel formation have been reported in PRRX1/PRRX2 double mutant mice. Specifically, these mice display abnormal architecture of the aortic arch, an anomalous retro-oesophageal right subclavian artery, and an elongated ductus arteriosus.
The second locus was identified on chromosome 7q31. The most significant SNP at this locus was intronic to the CAV1 gene. \textit{CAV1} encodes caveolin-1, a transmembrane scaffolding protein which plays an important role in signal transduction. Interestingly, knockout of this gene in a murine model results in a dilated cardiomyopathy.\textsuperscript{187} Further, \textit{CAV1} is selectively expressed in the atria and has been demonstrated to regulate the activity of the potassium channel \textit{KCNH2}.\textsuperscript{188}

The third locus was identified on chromosome 14q23. The signal at this locus is intronic to the \textit{SYNE2} gene. \textit{SYNE2} encodes a nuclear membrane protein, nesprin-2. This protein anchors the nucleus to the cytoskeleton, thereby maintaining position and structural integrity of the nucleus. \textit{SYNE2} is abundantly expressed in cardiac and skeletal muscle tissue. Of note, mutations in \textit{SYNE2} have been reported to segregate with disease in pedigrees with a progressive muscle-wasting disorder termed Emery-Dreifuss muscular dystrophy.\textsuperscript{189}

The fourth locus was identified on chromosome 9q22. The signal at this locus is situated within an open reading frame (\textit{C9orf3}). Potential candidate genes at this locus include \textit{FBP1} and \textit{FBP2}. However, while these genes play an important role in gluconeogenesis, deficiency of \textit{FBP1} does not appear to be associated with a cardiac phenotype.\textsuperscript{190}
The fifth locus was identified on chromosome 15q24. The SNP at this locus is intronic to the HCN4 gene. HCN4 encodes the hyperpolarisation-activated cyclic nucleotide-gated (HCN4) channel which underlies the funny ($I_f$) current. $I_f$ is an important current underlying the generation of spontaneous action potentials in cardiac pacemaker cells. Of note, mutations in the HCN4 gene have been reported to cause dysfunction of the SA node pacemaker.\textsuperscript{191}

The sixth novel locus was located on chromosome 10q22. Two intriguing candidate genes are located in the vicinity of the signal SNP. The first, SYNPO2L is located 5 kb upstream while the second MYOZ1 is located 20 kb upstream. The proteins encoded by these genes localise to the Z-disc in cardiac and skeletal muscle.\textsuperscript{192} Based on this observation, it has been postulated that these genes play critical roles in muscle function. Consistent with this hypothesis, in a murine model with pressure overload, knockout of MYOZ1 is associated with cardiac hypertrophy and unregulated calcineurin activity.\textsuperscript{193}
Chapter 1. Introduction

Table 3 - Summary of susceptibility loci identified by GWAS

<table>
<thead>
<tr>
<th>Locus</th>
<th>SNP</th>
<th>Nearest gene</th>
<th>Location of SNP relative to nearest gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>1q21</td>
<td>rs6666258</td>
<td>KCNN3</td>
<td>Intronic</td>
</tr>
<tr>
<td>1q24</td>
<td>rs3903239</td>
<td>PRRX1</td>
<td>46 kb upstream</td>
</tr>
<tr>
<td>4q25</td>
<td>rs6817105</td>
<td>PITX2</td>
<td>150 kb upstream</td>
</tr>
<tr>
<td>7q31</td>
<td>rs3807989</td>
<td>CAV1</td>
<td>Intronic</td>
</tr>
<tr>
<td>9q22</td>
<td>rs10821415</td>
<td>C9orf3</td>
<td>Intronic</td>
</tr>
<tr>
<td>10q22</td>
<td>rs10824026</td>
<td>SYNPO2L</td>
<td>5 kb upstream</td>
</tr>
<tr>
<td>14q23</td>
<td>rs1152591</td>
<td>SYNE2</td>
<td>Intronic</td>
</tr>
<tr>
<td>15q24</td>
<td>rs7164883</td>
<td>HCN4</td>
<td>Intronic</td>
</tr>
<tr>
<td>16q22</td>
<td>rs2106261</td>
<td>ZFHX3</td>
<td>Intronic</td>
</tr>
</tbody>
</table>

Figure 14. Manhattan plot of meta-analysis results for genome-wide association of AF. The $-\log_{10}(p$ value) is plotted against the physical position of each SNP on each chromosome. The threshold for genome-wide significance, $p <5\times10^{-8}$, is indicated by the dashed line. From: Ellinor et al. Nat Genet. 2012;44(6):670-5.\textsuperscript{185}
Figure 15. Representative illustration of a cardiomyocyte demonstrating protein products of genes implicated as conferring increased risk of AF in GWAS. From: Mahida et al. J Cardiovasc Electrophysiol. 2012;23(12):1400-6.¹⁹⁴

1.4 Hypothesis and Aims

1.4.1 Aim 1: Characterisation of the mechanistic link between the KCNN3 gene at 1q21 locus and AF

GWAS provide a unique opportunity to identify previously unsuspected genes and biological pathways that may play an important role in disease pathogenesis. However, it is important to note that while strong association signals have been identified and convincingly replicated in GWAS, they are just that – association
Chapter 1. Introduction

These signals may point the way towards disease causing variants however they are unlikely themselves to be causative variants. Additional exploration is therefore necessary to move from simple association to mechanism. A range of compelling candidate genes have been identified at the GWAS risk loci for AF. Therefore, one strategy for post-GWAS analysis is to focus on these genes and to further elucidate their potential functional role in AF pathogenesis using animal models.

As discussed in Section 1.3.6.1.2.3, in a meta-analysis of GWAS in patients with lone AF, a susceptibility locus was identified on chromosome 1q21. The risk variants at this locus are clustered in the KCNN3 gene. KCNN3 encodes a potassium channel (SK3) that is abundantly expressed in the heart and therefore represents a compelling candidate gene for AF. However, the potential mechanism by which genetic variation at the KCNN3 locus influences susceptibility to AF is unclear. We therefore sought to further clarify the relation between the SK3 channel and AF.

The SK3 channel, which is also known as KCNN3, KCa2.3 or SKca3, is a small-conductance, voltage-independent, calcium-activated potassium channel. In the following section, we provide an overview on calcium-activated potassium channels before discussing SK channels in more detail.
Chapter 1. Introduction

1.4.1.1 Calcium-activated potassium channels

The calcium-activated potassium channels are a major group of potassium selective ion channels. A unique property of these channels is that they are gated directly by calcium binding and therefore they provide a link between intracellular calcium transients and membrane potassium conductance.

Calcium-activated potassium channels are classified according to potassium conductance into three distinct groups; large-conductance (BK) channels, intermediate-conductance (IK) channels and small-conductance (SK) channels. The nomenclature for these channels varies according to different committees. The commonly used designations for these channels are summarised in Table 4.

**Table 4 – Nomenclature of calcium activated potassium channels**

<table>
<thead>
<tr>
<th>IUPHAR</th>
<th>HGNC</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>KCa1.1</td>
<td>KCNMA1</td>
<td>Slo, Slo1, BK</td>
</tr>
<tr>
<td>KCa2.1</td>
<td>KCNN1</td>
<td>SK\textsubscript{Ca}1</td>
</tr>
<tr>
<td>KCa2.2</td>
<td>KCNN2</td>
<td>SK\textsubscript{Ca}2</td>
</tr>
<tr>
<td>KCa2.3</td>
<td>KCNN3</td>
<td>SK\textsubscript{Ca}3</td>
</tr>
<tr>
<td>KCa3.1</td>
<td>KCNN4</td>
<td>IK\textsubscript{Ca}1</td>
</tr>
</tbody>
</table>

Abbreviations; BK, big-conductance K\textsuperscript{+} channel; SK, small-conductance K\textsuperscript{+} channel; IK, intermediate-conductance K\textsuperscript{+} channel. From: Gutman et al. Calcium-activated potassium channels, introductory chapter. IUPHAR database (IUPHAR-DB),http://www.iuphar-db.org/DATABASE/FamilyIntroductionForward.\textsuperscript{196}

The BK channels, also referred to as KCNMA1, Slo or Slo1, are encoded by the KCNMA1 gene. The unit conductance for BK channels is between 100 and 200
Chapter 1. Introduction

pico siemens (pS). The IK channels, also designated KCNN4, K_{Ca}3.1 or IK_{Ca}1, are encoded by the KCNN4 gene. The IK channels have a unit conductance of between 20-85 pS. The SK channels consist of a family of three members; SK1 (KCNN1, K_{Ca}2.1 or SK_{Ca}1), SK2 (KCNN2, K_{Ca}2.2 or SK_{Ca}2) and SK3 (KCNN3, K_{Ca}2.3 or SK_{Ca}3). SK channels display a significantly lower single channel conductance (2-20 pS).^{197, 198}

SK channel α-subunits have six transmembrane spanning domains and are thought to assemble as homotetramers.^{199} Previous studies have also demonstrated that the SK channel subtypes interact to form heteromultimers.^{200} SK channels are voltage-independent, and channel opening is strictly dependent upon intracellular calcium concentrations. These channels are also constitutively bound to calmodulin, which mediates gating in response to calcium binding.^{201} (Figure 16)

1.4.1.2 SK channel pharmacology

SK channels are highly homologous in terms of their biophysical properties and share a number of pharmacological properties. For instance, all three SK channel subtypes are susceptible to blockade by the bee venom toxin apamin. However, they display differential sensitivity to apamin. SK2 is highly sensitive, SK3 has intermediate sensitivity, while SK1 has a low sensitivity.^{202} Other SK channel blockers include the scorpion toxins scyllatoxin, P05, tamapin and organic compounds such as curare. A number of compounds have also been demonstrated to positively modulate SK channel function including 1-EBIO and NS309.^{203, 204}
1.4.1.3 SK channel expression and function

SK channels are expressed in excitable tissues such as the myocardium and neural tissue. The putative role of these channels in excitable tissues is discussed below.
1.4.1.3.1 **SK channels in neural tissue**

SK channels are abundantly expressed in the central and peripheral nervous system. The role of these channels in neural tissue has been investigated extensively. SK channels play a prominent role in the regulation of neuronal excitability and synaptic transmission. During neuronal action potentials, transient elevation in intracellular calcium results in the activation of SK channels. The resulting outward potassium current contributes to classical afterhyperpolarisation, which moves the membrane potential away from the threshold, thus making neurons less excitable for tens to hundreds of milliseconds (Figure 17). The overall effect of SK channel activation is to make neuronal cells less sensitive to excitatory inputs and inhibit neurotransmitter release.\(^{209}\)

In whole animal studies, variable modulation of SK channels had been reported to cause a number of cognitive changes.\(^{210}\) For instance pharmacological blockade of SK channels with apamin results in enhanced learning consolidation, improved spatial memory and navigation, enhanced object recognition, and habituation.
Chapter 1. Introduction

Figure 17: Neuronal action potential illustrating afterhyperpolarisation following an action potential. Abbreviations; RMP, resting membrane potential; APamp, amplitude of action potential; AP95%, duration of action potential at 95% repolarisation; t, latency following stimulation; AHPamp, amplitude of afterhyperpolarisation; AHP80%, duration of afterhyperpolarisation until 80% recovery to baseline. From: Hogan, Q et al. Anesth Analg. Vol. 107, No. 3, September 2008.

1.4.1.3.2 SK channels in the heart

The role of SK channels in cardiac electrophysiology is currently not clearly defined. The three SK channel subtypes display variable spatial distribution in the heart. Two of the channel subtypes, SK1 and SK2, are predominantly expressed in the atrium. SK3 on the other hand is equally expressed in the atrium and the ventricle. The expression profile of SK3 suggests that this channel may play an important role in the ventricular as well as the atrial action potential.
The majority of studies relating to SK channels in the heart have focused on the SK2 subtype. Li et al.\textsuperscript{212} demonstrated that in a murine model, selective knockout of the SK2 channel results in a significant prolongation of the action potential. The effect is most pronounced during the late phase (phase 3) of the action potential (Figure 18). In keeping with the expression profile of the SK2 channel, action potential prolongation was observed in atrial but not ventricular myocytes.

The role of the SK2 channel has also been investigated in AV nodal cells. Zhang et al.\textsuperscript{213} demonstrated that consistent with the findings in atrial tissue, SK2 channel knockout in AV nodal cells results in a prolongation of the action potential. Conversely, overexpression of the channel results in an abbreviated action potential. In addition to the effect on the action potential, SK2 channel overexpression results in a significant increase in the spontaneous firing rate of AV nodal myocytes.

As discussed in Section 1.4.1.3.1, in neural tissue, SK channels mediate prolonged afterhyperpolarisation of the membrane potential which reduces neuronal excitability and tonic firing frequency.\textsuperscript{209} However, based on the aforementioned studies, SK channels appear to have a distinct role in the cardiac action potential and contribute to the late phase of repolarisation rather than hyperpolarisation.
Figure 18: Action potential recordings from SK2 knockout homozygote and heterozygote mice demonstrating prolongation of the action potential. Homozygote mice (SK2Δ/Δ, dotted line) and heterozygote mice (SK2+/Δ, dashed line) display action potential prolongation as compared to wild type controls (WT, solid line). The most pronounced effect is seen during the latter phase of the action potential. From: Li N et al. J Physiol. 2009;587(5):1087-100.212

In atrial myocytes, shortening of the action potential and ERP is predicted to promote reentry, an important mechanism underlying AF.46 One would therefore expect that overexpression of the SK channels in the atrium increases susceptibility to AF. Paradoxically however, despite a prolongation of the atrial action potential, SK2 channel knockout mice have inducible atrial arrhythmias on programmed electrical stimulation.212 A potential explanation for this observation is that AF in these mice does not arise due to reentry, but rather results from a prolongation of the action potential and triggered activity.
1.4.1.3.3 Characterising the role of the SK3 channel in cardiac electrophysiology

The role of the SK3 channel in cardiac electrophysiology is largely unknown. Interestingly, SK3 has been demonstrated to form functional heteromultimeric complexes with SK2 \textit{in vitro}. Blockade of the SK2/3 complex results in prolongation of the action potential in isolated cardiomyocytes.\textsuperscript{200} Further, in a rabbit atrial model mimicking pulmonary venous ectopy, blockade of SK2 and SK3 with apamin attenuates the acceleration of repolarisation in response to burst pacing.\textsuperscript{214} It is plausible therefore that the SK3 channel also plays an important role in cardiac repolarisation and maintenance of electrical stability in the atrium. However, the SK3 channel has not been investigated in isolation. We hypothesised that alteration in the expression of the SK3 channel influences susceptibility to AF. In order to test our hypothesis, we characterised the cardiac electrophysiological phenotype of a mouse line with variable modulation of \textit{KCNN3} expression.

1.4.2 Aim 2: Identify novel mutations for AF using a combination of linkage analysis and exome sequencing

As discussed in Section 1.3.4.2, a popular hypothesis regarding the genetic basis of complex traits posits that common genetic variants underlie common diseases.\textsuperscript{101} This hypothesis forms the basis of GWAS, which typically focus on very common variants (MAF $\sim$ 5\%). GWAS have successfully identified numerous variants underlying a range of common diseases, including AF. Despite these successes however the reported variants only account for a small fraction of the heritability of these diseases.\textsuperscript{215}
In recent years, evidence has emerged to suggest that the common variants identified by GWAS (depicted in peach in Figure 19) represent one extreme of a spectrum of variants underlying complex traits. A significant proportion of the remaining or ‘missing’ heritability is likely to be accounted for by low frequency variants (MAF <5%, depicted in red in Figure 19) and rare variants (MAF 0.5-5%, depicted in blue in Figure 19). Therefore, common diseases may be more like Mendelian traits than has been proposed by the ‘common disease-common variant’ hypothesis.


Based on the above discussion, rare and deleterious mutations in isolated pedigrees could potentially be important drivers of the heritability of AF in the general population. Therefore, pedigree-based genetic studies may represent an effective
strategy for the identification of some of the missing heritability of AF. As discussed in Section 1.3.5, traditional methods of gene discovery in Mendelian pedigrees such as linkage analysis have identified multiple rare mutations underlying AF. However, a large proportion of rare and low frequency variants are unlikely to be amenable discovery using linkage analysis alone. Recently, exome sequencing has emerged as a potentially powerful technique for the identification of mutations underlying Mendelian diseases. The following section contains a more detailed description of the technique.

1.4.2.1 Pedigree-based exome sequencing

Exome sequencing uses next-generation sequencing technology to perform simultaneous sequencing of the entire protein coding region of the genome. In comparison to Sanger sequencing, which has widely been used in genetics research for nearly three decades, next generation sequencing has allowed a scaling up of sequencing by orders of magnitude. The latest machines can sequence up to 55 billion bases in approximately 10 days.\textsuperscript{217}

Despite the fact that the exome consists of only 1\% of the genome, the majority of causative mutations underlying familial disorders are predicted to reside in the coding region of the genome.\textsuperscript{218} Therefore, exome sequencing represents an efficient strategy for the identification of causative mutations in Mendelian pedigrees. Exome sequencing has recently been used to successfully identify causative mutations for a
Chapter 1. Introduction

range of Mendelian diseases\textsuperscript{219-221}.

Exome sequencing typically identifies in excess of 30,000 genetic variants per individual.\textsuperscript{222} Almost a third of these variants are nonsynonymous, disrupt splice sites, or result in insertions or deletions. Therefore, the identification of a variant that appears to be pathogenic is not sufficient to assign causation and the identification of the causative mutation can be a challenge. A number of different strategies can be used to narrow the search for the causative mutation. In large pedigrees with multiple affected individuals, linkage analysis maybe particularly effective for filtering the number of candidate variants. As discussed previously, linkage analysis is a technique for mapping the chromosomal location of causative mutations that underlie Mendelian diseases. The basic principle of linkage analysis involves investigating the co-segregation of a disease locus with a series of markers of known position distributed throughout the genome. Markers that segregate with the disease phenotype, and hence lie in close proximity to the disease locus, are used to define the boundaries of the disease locus.\textsuperscript{99} When used in combination with exome sequencing, linkage analysis is likely to narrow the search for the causative mutation to a handful of variants.

Another approach used to filter the number of candidate variants in Mendelian pedigrees is to perform exome sequencing of multiple affected individuals.

Sequencing the most distally related affected family members and prioritising of
overlapping variants is a particularly economical strategy.\textsuperscript{217} The further the genetic distance between two members of a family, the fewer the genetic variants that they share. Therefore, this approach is predicted to significantly reduce the number of candidate variants. It is important to note however, that even distant relatives are likely to share too many potential causative variants.\textsuperscript{217} Further prioritisation of variants is based on their potential functional role in biological pathways and their expression profiles. Once a small set of promising candidate variants has been identified, segregation analysis in other family members can be performed using traditional Sanger sequencing.

\textbf{1.4.2.2 Exome sequencing in AF pedigrees}

Over the past 11 years, the Massachusetts General Hospital (MGH) AF study has recruited approximately 1200 individuals and family members with early-onset AF. Eight hundred of these are probands while a further 400 are family members. In the present study, we took advantage of two large, multigenerational families with AF from the MGH-AF study in an attempt to identify novel causative mutations for the arrhythmia. The first is a three-generation family with an autosomal dominant pattern of inheritance of AF. The second is a three-generation family with an autosomal dominant pattern of inheritance of a complex phenotype of AF and atrioventricular septal defects. We hypothesised that performing pedigree-based exome sequencing will result in the identification of novel mutations underlying AF.
Chapter 1. Introduction

1.4.3 Summary of hypothesis and aims

In a GWAS of lone AF, a novel susceptibility locus was identified on chromosome 1q21 at the KCNN3 gene. KCNN3 encodes a calcium-activated potassium channel and is a compelling candidate gene for AF. However, the mechanistic link between this gene and AF is unclear. We hypothesised that altered expression of the KCNN3 gene influences susceptibility to AF.

Despite the successes of GWAS, a large proportion of the heritability of AF has yet to be uncovered. In recent years, the role of rare genetic variants in the pathogenesis of complex traits like AF has been increasingly explored. With the emergence of next-generation sequencing techniques, it has become possible to simultaneously sequence the entire protein coding region of the genome, or the exome. Exome sequencing represents a potentially powerful technique for the identification of disease causing variants in Mendelian pedigrees. We hypothesised that exome sequencing in two AF pedigrees with uncover novel mutations underlying the arrhythmia. We proposed to test our hypotheses though the following specific aims:

Aim 1 - Characterise the cardiac electrophysiological phenotype of a mouse line with variable modulation of KCNN3 expression.

Aim 2 - Identify causative mutations for AF using linkage analysis and exome sequencing.
CHAPTER 2.
METHODS

2.1 Characterisation of the cardiac electrophysiological phenotype of a mouse line with variable modulation of Kcnn3 expression

2.1.1 Generation of SK3\textsuperscript{T} mice

A mouse strain (SK3\textsuperscript{T}) with overexpression of Kcnn3 or inducible knockout of the gene was acquired from the Jackson Laboratory (ME, USA). The SK3\textsuperscript{T} mouse was originally developed in Dr. Jonathan Adelman’s laboratory at the Vollum Institute (Portland, Oregon) by site-specific insertion of a tetracycline-based regulatory cassette upstream of the translation initiation site of Kcnn3.\textsuperscript{223} A schematic representation of the regulatory cassette and the proposed mechanism through which it modulates Kcnn3 expression is illustrated in Figure 20. The cassette contains three functional modules. The first of the modules, located at the 5’ end of the cassette, is a binary tetracycline transactivator (tTA) protein. The tTA protein is composed of the activating domain of the herpes simplex viral protein VP16 and a tet repressor. The tTA protein is preceded by an adenovirus tripartite leader sequence which results in high translation initiation efficiency. The tTA protein is followed by a polyadenylation transcriptional termination sequence from SV40 (SV40 polyA).

The second module of the cassette consists of selectable markers flanked by loxP binding sites. The selectable markers consist of a bacterial neomycin resistance gene which is driven by the herpes simplex virus thymidine kinase promoter (TKprom) and
Chapter 2. Methods

the yeast URA3 gene. The selectable markers are followed by a transcription termination signal from the human growth hormone (polyadenylation hGH polyA sequence). The third module consists of five copies of the tet operator fused to a cytomegalovirus minimal promoter (tet<sub>05</sub>CMV).

*Kcnn3* expression is under the transcriptional control of the tet<sub>05</sub>CMV promoter. At baseline, binding of tTa to the tet<sub>05</sub>CMV promoter strongly activates expression of *Kcnn3* (depicted as solid and open circles in Figure 20). In turn, the native promoter of the Kcnn3 gene drives expression of tTA. Administration of doxycycline (depicted as circles with crosses in Figure 20) interferes with the interaction between the tTA protein and the CMV promoter, thereby abolishing *Kcnn3* expression.

Following the acquisition of the *SK3<sup>+/T</sup>* mice from Jackson Laboratories, we expanded the line in our laboratory. *SK3<sup>+/T</sup>* mice were crossed to generate *SK3<sup>T/T</sup>* mice, *SK3<sup>+/T</sup>* mice and wild type (WT) siblings. Mice were maintained in a standard facility with controlled temperature and humidity with 12-hour light and dark cycles. Food and water was available *ad libitum*. All animal experiments were approved by the Partners Subcommittee on Research Animal Care (SRAC) and were conducted in compliance with the regulations published in the US National Institute of Health *Principles of Laboratory Animal Care* (NIH Publication No. 85-23, revised 1996).
Figure 20: Illustration of the $SK3^T$ regulatory cassette. The cassette consists of three functional modules. The first is a binary tetracycline transactivator (tTA) protein (solid and open circles). This sequence is preceded by an adenovirus tripartite leader sequence and flowed by a polyadenylation/transcriptional termination sequence from SV40 (SV40 polyA). The second functional module consists of the selectable markers (bacterial neomycin (G418) resistance gene driven by the herpes simplex virus thymidine kinase promoter and the yeast URA3 gene). The markers are flanked on either side by loxP sites and followed by the polyadenylation/transcription termination signal from the human growth hormone gene (hGH polyA). The third consists of a fusion between a CMV minimal promoter and $tet_{os}$ CMV. tTA expression is driven by the native promoter which in turn induces $Kcnn3$ expression via the $tet_{os}$ CMV promoter. Doxycycline (crossed circles) interferes with this interaction and ablates $Kcnn3$ expression. From: Bond et al. Science. 2000;289(5486):1942-6.\textsuperscript{223}

2.1.2 Ear tagging

Ear tagging and tail biopsies (described in the next section) were performed simultaneously in 21- to 28-day-old mice. Mice were anaesthetised with isoflurane (VetOne, USA) to minimise discomfort and stress during the tagging process.
Chapter 2. Methods

Isoflurane was delivered via an isoflurane vapouriser driven by an oxygen source with a flow meter. Anaesthesia was induced by administering 5 % Isoflurane into an induction chamber. Anaesthesia was subsequently maintained with 1-2 % Isoflurane in 95 % oxygen delivered via a modified Bain circuit mask’s inner tube. Briefly, a modified Bain circuit is a non-rebreathing circuit with an inner tube which delivers Isoflurane and a larger outer tube which evacuates waste gases. Excess anaesthetic vapour was scavenged away by drawing waste gases through a canister of activated charcoal. Mice were tested for adequate depth of anaesthesia by toe-pinchat reflex. Proper anaesthesia was deemed to have been attained when pinching of the paw consistently failed to elicit a reflex.

Sterile ear tags were placed on sanitised ear tag appliers. The tag was placed just above the skin fold approximately 3 mm from the edge of the pinna using specifically designed pliers (Harvard Apparatus, MA, USA). The ear tags were imprinted with four digits by the manufacturers. Following ear tagging, the animals were released back into their cages and monitored for pain for a period of half an hour. Instruments were sterilised between procedures by washing with ethanol and water.

2.1.3 Mouse tail biopsies

As mentioned in the previous section, tail biopsies were performed at the same time as ear tagging in 21- to 28-day-old mice. Mice were anaesthetised with isoflurane to
minimise discomfort and stress during the procedure. The procedure for induction and maintenance of anaesthesia is described in detail in Section 2.1.2.

The tail biopsy was performed using a sterile pair of scissors. Tail skin was disinfected with alcohol prior to incising the tip. A small portion of the distal tail (~5 mm) was removed for the procedure. Adequate haemostasis was achieved with local pressure using sterile gauze. 1% lidocaine with epinephrine or silver nitrate was used in cases where bleeding was not controlled. Following the tail biopsy, animals were released back into their cages and monitored for pain and to ensure adequate haemostasis had been achieved. Instruments were sterilised between procedures with ethanol.

### 2.1.4 Isolation of genomic DNA and genotyping

The Gentra Puregene kit (Qiagen, CA, USA) was used for isolation and purification of DNA. 300 μl of cell lysis solution and 1.5 μl of proteinase K were added to each 5 mm tail piece. The samples were incubated at 55°C overnight. 1.5 μl of RNAse A solution was subsequently added to the lysate and further incubation performed for one hour at 37°C. 100 μl of protein precipitation solution was added and the mixture vortexed at high speed for 20 seconds. The samples were centrifuged for 1 minute at 13000 x g to form a protein pellet. The supernatant was decanted into a new 1.5 ml tube containing isopropanol. The precipitated DNA was spun down for one minute at 13000 x g. The supernatant was discarded and 300 μl of 70% ethanol was added to wash the DNA pellet. Further centrifugation was performed for 1 minute at 13000 x g. The supernatant was discarded and the DNA pellet air dried for 10 minutes. 50 μl of
DNA hydration solution was added and the samples left overnight at room temperature to allow dissolution of DNA. DNA concentration was subsequently measured using the NanoDrop™ 1000 Spectrophotometer (Thermo Scientific, FL, USA). The DNA had an absorbance peak centred at a wavelength of 260 nm with a 260/280 ratio of approximately 1.95. All samples were diluted to a concentration of 20 ng/μl.

Genotyping of the mice was performed using a protocol from Jackson Laboratories (MA, USA). PCR primer sequences were obtained from Jackson Laboratories (Appendix Table A.1). The wild type and mutant allele had a common forward primer and unique reverse primers. The common forward primer was designed to anneal to a C57BL/6 genomic sequence approximately 100 bp upstream of the non-genomic selection cassette. As discussed in Section 2.1.1, in the SK3T line, a non-genomic selection cassette is inserted just upstream of the translation initiation site of Kcnn3. The mutant reverse primer was designed to anneal to a sequence within the non-genomic selection cassette and therefore specifically amplified the mutant allele. The wild type reverse primer was designed from the published wild type genomic sequence for the C57BL/6 strain (UCSC genome database). The wild type and mutant amplicons differ in size to allow detection by gel electrophoresis (mutant amplicon 140bp, wild type amplicon 432bp).
Chapter 2. Methods

The PCR mixture and PCR conditions for genotyping are outlined in Appendix Tables A.2 and A.3 respectively. PCR products were separated by electrophoresis on 2% agarose gels and visualised using ultraviolet light. A single band of 140 bp was seen with SK3\textsuperscript{T/T} mice, a single band of 432 bp was seen with WT mice and two bands (140 bp and 432 bp) were seen with SK3\textsuperscript{T/+} mice.

2.1.5 Administration of doxycycline

Mice that were selected for doxycycline induced abolition of \textit{Kcnn3} expression were fed water medicated with doxycycline (2mg/ml, Sigma-Aldrich, SL, USA) after they had been weaned. In order to monitor adequate fluid intake, the medicated mice were housed in separate cages and the water bottles were weighed on a weekly basis. The medicated water bottles were also changed on a weekly basis. Doxycycline administration was continued in the selected mice until they were euthanised.

In view of the fact that \textit{Kcnn3} expression was not abolished despite administration of doxycycline medicated water, we replaced the medicated water with custom made doxycycline food pellets (Bioserv, NJ, USA). The pellets have a green food colouring to allow tracking of which animals are fed with medicated chow. As with the medicated water, food intake was monitored and food stocks replenished on a weekly basis.
2.1.6 Euthanasia for tissue collection

Mice were anaesthetised with isoflourane to minimise discomfort and stress during the procedure. The procedure for induction and maintenance of anaesthesia is described in detail in Section 2.1.2. Once adequate anaesthesia had been achieved, euthanasia was performed by a sodium pentobarbital overdose. Sodium pentobarbital 100 mg/kg (Hospira, IL, USA) and 100 units of heparin (APP Pharmaceuticals, IL, USA) was administered into the peritoneum. Once animals were properly injected, sufficient time was allowed for them to develop respiratory arrest.

Following respiratory arrest, the heart was rapidly excised. A midline sternotomy was performed to open the thoracic cavity. Incisions were made in the ascending aorta, the vena cava and the pulmonary arteries to explant the heart. The heart was transferred to a petri dish and washed with cold phosphate buffered saline.

For samples to be used for histological analysis, the whole heart was immediately placed in a prefilled formalin specimen container (Fisher Scientific, IL, USA) and stored at 4°C. For samples to be used for quantitative PCR and Western blot analysis, the four chambers of the heart were resected, placed in separate tubes, and immediately snap frozen in liquid nitrogen. The samples were subsequently stored at -80°C.
2.1.7 Quantitative PCR

Real-time reverse transcriptase PCR (RT-PCR) was used to quantify relative \textit{Kcnn3} expression in the four cardiac chambers. Cardiac tissue was homogenised and total RNA isolated using a Trizol reagent (Invitrogen, NY, USA) according to manufacturer’s instructions. The tissue was cut into small pieces and homogenised using a pestle. 1 ml of Trizol reagent was added to 50-100 mg of tissue. 200 μl of chloroform was added to the Trizol reagent. The mixture was vortexed for 20 seconds and allowed to sit for 2-3 minutes after which the samples were centrifuged at 12000 x g for 15 minutes at 4°C. The upper aqueous phase containing the RNA was transferred into a new tube and 500 μl of isopropanol added to precipitate the RNA. The samples were incubated for 10 minutes at room temperature and centrifuged at 12000 x g for 10 minutes at 4°C. The supernatant was discarded and the RNA pellet washed by adding 1 ml of ethanol, vortexing, and centrifuging at 7500 x g for 5 minutes at 4°C. The ethanol was discarded and the RNA pellet air dried for 5 minutes. The RNA was dissolved in DEPC water and the concentration measured using a NanoDrop1000 Spectrophotometer (Thermo Scientific, FL, USA). The A260/A280 ratio was < 1.6.

The Superscript III kit (Invitrogen, NY, USA) was used to perform first-strand cDNA synthesis. The reaction mixture listed in Appendix Table A.4 was used for the first-step of cDNA synthesis. Following incubation for 5 minutes at 65°C the first-step reaction mixture was added to the second-step cDNA synthesis mix (Appendix table A.5). The thermal cycling parameters for second-step cDNA synthesis are listed in Appendix.
Table A.6. Following cDNA synthesis, RNase H was added to each sample and incubation performed for 20 min at 37°C. cDNA concentration was measured using a NanoDrop1000 Spectrophotometer (Thermo Scientific, FL, USA) and all samples were diluted to a concentration of 50 ng/μl.

RT-PCR was performed using the GeneAmp PCR System 9600 thermocycler (Perkin–Elmer, CT, USA). Primer pairs for Kcnn3 and β actin (Appendix Table A.7) were designed to amplify unique DNA sequences of approximately 150 bp within the exons of each gene using the Ventor NTI software (Invitrogen, NY, USA). The primers spanned intronic boundaries in order to avoid amplification of contaminating genomic DNA. The SYBR® Green PCR Master Mix was used for RT-PCR. The components of the reaction mixture and the thermal cycling conditions are outlined in Appendix Tables A.8 and A.9 respectively. All RT-PCR reactions were performed in triplicate.

The $2^{-\Delta\Delta CT}$ method was used to calculate Kcnn3 expression in SK3$^{+/T}$ and SK3$^{+/T}$ mice relative to WT control mice. The average cycle threshold ($C_T$) was calculated for each triplicate sample. Analysis was performed by initially normalising the $C_T$ for each chamber in SK3$^{+/T}$ and SK3$^{+/T}$ mice to the internal control and subsequently comparing it to the $C_T$ for the corresponding cardiac chamber in WT mice. The β-actin gene was used as an internal control. Data were presented as a fold change in Kcnn3 expression in SK3$^{+/T}$ and SK3$^{+/T}$ mice relative to WT mice. An example of the calculation of $2^{-\Delta\Delta CT}$ for the left ventricle (LV) in SK3$^{+/T}$ mice is demonstrated below:
\[ \Delta C_T (SK3^{T/T}) = C_T (SK3^{T/T} \text{LV } Kcnn3) - C_T (SK3^{T/T} \text{LV } \beta\text{-actin}) \]

\[ \Delta C_T (WT) = C_T (WT \text{LV } Kcnn3) - C_T (WT \text{LV } \beta\text{-actin}) \]

\[ \Delta \Delta CT = \Delta C_T (SK3^{T/T}) - \Delta C_T (WT) \]

Fold difference = \( 2^{\Delta \Delta CT} \)

### 2.1.8 Histological staining and immunohistochemistry

The procedure for euthanasia and tissue collection is described in Section 2.1.6. The explanted hearts were placed in 10 % neutral buffered formalin for 24 hours for formalin fixation. Dehydration of the samples was performed by sequentially incubating in increasing concentrations of ethanol (30 %, 50 %, 70 %, 80 %, 90 %, 95 %, and 100 %) for two hours at a time. The samples were then sequentially transferred though a 50:50 solution of ethanol and toluene, 100 % toluene, and a 50:50 mixture of toluene and paraffin, again for two hours per solution. The sample block was incubated at 56-58°C in melted paraffin 2-3 hours to embed in paraffin.

A microtome was used to cut 5 micron thick slices. The sections were performed in the same area in \( SK3^{T/T} \) and WT hearts. The tissue sections were mounted onto poly-L-lysine coated slides and the slides dried for 45 minutes at 60°C in an oven. The slides were subsequently deparaffinised in xylene (three washes for five minutes at a time) and rehydrated using an ethanol gradient (100 %, 95 % and 80 % ethanol washes for three minutes per wash).
2.1.8.1 Haematoxylin Eosin staining

The tissue sections were incubated in Harris haematoxylin solution (Sigma-Aldrich, SL, USA) for 8 minutes and washed in running water. 1 % acid alcohol (1 ml HCl in 100 ml ethanol) was added for 30 seconds for differentiation. 0.2 % ammonia water was added for 30 seconds for bluing. The section was washed in water and rinsed in 95 % ethanol. Eosin-phloxine B solution was added for 30 seconds to counterstain. The section was dehydrated though four changes of ethanol (95 %, 95 %, 100 % and 100 % - 5 minutes each) and cleared in three changes of xylene (5 minutes each). The sections were subsequently mounted and imaged on a Leica DMIRB fluorescent microscope with an x40 water lens.

2.1.8.2 Staining for fibrosis

Fibrosis staining was performed using the Sigma Accustain Trichrome Stain Kit (Sigma-Aldrich, SL, USA) according to manufacturer’s instructions. Briefly, the deparaffinised and rehydrated sections were re-fixed in Bouin’s solution (at room temperature overnight). The sections were sequentially stained with Weigert’s iron haematoxylin working solution and subsequently in Biebrich scarlet-acid fuchsin solution for 5 minutes. Phosphomolybdic-phosphotungstic acid solution was added for 10 minutes. Distilled water was used to wash slides after each step. The sections were then stained in aniline blue solution for 5 minutes, rinsed and differentiated in 1 % acetic acid solution for 5 minutes. The sections were dehydrated, cleared and mounted as
Chapter 2. Methods

described above. Slides were imaged using a fluorescence microscope with an x40 water lens (Leica DMIRB).

2.1.8.3 Immunohistochemistry

Heat-induced antigen retrieval was performed by heating to 95-100°C in 1 mM EDTA (pH 8.0) for 8 min. After cooling, endogenous peroxidases were blocked by incubation with 1 % H₂O₂ in TBS (50 mM Tris, 150 mM NaCl, 0.05 %) for 20 min followed by rinses in dH₂O (2 x 5 min) and TBS (2 x 5 min). The sections were blocked for 10 minutes in 1 % bovine serum albumin (BSA) in TBS. The sections were incubated for 2 hours in connexin-43 primary antibody (1:1000, Abcam, MA, USA) at room temperature. The sections were washed in TBS (3 times for 5 minutes) and subsequently incubated for 1 hour at room temperature in a HRP-linked secondary antibody (1:100, Abcam, MA, USA). Following incubation, the sections were again washed in TBS as above. Connexin-43 was visualised using a DAB kit (Vector Laboratories, CA, USA) following manufacturer’s instructions. Slides were counterstained for nuclei with Mayer’s haematoxylin solution (Sigma-Aldrich, SL, USA) and imaged using a fluorescent microscope with a x40 water lens (Leica DMIRB).

2.1.9 Echocardiography

Echocardiograms were performed in the absence of anaesthesia in order to avoid the confounding effects of anaesthesia on heart rate, contraction and autonomic reflex control. During the three days preceding the echocardiogram, mice were trained by
Chapter 2. Methods

simulating the procedure on at least two separate occasions. Gentle manual restraint was used to hold the mouse during the procedure as previously described. The mouse was picked up by the nape of the neck and held firmly in the supine position in the palm of one hand.

On the day of the procedure, hair was removed from the thoracic area and ultrasound transmission gel was applied to the precordium. The mouse was placed in a left lateral decubitus position and the transducer placed on the left hemithorax. A GE Vivid 5 with a 15L8 15 MHz linear array transducer at a frame rate of 100 frames/sec was used to perform echocardiography on 3-month-old mice. The echocardiography machine allows high-resolution imaging with 0.25 mm axial resolution and 0.35 mm lateral resolution. Data acquisition was performed in real time and analysis performed off-line.

Echocardiographic imaging of the heart was performed using a standard protocol. The heart was first imaged in the two-dimensional parasternal long axis view. The transducer was subsequently rotated clockwise approximately 45° to obtain a parasternal short axis view at the level of the mitral valve chordae. In this view, a cursor for MMode was aligned perpendicular to the left ventricular posterior wall and the interventricular septum. MMode images were recorded in order to measure left ventricular chamber dimensions.
Chapter 2. Methods

During off-line image analysis, the following left ventricular chamber dimensions were measured; left ventricular end diastolic dimension (LVEDD), left ventricular end systolic dimension (LVESD), left ventricular posterior wall dimension in diastole (PWD), and interventricular septal dimension in diastole (IVSD). End diastole was defined as the maximum dimension in diastole. End systole was defined as the most anterior systolic excursion of the posterior wall. In accordance with the guidelines from the American Society of Echocardiography, all measurements were performed from the leading edge to the leading edge. 226

Left ventricular fractional shortening (FS) and ejection fraction (EF) were calculated as follows:

\[
FS (\%) = \left( \frac{LVEDD - LVESD}{LVEDD} \right) \times 100
\]

\[
EF (\%) = \left( \frac{LV \text{ end diastolic volume} - LV \text{ end systolic volume}}{LV \text{ end diastolic volume}} \right) \times 100
\]

### 2.1.10 In vivo electrophysiology studies

Mice were anaesthetised with isoflurane to minimise discomfort and stress during the procedure. The procedure for induction and maintenance of anaesthesia is described in detail in Section 2.1.2. A rectal temperature-sensing probe (Harvard Apparatus, MA, USA) was used to monitor the core temperature of the mouse. A
homoeothermic blanket was placed over the mouse in order to maintain the core temperature between 36-38°C.

The technique for performing a tracheostomy has previously been described by Moldestad et al. Briefly, the mouse was placed in a supine position and secured to a platform. The head was tethered to the platform using adhesive tape. A skin incision was made from the lower jaw to the thoracic aperture. Under an operating microscope (Olympus SZ61, PA, USA), blunt dissection was used to separate the lobes of the thyroid gland and the sternohyoid muscles to expose the trachea. A tracheostomy was made by performing an incision between two tracheal rings. A tracheal tube fashioned from a 22GA angiocath (Hamilton Medical Products, NV, USA) was inserted through the tracheotomy and connected to a small animal ventilator (Minivent type 845, Harvard Apparatus, MA, USA). Correct positioning of the tube was confirmed by symmetrical chest expansion. The tube was subsequently secured in place using a silk surgical suture around the trachea. Ventilation parameters were maintained at 130 breaths/min with a tidal volume of 100 µl. The ventilator was connected to the isoflurane vaporiser driven by an oxygen source with a flow meter and to a scavenging unit.

A jugular cutdown was used to gain access to the right internal jugular vein (Figure 21). The section of the jugular vein cranial to the clavicle (approximately 5mm) was mobilised using blunt dissection. A small incision was made in the jugular vein and an
octapolar catheter (Millar Instruments, TX, USA) was inserted into the vein and
advanced such that the distal electrodes were in the right ventricle while the proximal
electrodes were in the right atrium. The ideal catheter position was determined by the
intracardiac electrograms and pacing thresholds. The catheter had an electrode length
of 0.5 mm with an interelectrode distance of 0.5 mm. A ligature was tied over the
catheter to achieve haemostasis and to secure the catheter in place.

![Figure 21: Photographs demonstrating the electrophysiology catheter and surgical field. A) View of the jugular dissection and electrophysiology catheter. B) Close-up view of jugular cutdown with distal and proximal sutures. From: Sastry et al. JACC 2006;48(3):559-65.](image)

The catheter was connected to a data acquisition system and a stimulator. Bipolar
electrograms were obtained from the right atrium, atroventricular junction and the
right ventricle. 27-gauge needle ECG electrodes were inserted subcutaneously in the
lead I and II positions to allow for simultaneous ECG recording. The intracardiac
electrograms and surface ECG signals were amplified and filtered using an EVR OctalBioamp recorder (Harvard Apparatus, MA, USA). Sampling of the signals was performed at 2 kHz and the signals were filtered and amplified at 30-500 Hz. The AD systems data acquisition system (LabChart, AD Instruments, CO, USA) was used for display and analysis of the electrograms.

A Medtronic stimulator was used to perform programmed stimulation. The stimulator was modified by manufacturers to allow for coupling intervals as short as 10 ms. The diastolic atrial and ventricular pacing thresholds were measured as the lowest output voltage that resulted in 1:1 capture. Programmed stimulation was performed with 1 ms pulse widths at twice the diastolic threshold. In both the atrium and the ventricle, the distal pair of electrodes was used to perform stimulation while the proximal pair was used to record the electrograms.

The extrastimulus method was used to measure the AV node effective refractory period (AVERP) and the effective refractory period of atrial (AERP) and ventricular tissue (VERP). Atrial and ventricular premature extrastimuli (S2) were delivered after drive trains of eight stimuli at 100 ms (S1). The S1-S2 coupling intervals were progressively reduced by 5-10 ms until the tissue became refractory. A more precise measurement of the ERP was subsequently made by reducing the S1-S2 coupling intervals by 1 ms.
AVERP was defined as the longest S1-S2 interval that failed to conduct to the ventricle from the atrium. AERP was defined as the longest S1-S2 interval that failed to capture the atrium. VERP was defined as the longest S1-S2 interval that failed to capture the ventricle.

The Wenckebach cycle length was measured by progressively faster atrial pacing rates. Atrial pacing was performed for five second periods at progressively higher rates until first degree Wenckebach heart block developed. The Wenckebach cycle length was defined as the minimum cycle length that maintained 1:1 AV conduction.\textsuperscript{229}

SA node function was measured indirectly by measuring the sinus node recovery period (SNRP) following 30 ms of pacing at three cycle lengths (120 ms, 100 ms and 80 ms). The SNRP was defined as the interval between the last paced P wave and the first return sinus P wave. The maximum sinus return cycle from the three drive trains was used to calculate the corrected SNRP (cSNRP). cSNRP was calculated in order to correct for the sinus cycle length. cSNRP was calculated by dividing the absolute SNRP by the sinus cycle length (SNRP/sinus cycle length).\textsuperscript{229}

Burst pacing and single, double and triple extrastimuli were used to assess for inducibility of atrial and ventricular arrhythmias. Atrial burst pacing was performed with eight 50 ms and four 30 ms burst episodes up to a total burst pacing duration of
Chapter 2. Methods

one minute. Ventricular burst pacing was performed with eight 70 ms burst episodes up to a total burst pacing duration of one minute. Programmed induction of arrhythmias was performed with single, double and triple extrastimuli up to a minimum coupling interval of extrastimuli of 10 ms (in both the atrium and the ventricle). Immediately after data was acquired, animals were euthanised by overdose of sodium pentobarbital injection. Offline analysis was performed using the LabChart software to measure the intracardiac intervals.

2.1.11 High resolution optical mapping

A Langendorff perfusion apparatus was used to perfuse isolated mouse hearts according to previously published methods (Figure 22). The apparatus is designed to keep excised hearts alive though retrograde perfusion of the aorta with an oxygenated physiological salt solution. In addition to the salt solution, the heart was perfused with Blebbistatin (10 μM, Tocris Bioscience, MO, USA) and di-4-ANEPPs (Invitrogen, NY, USA). Blebbistatin causes electro-mechanical uncoupling of the isolated heart and therefore eliminates movement artefacts due to cardiac muscle contraction. di-4-ANEPPs is a membrane-bound dye that allows tracking of changes in cardiac membrane potential by emitting florescence.

High-resolution optical mapping of the Langendorff perfused hearts was performed according to previously published protocols. The main components of the optical mapping setup consist of a light source and an optical sensor (Figure 23). An excitation
beam from the light source is focused onto the epicardial surface of the heart. The interaction between the excitation beam and di-4-ANEPPs results in emission of florescence, which is collected by the optical sensor. The florescence is proportional to the transmembrane voltage and is used to derive a map of action potentials and propagation of electrical activity on the epicardial surface of the heart. Activation and repolarisation patterns are presented as isochronal maps.

Figure 22: Schematic representation Langendorff apparatus for perfusion of isolated mouse hearts. The perfusion solution is delivered via an aortic cannula at a constant flow rate. The perfusion solution is delivered from a reservoir which is provided with oxygen and carbon dioxide. The reservoir is water jacketed to maintain constant temperature. Devices are connected to a computer which allows recording and analysis of data. From: Skrzypiec-Spring et al. Journal of Pharmacological and Toxicological Methods. 2007;55:113-126.
Mice were anaesthetised with isoflurane to minimise discomfort and stress during the procedure. The procedure for induction and maintenance of anaesthesia is described in detail in Section 2.1.2. Once adequate anaesthesia had been achieved, euthanasia was performed by a sodium pentobarbital overdose, and the heart was rapidly excised (described in detail in Section 2.1.6). The explanted heart was bathed in an arrest solution in order to limit ischaemia (118.0 mM NaCl, 18.8 mM KCl, 1.2 mM MgSO$_4$, 2.55 mM CaCl$_2$, 24.9 mM NaHCO$_3$, 1.2 mM KH$_2$PO$_4$, 11.1 mM glucose, 1000 u heparin, pH = 7.35±0.05). The aortic root was slipped over an aortic cannula using forceps and secured using a silk surgical suture.

The cannulated heart was transferred to the Langendorff perfusion apparatus and retrogradely perfused with Krebs-Henseleit solution (118.0 mM NaCl, 4.7 mM KCl, 1.2 mM MgSO$_4$, 2.55 mM CaCl$_2$, 24.9 mM NaHCO$_3$, 1.2 mM KH$_2$PO$_4$, 11.1 mM glucose, pH=7.35±0.05), Blebbistatin and di-4-ANEPPs. The temperature of the perfusate was maintained at 37°C by a water-jacketed system. A sintered glass bubbling device with a high oxygen concentration (95 % O$_2$ and 5 % CO$_2$) was used to gas the perfusion solution. The perfusate was delivered at a constant flow rate (2 ml/min) sufficient to achieve 90 mmHg of perfusion pressure. A rotameter was connected to the inflow line proximal to the aortic cannula to monitor coronary flow of the perfusate. Perfusion was continued for a period of 30 minutes in order to motion arrest and stain the heart.
**Figure 23**: Components of an optical mapping setup. The isolated heart is depicted on the left, the system of optics in the centre, and the detector on the right. Selected wavelengths of light from the light source are allowed to pass through the excitation filter. The dichroic mirror directs light towards the heart. The voltage-sensitive dye molecules fluoresce in response to the excitation light. The degree of fluorescence is proportional to the membrane voltage. The fluorescent light from the heart passes through the dichroic mirror as it has a longer wavelength. The light is focused onto a detector. From: Rosenbaum, D. Optical Mapping of Cardiac Excitation and Arrhythmias: A Primer. In Rosenbaum D, Jalife J, ed. Optical Mapping of Cardiac Excitation and Arrhythmias. Oxford: Blackwell Futura Publishing; 2001: 2-7.234

Once the heart was adequately motion arrested and stained, three custom-made platinum electrodes were attached to the epicardial surface of the ventricle. One electrode was used to pace the heart from the left ventricular apex while the other two were used as sensing electrodes. The cardiac rhythm was monitored throughout the experiment. A Medtronic stimulator was used to pace the heart. The diastolic ventricular pacing threshold was measured as the lowest output voltage that resulted
in 1:1 capture of the ventricle. Pacing was performed with 1 ms pulse widths at twice the diastolic threshold at a cycle length of 150 ms. The epicardial electrograms were amplified and filtered using an EVR OctalBioamp recorder (Harvard Apparatus, MA, USA). Sampling of the signals was performed at 2 kHz and the signals were filtered and amplified at 30-500 Hz. The LabChart data acquisition system (AD Instruments, CO, USA) was used for display and analysis of the electrograms.

A halogen light source (exfo excite 150 W, filtered at 520 ± 45 nm) was used to excite fluorescence. Emissions above 610 nm were collected and focused onto an 80 × 80 CCD camera (RedShirt Imaging SMQ Camera and Macrooscope IIA) using a 50 mm 2.7x lens (NA 0.9). Data sampling was performed at 2,000 frames/second with a filter setting of 1 kHz. The fluorescence signals were amplified, digitised, processed, and analysed to convert them into electrical signals.

A specifically designed program written in Matlab (Mathworks, Cambridge, UK) was used to perform off-line data analysis. Fluorescence signals from individual pixels on the epicardial surface of the heart were analysed. After subtracting background fluorescence levels, positive changes in fluorescence were correlated to depolarisation of the membrane potential. APD$_{80}$, the time taken to 80 % repolarisation was calculated from the corresponding times of fluorescence decay. Composite data from the all the pixels was used to derive conduction velocity, APD dispersion and to generate isochronal maps. All data were averaged over 20 cycles.
2.1.12 Implantable loop recording

Continuous ambulatory monitoring was performed in one-month-old SK3<sup>T/T</sup> mice and WT controls. The procedure was performed using previously established techniques for implantation of telemetry devices.<sup>235</sup> Mice were anaesthetised with isoflurane to minimise discomfort and stress during the procedure. The procedure for induction and maintenance of anaesthesia is described in detail in Section 2.1.2. Mice were continuously monitored during the implantation procedure. Core temperature was monitored using a rectal temperature-sensing probe (Harvard Apparatus, MA, USA) and was controlled between 36-38°C using a homoeothermic blanket.

The device was implanted into the abdominal cavity under sterile conditions. Local anaesthetic was applied to the implant site (2 % lidocaine, MP Biomedicals, CA, USA). A 2 cm midline abdominal incision was made and a subcutaneous pocket was fashioned in the lateral abdominal wall using blunt dissection. An ETA-F10 monitor (Data Sciences International, MN, USA) was implanted in the subcutaneous abdominal pocket. Telemeter leads were tunnelled and anchored to the pectoral muscle in a modified lead II position using a nonabsorbable suture (Figure 24). The subcutaneous pocket was closed, also using a nonabsorbable suture.
Figure 2. Methods

Figure 24: Schematic representation of implantation of telemetry device with leads in a modified lead II position. From: DSI surgical implantation manual. http://www.datasci.com

Buprenorphine was administered subcutaneously (0.1 mg/kg) immediately postoperatively and every 8 hours for the first 72 hours. The animals were monitored for signs of pain and when necessary, further doses of analgesia were administered. Data collection was initiated by placing a magnet over the subcutaneous device.

Telemetry data was recorded continuously via a receiver placed under the mouse cage. The ECG signal was sampled continuously at a rate of 2000 Hz. As illustrated in Figure 25, the receivers were connected to a data exchange matrix and to a computer where the data was stored. The Dataquest ART software (Data Sciences International, MN, USA) was used to analyse the raw ECG traces.
2.1.13 Statistical analysis

Data analysis for *in vivo* electrophysiology data and echocardiographic data was performed using Excel (Microsoft; Redmond, WA, USA). Data were presented as mean values ± standard deviation. A 2-tailed *t* test was used to perform comparisons between groups. A *p* value of < 0.05 was considered to be statistically significant. Chi square analysis was used to determine whether *SK3*T mice follow Mendelian inheritance ratios. In order to estimate the incidence of sudden death, Kaplan-Meier survival analysis was performed using SAS software (SAS Institute Inc., NC, USA).
Chapter 2. Methods

Mean heart rate data was collected using the Dataquest ART analysis software (version 1.10, Data Sciences International, MN, USA). For analysis of PR and RR intervals at 3 months of age, raw waveform data was exported and analysed using ChartPro software (ADinstruments, CO, USA). Final data analysis was performed using SigmaPlot 4.1 (SPSS, IL, USA) and Stata 12 (StataCorp LP, TX, USA).

2.2 Aim 2 - Identify causative mutations for AF using linkage analysis and exome sequencing.

2.2.1 Clinical evaluation

Two AF pedigrees were evaluated in the present study. The first pedigree (AF-325) spans three generations and has an autosomal dominant pattern of inheritance of AF. The proband was initially enrolled into the Women’s Health Study in 2009. Between 2009 and 2012, a total of 57 members from family AF-325 were recruited. Nine family members had AF. The second pedigree (AF-435) is a three-generation family with an autosomal dominant pattern of inheritance of a complex phenotype of AF and atrioventricular septal defects. Between 2010 and 2012, a total of 17 individuals from family AF-435 were recruited. Four of the family members were affected. The study was approved by the Institutional Review Board at Massachusetts General Hospital. Written informed consent was obtained from all study participants. Members of the
family were evaluated by taking a detailed medical history and performing physical
examination, ECG and echocardiography. For patients who had previously undergone
investigations relevant to the phenotype, previous records were obtained from the
primary care physician and/or the investigating cardiologist.

2.2.2 12-lead electrocardiograms

Standard 12-lead electrocardiograms were performed at 10 mm/mV and a speed of
25 mm/s (Hewlett-Packard Corporation, CA, USA). The chest and limb leads were
placed in the standard positions. Manual measurements of the PR interval, QRS
duration, RR interval, QT interval and the P and R wave axis were made. The presence
of interventricular conduction delay (right bundle branch block or left bundle branch
block) and abnormalities in the ST segments or T waves were also recorded.

2.2.3 Echocardiography

Echocardiographic analysis was performed using a 2.5 MHz transducer connected to a
Vivid-7 ultrasound machine (GE Healthcare, PA, USA). Echocardiography was
performed with the patients lying in the left decubitus position. Images were acquired
in the four standard views; parasternal long axis, parasternal short axis (at the level of
the papillary muscles), apical four chamber, and apical two chamber. Pulsed wave
Doppler, continuous wave Doppler and colour flow mapping were used for analysis of
valve function and atrioventricular septal defects.
Offline analysis of echocardiographic data was performed using a dedicated workstation using EchoPAC software (GE Healthcare, PA, USA). The following dimensions were measured; left atrial dimension, left ventricular end diastolic dimension (LVEDD), left ventricular end systolic dimension (LVESD) and right ventricular end diastolic dimension at the level of the tricuspid annulus. Forward flow velocities through all four valves were measured using continuous wave Doppler in order to derive pressure gradients across the valves. Potential atrial or ventricular septal defects were detected using colour flow mapping and continuous wave Doppler to measure flow velocity across the defects.

Left ventricular systolic function was measured as fractional shortening (FS) and ejection fraction (EF) using the following formulas:

\[
FS (\%) = \frac{LVEDD - LVESD}{LVEDD} \times 100
\]

\[
EF (\%) = \frac{(LV \text{ end diastolic volume} - LV \text{ end systolic volume})}{LV \text{ end diastolic volume}} \times 100
\]

2.2.4 DNA extraction

Peripheral venous blood samples were obtained and stored in vacutainers containing EDTA (BD, NJ, USA). DNA was extracted from 10 ml of peripheral blood using the Qiagen DNA extraction kit (Qiagen, CA, USA) according to manufacturer’s instructions.
Chapter 2. Methods

Briefly, 30 ml of red cell lysis solution was added to a 10 ml sample of whole blood. The mixture was allowed to stand at room temperature for 5 minutes and subsequently centrifuged for 2 minutes at 2000 x g to form a white cell pellet. The supernatant was discarded, 10 ml of cell lysis solution was added, and the mixture was allowed to stand at room temperature overnight. 50 µl of RNAse A was added in order to digest RNA. Following an incubation period of 15 minutes with RNAse A, the sample was cooled for 3 minutes on ice. 3.3 ml of protein precipitation solution was added and the sample vortexed at high speed for 20 seconds. Centrifugation was performed for 5 minutes at 2000 x g to pellet the precipitated protein. The supernatant containing the DNA was transferred to a fresh tube containing 10 ml of isopropanol. Centrifugation was performed for 2 minutes at 2000 x g to pellet the DNA. The supernatant was discarded and the DNA pellet washed with 10 ml of 70 % ethanol. Centrifugation was performed for 1 minute to re-pellet the DNA and the ethanol supernatant discarded. The pellet was air dried and resuspended in 1 ml of DNA hydration solution.

DNA concentration was measured using a Nanodrop spectrophotometer (Thermo Scientific, FL, USA) at the absorbance ratio $A_{260}/A_{280}$ according to manufacturer’s instructions. The quality of the DNA samples was analysed by running 1 µg of DNA in a 1 % agarose gel. The DNA samples were diluted to a concentration of 25 ng/µl in deionised water and stored at 4°C.
2.2.5 Candidate-gene screening in pedigree AF-435

Candidate-gene screening of four transcription factor genes previously reported to cause atrial and ventricular septal defects (\textit{GATA4}, \textit{NKX-2.5}, \textit{TBX5} and \textit{TBX20})\textsuperscript{238} was performed.

PCR was used to amplify all exons and flanking introns of the four candidate genes. Larger exons were divided into two or three amplicons. The average size of each amplicon was 200-300 bp. Primers were designed using the genomic sequences from the University of California Santa Cruz (UCSC) website (hg19, http://genome.ucsc.edu/) and the ExonPrimer website (http://ihg.gsf.de/ihg/ExonPrimer). The primer sequences are summarised in Appendix Table A.10. Synthesis of the primers was performed by the MGH sequencing core facility.

PCR products were amplified in a reaction volume of 25 µl. 50 ng of DNA (25 ng/µl) was added to a PCR reaction mixture consisting of 13.8 µl of deionised water, 1 µl of dimethyl sulfoxide (DMSO), 2.5 µl of 10x buffer (200 mM Tris-HCl, (pH 8.4), 500 mM KCl, Invitrogen, NY, USA), 0.75 µl of both forward and reverse primers (100 nm) and 0.2 µl of Taq DNA polymerase (Invitrogen, NY, USA).

PCR was performed for 35 cycles using a Biorad S1000 PCR thermal cycler (Biorad, CA, USA). The cycling protocol for PCR amplification is summarised in Appendix Table
A.11. The optimal annealing temperature for each set of primers was determined by performing PCR with an annealing temperature gradient (50-60°C).

Analysis of the quality of the amplification product was performed using gel electrophoresis in a 2% agarose gel in TAE buffer (40 mM Tris-OH, 20 mM Acetic Acid, pH 7.8). The gel contained ethidium bromide (1 µg/ml) to allow visualisation of DNA. 1 µl of the DNA product was mixed with 4 µl of loading buffer (50 mM EDTA, 0.2% SDS, 50% glycerol, 0.05% w/v bromphenol blue) and run on the gel for 30 minutes in TAE buffer with a voltage of 120 mV. A 1 kb ladder was run for reference (Biorad, CA, USA). The DNA was visualised under a UV transilluminator (Biorad, CA, USA).

Sanger sequencing of the PCR product was performed by the MGH sequencing core facility. The BigDye Terminator v3.1 Cycle Sequencing Kit (PE-Applied Biosystems, NY, USA) was used for bidirectional sequencing according to manufacturer’s instructions. Briefly, successive rounds of denaturation, annealing, and extension were performed in a thermal cycler resulting in linear amplification of extension products (Figure 26). The reaction mixture consisted of the DNA template, a sequencing primer, DNA Polymerase, the four dNTPs and four fluorescently labelled 2',3'-dideoxynucleotides (ddNTPs). ddNTPs are designed to terminate chain elongation. Therefore, the extension step produces a series of DNA fragments which terminate at different labelled bases. Following heat denaturation, samples were loaded onto a sequencer.
Chapter 2. Methods

Analysis of the sequencing product was performed using an automated DNA sequencer.

The raw sequencing data from the MGH DNA core website was analysed using the Contig Express software (Invitrogen, NY, USA). The software assembles fragments of DNA into contiguous sequences and allows comparison with a reference sequence. The genomic sequence from the UCSC genome browser was used as the reference sequence (hg19, http://genome.ucsc.edu/). The effect of the mutation on the protein coding sequence was analysed using the Vector NTI software (Invitrogen, NY, USA).


2.2.6 Exome sequencing

Exome sequencing was performed in one individual from pedigree AF-325 (IV-1) and three individuals from pedigree AF-435 (III-5, III-15 and III-16) by Perkin Elmer
Chapter 2. Methods

sequencing (Waltham, MA, USA). The following section contains a brief background on exome sequencing followed by a description of the methods.

Exome sequencing is a technique that utilises massively parallel sequencing technology to perform targeted sequencing of the coding region of the genome. The subsegment of the genome that constitutes the exome is captured from fragmented DNA using a series of specifically designed probes. There are two main capture methods used in exome sequencing; solid phase hybridisation and liquid phase hybridisation. Solid-phase hybridisation involves the use of microarrays or filters. These solid platforms have the exome-specific DNA probes affixed to them. Liquid-phase hybridisation on the other hand involves the use of biotinylated probes rather than probes attached to a solid matrix. Liquid phase hybridisation was used in the present study and is the focus of the following discussion.

The technique for liquid-phase exome capture involves fragmentation of total human DNA followed by ligation with sequencing adapters. The adapters are designed to allow for PCR amplification with universal primers at a later stage. The fragmented, adapter-ligated DNA, referred to as a ‘pond’ of DNA, is mixed with biotinylated cRNA probes, referred to as the ‘bait’. The target DNA that has hybridised to the complementary probes, also referred to as the ‘catch’, is pulled down onto streptavidin-coated magnetic beads (Figure 27). Following hybridisation, non-targeted DNA sequences are washed away. Universal primers are used to amplify the
targeted sequences. The enriched target DNA is subsequently eluted and sequenced using next-generation sequencing technology.

**Figure 27:** Pictorial representation of liquid-phase exome capture. The bait probes are depicted in light blue and black. The probes are biotinylated (asterisk). Following fragmentation, the bait probes hybridise to the genomic sequences of interest. Streptavidin beads are added to allow physical separation. Following washing, the DNA of interest is eluted. Adapted from: Teer JK et al. Hum Mol Genet. 2010;19(R2):R145-S1.240

The basic principle of next-generation sequencing technology is sequencing-by-synthesis.242 The first step before sequencing involves amplification of the template DNA. The templates are covalently bound to glass surfaces and amplified using universal primers. This process results in clonal amplification of the original DNA
Chapter 2. Methods

template resulting in the formation of clusters of more than 1000 identical copies. In the case of the Illumina sequencer, bridge amplification is used to generate clusters of identical single-stranded DNA templates for the sequencing step (Figure 28). Sequencing-by-synthesis involves simultaneous 'reading' of the clusters generated in the previous step. Paired-end sequencing, which involves sequencing of approximately 100 bases from each end of a DNA fragment, is typically used to sequence the DNA. The sequencing primers anneal to adapters on opposite ends of the DNA fragments. The nucleotides used for the sequencing step are bound to fluorescently labelled bases. The fluorescent labels are unique for each base. As the nucleotides are incorporated during synthesis, excitation of the labels using a light source results in emission of fluorescence which is captured by a charge-coupled device (Figure 28).

Fluorescence intensity data is used to infer the nucleotide sequence of the short reads. A quality score is usually assigned for each base call using base-calling algorithms. A number of different software programs can be used for base calling. The programs are designed to correct for potential sources of base calling errors associated with the Illumina sequencer including mixing of clusters, phasing (which may arise when strands either lag behind or jump ahead in a single cycle), and decay of signal intensity due to improper chemistry and imperfect optical sensors.
Figure 28: Illumina HiSeq 2000 sequencing steps. The DNA ‘reads’ are immobilised onto the surface of the flowcell (step 1). Bridge PCR is used to amplify the DNA template and generate clusters of templates (steps 2 and 3). Cyclic reversible termination is performed using fluorophore bound deoxynucleotides (steps 4, 5 and 6). Imaging of the array is performed and four colour sequences are used to decode the sequences. Reproduced with permission from: http://www.eurofinsdna.com/244
Chapter 2. Methods

Analysis of the raw exome sequencing reads is divided into two stages. The first stage involves aligning the sequencing data to the reference human genome. A number of different tools, such as the Burrows-Wheeler Aligner (BWA), are available for alignment.\textsuperscript{245} A certain number of mismatches is usually allowed during alignment as overly stringent alignment protocols may compromise read depth.\textsuperscript{242} The alignment data is commonly produced in a standard Sequence Alignment/Map (SAM) format. SAM is a generic format for storing read alignment data against reference sequences. The format is used by other software packages for post processing of the alignment data.\textsuperscript{246}

Accurate alignment is critical in order to identify genetic variants. The second stage of analysis is the post-alignment identification or ‘calling’ of genetic variants using calling algorithms.\textsuperscript{245} Downstream analysis is usually performed using software programmes such as The Genome Analysis Toolkit (GATK) and SAMtools. These software packages allow manipulation and parsing of alignments. They can be used to generate per base position information in a pileup format and allow calling of SNPs and insertion-deletion variants.\textsuperscript{246} Due to the sheer volume of data, the alignment and base calling steps commonly require clusters of computers.

Exome sequencing typically identifies up to 30,000 variants per individual.\textsuperscript{222} Almost a third of these variants are nonsynonymous, disrupt splice sites, or result in insertions and deletions. Therefore, the identification of a variant that appears to be pathogenic
is not sufficient to assign causation. Comprehensive bioinformatic analysis is necessary to filter the number of candidate variants for further analysis. Given that the phenotypes under consideration in pedigree-based studies are likely to be caused by rare and highly penetrant mutations, the first level of filtering commonly involves removal of polymorphisms that have been reported in publically available databases. Using this strategy, up to 90% of variants can be excluded from downstream analysis. Further computational analysis can be used to prioritise variants based on their predicted pathogenicity. Variants that are predicted to be benign can be removed from further analysis using this strategy. The next stage of filtering usually involves analysis of segregation of candidate variants with disease. For autosomal dominant traits, the disease variant must segregate with the disease phenotype.

2.2.7 Target enrichment and sequencing

The Illumina genomic sample preparation kit was used to prepare a DNA fragment library at Perkin Elmer sequencing (Waltham, MA, USA). Purified genomic DNA from peripheral blood was randomly sheared into small fragments. The shearing step yielded DNA fragments of approximately 250 base pairs. The DNA fragment ends were repaired by removing 3’ overhangs and filling 5’ overhangs. Nontemplated terminal A residues were subsequently added to the DNA fragments. The DNA was ligated to standard sequencing adapters (single reads) for the Illumina sequencer. In order to confirm successful adapter ligation and to add custom synthesised barcode sequences
(which act as sample signatures for subsequent steps), PCR amplification was performed for 4 cycles using a high fidelity polymerase and custom synthesised primers. A further 12 cycles of PCR was performed using standard Illumina PCR primers to increase the yield of the ‘pond’ of genomic DNA.

DNA from the adapter ligation reaction with unique barcode sequences was denatured by heating and was subsequently hybridised to biotinylated cRNA baits for target enrichment (Agilent SureSelect XT target enrichment system V4). The enrichment and capture process is summarised in Figure 29. The Aligent target enrichment system is designed to target 20,965 genes and 334,378 exons. The approximate bait length was 200 bases. Streptavidin Dynabeads were added to the hybridisation mixture. The beads were pulled down and washed after which the DNA was eluted. The eluted DNA was transferred to the amplification step.

An Illumina HiSeq 2000 platform was used to amplify and sequence the captured library. The DNA fragments were washed though the lanes of the sequencer where they covalently bonded to the glass surface. Three patient samples, each with unique barcodes, were run per lane. Amplification was performed using bridge PCR. The standard Illumina PCR primers were mixed with the template. Both the template and the primers were immobilised on the glass surface of the flow cell. The magnitude of enrichment of the PCR product was measured using the Aligent 2000 Bioanalyzer (Aligent, CA, USA).
The clusters were sequenced using cyclic reversible termination. Sequencing primers designed to anneal to the adapter sequences on either side of the DNA fragments were used. 3' terminated and fluorescence-labelled nucleotides for all four bases (ddATP, ddGTP, ddCTP, ddTTP) were washed though the lanes of the flowcell. Red and
green lasers were used to illuminate the fluorophores and imaging was performed using a CCD camera.\textsuperscript{248} Optimal wavelengths for each of the four fluorophores were used to excite fluorescence and each lane was imaged four times. Finally, chemical cleavage was performed to remove the fluorophore and the blocking group in preparation for the next sequencing cycle. Sequencing was performed for 50 PCR cycles.

The Illumina pipeline was used to perform base calling. The FIRECREST program was used to convert raw data into digital intensity files. Further downstream, the BUSTARD program was used to convert digital intensities into sequences in three steps.\textsuperscript{243} The first step involved conversion of the digital intensities into concentration (or number of templates). The second step involved renormalisation of the concentration against an average concentration across a tile in order to minimise signal decay problems. Finally, the renormalised concentrations were converted into sequences using a Markov model. A base specific quality score was generated for each position. The quality scores varied between -40 and +40. Bases with a score of greater than +10 were considered acceptable.\textsuperscript{250} Redundant read pairs and low quality reads were excluded from further analysis.

### 2.2.8 Analysis of exome sequencing data

Reads from different samples were identified by matching the unique barcodes on the reads. The raw short reads were aligned against the human reference genome (hg19
NCBI Build 37, 2009) using the BWA software (version 0.5.8c). A maximum of four mismatches were allowed between the read and the alignment reference sequence. Reads that aligned at multiple locations of the genome were removed. The number of unique alignments at each base position was used to determine coverage. Aligned data was generated in SAM format and subsequently compressed to a binary format (BAM) in order to facilitate downstream analysis.\textsuperscript{251}

GATK and SAMtools were used for variant calling. SAMtools was used to eliminate duplicate reads and to generate pile-up style files for initial variant calling. GATK was subsequently used for advanced variant calling. Where putative insertions or deletions were identified, multiple local realignments were performed in the absence of mismatches in order to exclude alignment artefacts. The base quality score recalibrator of GATK was used to perform quality score recalibration. The aim of this step is to correct for bias introduced by base calling quality calculations which may arise due to factors such as position within the sequence read, probability of mismatching to the reference genome, and both preceding and current nucleotide calls.\textsuperscript{252}

The next step was to filter unmapped reads, taking into account the improved quality values. The GATK Unified Genotyper was used to identify SNPs and insertion deletions and also to perform quality score recalibration of variants. SNPs with a quality per depth score of less than 5 or a quality score of less than 30 were removed.\textsuperscript{253} The
minimum required coverage for a SNP to be included for further analysis was 20x. The data processing pathway is summarised in Figure 30.

![NGS Data Processing Diagram](image)

**Figure 30:** Pathway for processing of exome sequencing data. From: http://www.broadinstitute.org.  

### 2.2.9 Bioinformatic analysis

We identified in excess of 18,000 variants in each exome. Filtration and prioritisation of the variants was performed at three levels according to previously published methods. The first filtering level is the genetic level, the second is the variant-gene level and the third is based on knowledge of the function of the gene. Potential levels of filtering are summarised in Figure 31.
Chapter 2. Methods

At a genetic level, the pattern of inheritance for both pedigrees under consideration is autosomal dominant. Therefore, all individuals who are definitively affected are expected to harbour the causative mutation. On the basis of this assumption, all variants that did not segregate with disease were excluded from further analysis.

At a variants level, based on the fact that both pedigrees have Mendelian inheritance patterns, it is very unlikely that common polymorphisms encountered in the general population underlie the phenotype. Therefore, all the variants identified in publically available databases (dbSNP, 1000 Genomes Project, exome variant server) were removed from further analysis. Further, it is likely that Mendelian diseases are caused by variants that result in a change in the amino acid sequence i.e. nonsynonymous variants or insertion-deletion variants. Therefore, synonymous variants were excluded. The SIFT (version 4.0, http://sift.jcvi.org/) program was used to remove all synonymous variants.

We also performed analysis of evolutionary conservation of the altered amino acids using the Phylop tool. PolyPhen-2, and SNAP, were used to predict the effect of the amino acid substitution in protein structure. Variants affecting amino acids that are highly conserved though evolution and those predicted to have a significant effect on protein structure were prioritised for further analysis.
Chapter 2. Methods

Figure 31: A stepwise approach to filtration and prioritisation of variants identified by exome sequencing. From: Li, M.X, et al. Nucleic Acids Res; 2012;40(7):e53.255

At a knowledge level, variants were prioritised based on their expression profile and their involvement in biological functional pathways. Information on the function of identified genes and their potential involvement in functional pathways was also
obtained from resources such as the UCSC genome browser, Entrez Gene, Gene Ontology and KEGG pathway database.

The final results files annotated all the coding variants and generated files with the position of the variant, the rs ID from dbSNP, conservation scores, classification (synonymous, missense, nonsense or insertion/deletion), amino acid changes and polyphen classification.

2.2.10 Confirmation of variants

After filtering the variants identified by exome sequencing, a variant in the GATA6 gene was prioritised for further analysis. Sanger sequencing was initially performed to reconfirm the presence of the variant and subsequently to perform segregation analysis amongst the other members of the family. The protocol for confirmation of variants is identical to that outlined for candidate gene screening (outlined in Section 2.2.5). The primer sequences used for confirmation of the GATA6 R585L variant included in Appendix Table A.12.

2.2.11 Mutagenesis and cell culture

The mammalian GATA6 expression vector (pCMV6 Xl6 - GATA6, Origene, MD, USA) was used for functional studies (Figure 32). Site-directed mutagenesis was used to introduce mutations into the wild type GATA6 construct according to manufacturer’s
instructions (QuikChange kit, Stratagene, CA, USA). The method for site-directed mutagenesis is summarised in Figure 33.

**Figure 32:** pCMV6-XL6 vector design. From: http://www.origene.com/cdna/trueclone/vectors.mspx.\(^{256}\)

Mutagenic oligonucleotide primers were designed using the Vector NTI software (Invitrogen, NY, USA). Briefly, the genomic sequence from the UCSC genome browser (http://genome.ucsc.edu/) was uploaded into the Vector NTI program. The position of the mutation was annotated and the surrounding sequence of 25-45 bases used to design primers. Both the forward and reverse mutagenic primers contained the desired mutation. The primer length was adjusted to ensure a melting temperature of \(\geq 78^\circ\text{C}\). The Oligocalc website (http://www.basic.northwestern.edu/biotools/oligocalc.html) was used to estimate primer melting temperature. Oligocalc was also used to calculate likelihood of self dimerisation, secondary structure formation, and GC content of the primers. The
optimal GC content was set at 40%. The mutagenic primer sequences are included in Appendix Table A.13. The primers were synthesised by the MGH DNA core facility.

PCR with the mutant primers was used to synthesise mutant GATA6 constructs. The reaction mixture and thermal cycling conditions for the mutant strand synthesis reaction are included in Appendix tables A.14 and A.15. 1 µl of Dpn 1 restriction enzyme was subsequently added to the PCR product to digest the unmutated template for 1 hour. Transformation of DH5α cells (Invitrogen, NY, USA) was performed using a high competence protocol.257 DH5α cells (50 µl) were thawed on ice and 1 µl of the Dpn 1 treated DNA was added. The transformation reaction was incubated for 30 minutes and heat shocked for 45 seconds at 42°C. Cells were recovered in 0.5 ml of SOC media (0.5 % yeast extract, 2 % tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl2, 20 mM MgSO4, 20 mM glucose) for 1 hour in a shaker (37°C, 250 rpm). 250 µl aliquots of the transformants were plated on ampicillin-LB agar plates (100 µg/mL ampicillin, 1.5 % agar, 1 % Tryptone, 0.5 % Yeast Extract, and 0.1 % NaCl, pH 7) and incubated for 16 hours. Transformants were selected and grown for many generations in 5 ml of ampicillin-LB media (100 µg/mL ampicillin, 1 % Tryptone, 0.5 % Yeast Extract and 0.1 % NaCl, pH 7).
Figure 33: Overview of QuikChange site-directed mutagenesis method. From: QuikChange Site-Directed Mutagenesis Kit instruction manual (Catalog #200518).
The QIAprep Spin Miniprep Kit (Qiagen, CA, USA) was used to isolate plasmid DNA. Overnight cultures were centrifuged at 2000 x g for 20 minutes to form a bacterial pellet. The cells were resuspended in 250 µl of buffer P1. 250 µl of buffer P2 and 350 µl of buffer N3 were added sequentially. Centrifugation was performed at 1000 x g for 10 minutes. The supernatant was applied to a QIAprep column and vacuum suction applied. The column was washed with 500 µl of buffer PB and 750 µl of buffer PE. The DNA was eluted with 30 µl of elution buffer (10 mM Tris-Cl, pH 8.5). The concentration of DNA was measured using a NanoDrop 1000 Spectrophotometer (Thermo Scientific, FL, USA). The DNA had an absorbance peak centred at a wavelength of 260 nm with a 260/280 ratio of approximately 1.95.

Prior to performing functional experiments, the entire coding region of the mutant GATA6 clone was sequenced to ensure that there were no additional mutations. The sequencing primers are listed in Appendix Table A.16.

### 2.2.12 Functional analysis

The functional effect of the GATA6 mutation was analysed by performing luciferase assays.

#### 2.2.12.1 Plasmids

Three separate luciferase reporter constructs were used for functional analysis; natriuretic peptide precursor A promoter luciferase reporter construct (*NPPA*...
promoter-luc), α-myosin heavy chain promoter luciferase reporter construct (αMHC promoter-luc) and brain natriuretic peptide promoter luciferase reporter construct (BNP promoter-luc). The NPPA promoter-luc construct consists of a 500 bp 5´ flanking sequence of the rat NPPA promoter (-700 to -135 bp) subcloned into the pxp-2 vector. The construct was a kind gift from Prof. Mona Nemer, University of Ottawa. The αMHC promoter-luc construct consists of an approximately 3 kb 5´ flanking region of the rat α-MHC gene in a pAM3 vector. The 5´ region extends from -2560 to +421 bp. The construct was a kind gift from Dr Kaie Ojamaa, The Feinstein Institute for Medical Research, Hofstra NSLIJ School of Medicine. The BNP promoter-luc construct consists of a 2.5 kb (-2501 to +83) 5´ flanking sequence from the rat BNP promoter in a pGL2 vector. The construct was a kind gift from Prof. Molkentin, Cincinnati Childrens Hospital.

### 2.2.12.2 Cell culture and transfection

Human Embryonic Kidney 293 (HEK293) cells (Aligent Technologies, CA, USA) were used for the luciferase assays. The cells were stored in liquid nitrogen. Cells were thawed and grown in a culture flask under standard conditions (37°C and 5 % CO₂) in DMEM-high glucose medium (25 mM D-glucose) supplemented with Glutamax, 10 % fetal bovine serum (FBS), penicillin (100 units/ml) and streptomycin (100 µg/ml). The cells were maintained in cell culture plates and ‘split’ when 80 % confluent.
Chapter 2. Methods

For the assays, HEK293 cells were seeded in 96-well plates (approximately $5 \times 10^4$ cells per well in 0.5 ml of complete growth medium) and incubated for 24 hours at 37°C and 5 % CO$_2$. The cells were transiently co-transfected with either wild type or mutant GATA6 expression vectors in combination with one of the three reporter constructs (NPPA promoter-luc, αMHC promoter-luc, or BNP promoter-luc). The pRL-SV40 vector was also co-transfected in each group of cells as an internal control. Lipofectamine LTX and Plus reagent (Invitrogen, NY, USA) was used for transfection according to manufacturer´s instructions. Briefly, 800 ng of the expression plasmid and 400 ng of the reporter construct diluted in 100 µl of Opti-MEM were added to each well. After 30 minutes incubation, 1.75 µl of Lipofectamine LTX was added to the Opti-MEM (Invitrogen, NY, USA). Transfected cells were incubated at 37°C in a 5 % CO$_2$ incubator for 24 hours. Assays were performed in triplicate on separate occasions.

2.2.12.3 Luciferase assays

Luciferase assays were performed 24 hours post-transfection using the Dual-Luciferase Reporter Assay System (Promega, WI, USA) according to manufacturer´s instructions. The cells were washed with phosphate buffered saline (PBS) and lysed with 100 µl of passive lysis buffer for 15 minutes. 100 µl of luciferase assay reagent II was dispensed into each well of the 96-well plate. 20 µl of the cell lysate was added to each of the wells. The plate was placed in a luminometer (Perkin Elmer, IL, USA) and measurements were performed for 10 second periods per well. 100 µl of Stop & Glo Reagent was subsequently added to the reaction in order to quench the firefly
reaction while at the same time initiating the Renilla reaction. Measurements were repeated in 10 second periods in the luminometer (Figure 34).

![Diagram of luciferase reporter assay](http://www.piercenet.com)

**Figure 34:** Schematic representation of a luciferase reporter assay. From: http://www.piercenet.com.

### 2.2.12.4 Statistical analysis

All experiments were performed in triplicate. The normalised relative light units were calculated for each sample as the ratio of light units of firefly luciferase normalised to light units of renilla luciferase. For each promoter-reporter combination, the normalised relative units for the triplicates were averaged and expressed as a mean.
with standard error of the mean (SEM). For relative promoter activity of the mutant GATA6 and the wild type GATA6, data were expressed as fold difference in normalised relative units and SEM. The SEM was expressed as error bars. Analysis of data was performed using a two-tailed paired t test. A p value of ≤0.05 was regarded as evidence of statistical significance.

2.2.13 Linkage analysis in pedigree AF-325

Linkage analysis was performed in 19 individuals from pedigree AF-325. The following section provides a brief background on the principles of linkage analysis followed by more detailed methods.

Linkage analysis takes advantage of genetic recombination to localise the region of the genome that harbours the causative mutation in Mendelian pedigrees. The basic principle of linkage analysis involves investigating the co-segregation of a disease locus with a series of markers of known position distributed throughout the genome. Each typed marker is tested for linkage with the disease phenotype, and by implication, the disease locus.

Markers used for linkage analysis include SNPs and microsatellite repeats. Microsatellite repeat sequences are highly polymorphic because they vary significantly in the number of repeats. Microsatellite markers are distributed throughout the genome in non-coding regions.
In its simplest form, linkage analysis involves counting the number of recombinants and non-recombinants between individual markers and the disease phenotype and calculating the recombination fraction. Markers that segregate with the disease phenotype, and hence lie in close proximity to the disease locus, have a low recombination fraction. Therefore, the identification of a series of markers with a low recombination fraction can be used to define the boundaries of a disease locus. Of note however, such an approach is only possible if precise information on the inheritance patterns of markers is available. Specifically, individuals included in the analysis must fulfil two criteria; firstly, they must be informative for linkage i.e. genotype known and doubly heterozygote and secondly, they have to be phase known i.e. the ancestral origin of the observed alleles must be known in order to allow reconstruction of the haplotypes.

In most pedigrees, the available information is incomplete. For instance some family members may not have genetic data while in others, the phase of the alleles may not be known. In other words, it is not possible to determine which alleles are received as haplotypes from which parent. As a result precise inheritance patterns cannot be defined and multiple inheritance patterns can generate the observed data. In order to overcome this limitation, statistical methods have been developed to calculate the likelihood of linkage in Mendelian pedigrees. The statistical methods are discussed in more detail below.
2.2.13.1 The logarithm of odds (LOD) score

The most efficient statistical test to assess the likelihood of linkage between a disease locus and a marker locus is the logarithm of the odds LOD score. The LOD score compares the likelihood of two alternative hypotheses; the first is that the marker and disease locus co-segregate due to true linkage, while the second is that co-segregation occurs purely by chance.

The LOD score is a parametric linkage test. Parametric linkage analysis is performed under an explicit genetic model. In other words, for any given pedigree, the likelihood of generating the observed genotypic data is calculated under a number of assumptions; the underlying genetic model (e.g. autosomal dominant versus autosomal recessive), the frequency of the causative allele in the general population, the penetrance of the causative mutation, and the number of phenocopies. In comparison to non-parametric analysis, in which the mode of inheritance is not defined, parametric linkage analysis has greater power to detect linkage.

The method for calculation of the LOD score is outlined here. $L_{\text{Linkage}}$ is the likelihood of observing the pedigree data under the assumption that two loci are linked i.e. $\Theta$ is less than 0.5. $L_{\text{Null}}$ on the other hand is the likelihood of observing the data under the null hypothesis i.e. the two loci are not linked and segregate independently ($\Theta = 0.5$). The formula for calculation of the LOD score is as follows:
Chapter 2. Methods

\[ \text{LOD} = \log_{10} \left( \frac{L_{\text{Linkage}}}{L_{\text{Null}}} \right) \]

\[ L_{\text{Linkage}} = (1 - \Theta)^{NR} \times \Theta^R \]

\[ L_{\text{Null}} = 0.5^{(NR+R)} \]

\[ \text{LOD} = \log_{10} \left( \frac{(1 - \Theta)^{NR} \times \Theta^R}{0.5^{(NR+R)}} \right) \]

In the above formula, \( NR \) denotes the number of non-recombinants and \( R \) denotes the number of recombinants. The LOD score is a function of \( \Theta \). Therefore, the LOD score is calculated for a range of \( \Theta \) values (0.0, 0.1, 0.2, 0.3, 0.4). The value of \( \Theta \) that generates the highest LOD score is taken as the recombination fraction.\(^99\) A positive LOD score of 3 or more is conventionally considered as statistical evidence of linkage. A LOD score of 3 equates to 1000 times greater odds that two loci are linked as compared to the odds of the loci not being linked.

**2.2.13.2 Two-point versus multipoint linkage analysis**

Two-point linkage analysis involves an analysis of linkage between an individual marker and a disease locus. Multipoint linkage analysis involves simultaneous analysis of numerous loci and the trait locus.\(^96\) Therefore, all markers are used simultaneously to infer segregation with disease. Multipoint linkage analysis is a more powerful technique for the detection of linkage. Of note, in the case of two-point linkage analysis, the results may not vary significantly with a parametric or a nonparametric
Chapter 2. Methods

In the case of multipoint linkage analysis on the other hand, definition of the mode of inheritance is critical to perform parametric linkage analysis.

2.2.13.3 Computation of LOD scores

Performing linkage analysis in pedigrees with incomplete structure and incomplete informativeness of markers can be a challenge from a computational perspective. A number of computational programs designed to perform efficient linkage analysis are available including GENEHUNTER, VITESSE and SUPERLINK. The computational capability of a single computer is insufficient to perform linkage analysis in large pedigrees with incomplete data. Therefore, multiple parallelised computers are commonly used to perform the analysis.

As discussed above, multipoint linkage analysis is a more effective strategy for the identification of disease genes. Despite the use of efficient computer programs and clusters of high-performance computers, calculation of multipoint LOD scores with a large number of markers in complex pedigrees may not be possible due to prohibitive computing times. One approach to circumvent this problem is to perform preliminary genome-wide two-point linkage analysis to identify potential loci that are linked with disease. Multipoint linkage analysis can subsequently be performed with a limited, but dense set of markers to refine the locus.
Chapter 2. Methods

2.2.13.4 SNP genotyping

Genotyping of SNPs for linkage analysis was performed at the Broad Institute genomics core facility. The Affymetrix Genome-wide Human SNP array 6.0 (Affymetrix, CA, USA) was used to perform SNP genotyping. The array includes a total of 1.8 million genetic markers and 906,600 SNPs. The median intermarker distance for the SNPs on the Affymetrix 6.0 array is 680 bp. Briefly, purified genomic DNA was digested using restriction enzymes. The digested fragments were ligated to adapters that are designed to recognise fragment overhangs. Primers designed to anneal to the adaptor sequences were used to PCR amplify the adaptor-ligated DNA fragments. The amplified fragments were purified using beads made of polystyrene. They were then labelled and hybridised to the Affymetrix 6.0 array. A Birdseed genotype calling algorithm, which uses signal intensity to call SNP genotype, was used for analysis of results. Criteria for exclusion of SNPs from further analysis included; 1) SNPs with ≤95 % call rate, 2) SNPs with a minor allele frequency of <0.1, 3) monomorphic SNPs.

2.2.13.5 Microsatellite marker genotyping

Seven fluorescent dye-labelled microsatellite markers on chromosome 1 (D1S2868, D1S206, D1S2726, D1S252, D1S498, D1S484, and D1S196) from the ABI Prism Linkage Mapping Set (Applied Biosystems, CA, USA) were genotyped. Genotyping was performed at the MGH sequencing core facility.
Chapter 2. Methods

Briefly, PCR amplification of the tandem repeats was performed using specific fluoro-labelled primer sequences flanking the microsatellite makers from Applied Biosystems. The PCR products were electrophoretically separated on polyacrylamide gels and analysed on an automated ABI 3730 DNA Genetic Analyser (Applied Biosystems, CA, USA). Genescan 2.1 software (Applied Biosystems, CA, USA) was used for analysis.

2.2.13.6 Computation of LOD scores

Analysis of linkage data was performed using the SUPERLINK online program. Genotype data were formatted into a specific ‘linkage format’ for SUPERLINK using GDASPort (Affymetrix, CA, USA). Members of the family were encoded as affected, unaffected or phenotype unknown. Positions of the SNP markers on the genetic map were derived from the deCODE Genetics sex-averaged genetic map. The frequencies of the alleles were provided by Affymetrix.

Genome-wide two-point linkage analysis was initially attempted with 367,510 SNP markers. However, the use of such a dense set of markers was associated with very long computation times. We therefore filtered the markers in order to reduce the computational power required to perform the analysis. 14,844 marker SNPs were used for two-point linkage analysis. Parametric two-point LOD scores were computed under the following assumptions; 1) an autosomal dominant pattern of inheritance, 2)
complete penetrance of the causative variant, 3) a disease allele frequency of 0.0001, and 4) no phenocopies.

Two-point linkage analysis identified a locus with a peak LOD score of 2.1 on chromosome 1p21.1-q22. This locus was prioritised for further analysis. Multipoint linkage analysis was performed with 1,506 SNP markers on chromosome 1. The model for multipoint linkage analysis was identical to that used for two-point linkage analysis.

Two-point and multipoint linkage analysis was also performed using microsatellite markers distributed within and around the chromosome 1 locus. Two-point linkage analysis was performed using seven microsatellite markers (D1S2868, D1S206, D1S2726, D1S252, D1S498, D1S484, and D1S196). Initial attempts at performing multipoint linkage analysis with a set of seven markers were associated with very long computation times. We therefore reduced the number of markers to four (D1S206, D1S2726, D1S252 and D1S484).

2.2.14 Analysis of candidate variants in family AF-325

Linkage analysis in family AF-325 identified a locus on chromosome 1p21.1-q22. Candidate variants at this locus were identified by performing exome sequencing in one affected individual from the family. The methods for exome sequencing and prioritisation of candidate variants are discussed in Sections 2.2.6 - 2.2.9. Using these
techniques, potential causative variants were identified in five genes; \textit{C1orf62} (V84I), \textit{CELF3} (T240P), \textit{PMVK} (V96A), \textit{MRPS21} (L75V), \textit{FLG} (H1880N).

Sanger sequencing was performed to reconfirm the presence of the five variants identified by exome sequencing and to perform segregation analysis in other affected individuals from pedigree AF-325. Primers were designed using the UCSC website (hg19, http://genome.ucsc.edu/) and the ExonPrimer website (http://ihg.gsf.de/ihg/ExonPrimer). The primer sequences are summarised in Appendix Table A.17. Synthesis of primers was performed by the MGH sequencing core facility. The technique for PCR amplification and sequencing of the candidate variants is outlined in Section 2.2.5.
CHAPTER 3.
THE ROLE OF THE SK3 CHANNEL IN CARDIAC ELECTROPHYSIOLOGY

3.1 Introduction

In 2010, a GWAS of lone AF identified risk variants for the arrhythmia at a locus on chromosome 1q21. The risk variants at this locus are intronic to the KCNN3 gene. KCNN3 encodes a calcium-activated potassium channel (SK3) which is a compelling candidate gene for AF. The calcium-activated potassium channels are a major group of potassium selective ion channels. A unique property of these channels is that they are gated directly by calcium binding and therefore provide a link between intracellular calcium transients and membrane potassium conductance.

The role of SK channels in cardiac electrophysiology is currently not clearly defined. The three SK channel subtypes (SK1- SK3) display variable spatial distribution in the heart. Two of the channel subtypes, SK1 and SK2, are predominantly expressed in the atrium. SK3 on the other hand is equally expressed in the atrium and the ventricle.

In a murine model, selective knockout of the SK2 channel results in a significant prolongation of the action potential. In keeping with the expression profile of the SK2 channel, action potential prolongation was observed in atrial but not ventricular myocytes. Consistent with the findings in atrial tissue, SK2 channel knockout in AV nodal cells results in a prolongation of the action potential. Conversely,
overexpression of the channel results in an abbreviated action potential. In addition to
the effect on the action potential, SK2 channel overexpression results in a significant
increase in the spontaneous firing rate of AV nodal myocytes.

As discussed in Section 1.4.1.3.1, in neural tissue, SK channels mediate prolonged
afterhyperpolarisation of the membrane potential, which reduces neuronal
excitability and tonic firing frequency. However, based on the aforementioned
studies, SK channels appear to have a distinct role in the cardiac action potential and
contribute to the late phase of repolarisation rather than mediating hyperpolarisation.

The role of the SK3 channel in cardiac electrophysiology is currently largely unknown.
Interestingly, SK3 has been demonstrated to form functional heteromultimeric
complexes with SK2 in vitro. Blockade of the SK2/3 complex results in a
prolongation of the action potential in isolated cardiomyocytes. Further, in a rabbit
atrial model mimicking pulmonary venous ectopy, blockade of SK2 and SK3 with
apamin attenuates the acceleration of repolarisation in response to burst pacing. It
is plausible therefore that the SK3 channel also plays an important role in cardiac
repolarisation. However, the SK3 channel has not been investigated in isolation.

We hypothesised that variable expression of the SK3 channel creates a profibrillatory
substrate in the atrium. In order to test our hypothesis, we characterised the cardiac
electrophysiological phenotype of a mouse strain (SK3\textsuperscript{T}) designed to either overexpress Kcnn3 or to induce knockout of the gene.

### 3.2 Results

#### 3.2.1 Generation of SK3\textsuperscript{T} mutant mice

The SK3\textsuperscript{T} mouse line was generated in Dr. Adelman’s laboratory at the Vollum Institute by site-specific insertion of a tetracycline-based genetic switch just upstream of the translation initiation site of Kcnn3.\textsuperscript{223} Adelman et al previously reported that in brain tissue, homozygote mice with the mutant allele (SK3\textsuperscript{T/T}) have a three-fold upregulation of Kcnn3 expression at baseline. Upon administration of tetracycline (or doxycycline), Kcnn3 expression is almost completely eliminated.

Heterozygote mice (SK3\textsuperscript{+/T}) were acquired from Jackson Laboratories (ME, USA) and crossed to generate SK3\textsuperscript{T/T} mice, SK3\textsuperscript{+/T} mice and wild type (WT) siblings. Genotyping was performed using routine PCR. Gel electrophoresis of the PCR product demonstrated a 140 bp band in SK3\textsuperscript{T/T} mice, a 432 bp band in WT mice, and two bands (140 bp and 432 bp) in SK3\textsuperscript{+/T} mice. Heterozygote crosses in the SK3\textsuperscript{T} mouse line did not follow pure 1:2:1 Mendelian inheritance ratios. We analysed 17 litters which included 89 neonates from eight different SK3\textsuperscript{+/T} breeding pairs. As demonstrated in Table 5, out of the 89 neonates, 17 (19 %) were SK3\textsuperscript{T/T}, 44 (50 %) were SK3\textsuperscript{+/T} and 28 (31 %) were WT. The deviation from the expected Mendelian ratio was not statistically significant (X\textsuperscript{2}; p=0.26).
### Table 5 – Ratios from $SK3^{+/T} \times SK3^{+/T}$ cross

<table>
<thead>
<tr>
<th></th>
<th>WT</th>
<th>$SK3^{+/T}$</th>
<th>$SK3^{+/T}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>31</td>
<td>50</td>
<td>19</td>
</tr>
<tr>
<td>N</td>
<td>28</td>
<td>44</td>
<td>17</td>
</tr>
</tbody>
</table>

*Kcnn3* expression in $SK3^{T}$ mouse hearts was analysed by RT-PCR. Consistent with previous reports in brain tissue, $SK3^{+/T}$ mice displayed upregulation of *Kcnn3* expression in all four cardiac chambers as compared to WT controls. The most marked overexpression was seen in the ventricles ($SK3^{+/T}$ vs WT; RA, 1.8-fold; LA, 2-fold; RV, 10.1-fold; LV, 9.8-fold). $SK3^{+/T}$ mice also displayed upregulation of *Kcnn3* expression however not surprisingly, the upregulation was more subtle. Of note, we did not observe *Kcnn3* overexpression in the LA in $SK3^{+/T}$ mice ($SK3^{+/T}$ vs WT; RA, 1.8-fold; LA, 1-fold; RV, 1.5-fold; LV, 2.1-fold).

In the homozygote group treated with doxycycline ($SK3^{+/T} + DOX$), *Kcnn3* overexpression was silenced. In three out of four chambers, *Kcnn3* expression levels were suppressed relative to those observed in WT controls ($SK3^{+/T} + DOX$ vs WT; RA, 0.2-fold; LA, 0.8-fold; RV, 0.5-fold; LV, 1.6-fold). Of note however, in contrast to the previous reports in brain tissue from Adelman *et al*, we did not observe a complete elimination of *Kcnn3* expression in cardiac tissue. The results from RT-PCR are summarised in Figure 35A-C.
Figure 3.5: Relative expression levels of \textit{Kcnn3} in \textit{SK3}\textsuperscript{T} mice as measured by RT-PCR. mRNA levels were quantified in all four cardiac chambers and normalised to levels of β-actin. A) Relative expression of \textit{Kcnn3} in \textit{SK3}\textsuperscript{T/T} mice relative to WT control mice. B) Relative expression of \textit{Kcnn3} in \textit{SK3}\textsuperscript{+T/mice relative to WT control mice. C) Relative expression of \textit{Kcnn3} in \textit{SK3}\textsuperscript{T/T} + DOX mice relative to WT control mice.
Chapter 3. The Role of The SK3 Channel in Cardiac Electrophysiology

$SK3^{T/T}$ mice demonstrated normal growth and achieved similar body weights to wild type siblings ($SK3^{T/T}$ vs WT; $25.8 \pm 2.8 \text{ g vs } 26.3 \pm 1.7 \text{ g, p=0.6}$). Histological analysis was performed in three-month-old $SK3^{T/T}$ and WT mice. Hematoxylin-eosin staining did not reveal structural cardiac abnormalities in $SK3^{T/T}$ mice. Specifically, there was no evidence of abnormal cardiac hypertrophy, dilatation or myofibrillar disarray. Further, Masson’s trichrome staining revealed an absence of myocardial fibrosis (Figure 36). Taken together, these findings suggest that variable expression of $Kcnn3$ does not result in histological or morphological cardiac abnormalities.

![Figure 36: Histological analysis of $SK3^{T/T}$ and WT mouse hearts. A) Hematoxylin and eosin staining of heart sections from three-month-old $SK3^{T/T}$ and WT mice. B) Masson’s Trichrome staining of heart sections. Neither group demonstrated abnormalities of cardiac structure or abnormal fibrosis.](image-url)
Chapter 3. The Role of The SK3 Channel in Cardiac Electrophysiology

Echocardiographic analysis was performed in three-month-old SK3\textsuperscript{T/T}, SK3\textsuperscript{+/T} and WT mice. Four parameters were measured during echocardiography; left ventricular end diastolic dimension (LVEDD), left ventricular end systolic dimension (LVESD), interventricular septal dimension in diastole (IVSD), posterior wall dimension in diastole (PWD). Left ventricular fractional shortening (FS) and ejection fraction (EF) were derived from these measurements. There were no significant differences in left ventricular dimensions and systolic function in 3-month-old SK3\textsuperscript{T/T} and SK3\textsuperscript{+/T} mice compared to age-matched WT controls (Table 6). The mitral and aortic valves also had normal appearances in all three experimental groups.

Table 6. Echocardiographic measurements in 3-month-old mice

<table>
<thead>
<tr>
<th></th>
<th>LVEDD (mm)</th>
<th>LVESD (mm)</th>
<th>IVSD (mm)</th>
<th>PWD (mm)</th>
<th>FS (%)</th>
<th>EF (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT (n=15)</td>
<td>3.0 ± 0.2</td>
<td>1.5 ± 0.2</td>
<td>0.8 ± 0.1</td>
<td>0.8 ± 0.01</td>
<td>48 ± 8</td>
<td>85 ± 6</td>
</tr>
<tr>
<td>SK3\textsuperscript{T/T} (n=13)</td>
<td>3.2 ± 0.2</td>
<td>1.6 ± 0.1</td>
<td>0.7 ± 0.01</td>
<td>0.8 ± 0.01</td>
<td>50 ± 1</td>
<td>87 ± 1.1</td>
</tr>
<tr>
<td>SK3\textsuperscript{+/T} (n=5)</td>
<td>3.0 ± 0.4</td>
<td>1.5 ± 0.1</td>
<td>0.8 ± 0.02</td>
<td>0.8 ± 0.05</td>
<td>53 ± 4</td>
<td>89 ± 3</td>
</tr>
</tbody>
</table>

Abbreviations: LVEDD, left ventricular end diastolic dimension; LVESD, left ventricular end systolic dimension; IVSD, interventricular septal dimension in diastole; PWD, posterior wall dimension in diastole; FS, fractional shortening; EF, ejection fraction.

3.2.2 Sudden death in SK3\textsuperscript{T/T} mice

We observed a high incidence of sudden death amongst SK3\textsuperscript{T/T} mice. As demonstrated in the Kaplan-Meier survival curves in Figure 37, at a three-month time point, 7 out of 19 SK3\textsuperscript{T/T} males (32 %) died suddenly while there was no mortality observed amongst
the SK3\(^{+/+}\) and WT mice. The highest mortality was observed between the ages of 20 and 40 days.

As discussed above, SK3\(^{+/+}\) mice demonstrated normal growth and achieved similar body weights to wild type siblings. We did not observe any abnormal behaviour or appearance in the days preceding death. Interestingly, we found that doxycycline mediated silencing of SK3 expression resulted in a rescue of the sudden death phenotype. Out of 25 SK3\(^{{+/+}}\) males treated with doxycycline (depicted as SK3\(^{+/+}\) + DOX in the Kaplan-Meier curve), none died suddenly.

![Kaplan-Meier survival plot at 90 days.](image)

**Figure 37:** Kaplan-Meier survival plot at 90 days. Cumulative survival rate of male WT (n=42), SK3\(^{+/+}\) (n=55), SK3\(^{+/+}\) + DOX (n=25), and SK3\(^{+/+}\) (n=19) mice is plotted against age (days). SK3\(^{+/+}\) mice demonstrated a marked increase in the incidence of sudden death between the ages of 20 and 40 days.
3.2.3 Ambulatory monitoring in $SK3^T$ mice

In order to determine the mechanism of sudden death in $SK3^{T/T}$ mice, telemetry devices were implanted to perform continuous monitoring of the cardiac rhythm. Six $SK3^{T/T}$ mice and two WT control mice underwent ambulatory monitoring. As discussed in the previous section, we observed sudden death amongst $SK3^{T/T}$ mice from the age of 20 days. Therefore, 20 days represents the ideal time point to commence monitoring. However, due to constraints related to the size of the telemetry device, it was not possible to perform device implantation until the mice were at least one-month-old ($SK3^{T/T}$ vs WT; age at implant, 37±7 vs 34±0 days, p=0.58).

Out of the six $SK3^{T/T}$ mice, four died suddenly while the telemetry devices were in situ. However, due to premature battery depletion of one device, the cardiac rhythm at the time of death was only recorded in three mice. Of note, neither of the WT control mice died during the period of monitoring. The rhythm abnormalities in $SK3^{T/T}$ mice are discussed in more detail below.

The first $SK3^{T/T}$ mouse to die suddenly with the telemetry device in situ, $SK3^{T/T}$ (1), had the device implanted at the age of 39 days. Sudden death occurred after a monitoring period of 70 days. The cardiac rhythm recorded at the time of death was bradycardia with sinus arrest and advanced atrioventricular conduction block. A representative lead II ECG is illustrated in Figure 38B.
The second $SK3^{T/T}$ mouse, $SK3^{T/T}$ (2), had a telemetry device implanted at the age of 34 days. Sudden death occurred 80 days post-implantation. The cardiac rhythm recorded at the time of death was bradycardia with pronounced first-degree atrioventricular block and interventricular conduction delay. A representative lead II ECG is illustrated in Figure 38C.

The third mouse, $SK3^{T/T}$ (3), had a telemetry device implanted at the age of 28 days. Sudden death occurred 10 days post implantation. Of note, there were no complications following device implantation and the mouse did not display signs of infection in the days preceding death. The cardiac rhythm at the time of death was severe bradycardia with ventricular bigeminy. A representative lead II ECG is illustrated in Figure 38D.

The fourth mouse, $SK3^{T/T}$ (4), had a telemetry device implanted at the age of 39 days. Sudden death occurred 80 days following implantation. As mentioned above, due to premature depletion of the device battery, we were unable to record the cardiac rhythm at the time of death.

The two remaining $SK3^{T/T}$ mice, $SK3^{T/T}$ (5) and $SK3^{T/T}$ (6), did not die during ambulatory monitoring. $SK3^{T/T}$ (5) had a telemetry device implanted at the age of 48 days and was monitored for a total period of 80 days. $SK3^{T/T}$ (6) had a telemetry device implanted at
the age of 34 days and was monitored for a period of 115 days. Monitoring was continued until depletion of the battery in the device.

In addition to the episodes of heart block and severe bradycardia at the time of sudden death, all six $SK3^{T/T}$ mice displayed recurrent episodes of atrioventricular dissociation (Figure 39). More specifically, we observed slowing of the sinus rate with a junctional escape rhythm. These episodes occurred both at rest and during periods of activity. Of note, the two wild type mice, WT (1) and WT (2), did not display atrioventricular dissociation.

Based on the findings of bradycardia and heart block in $SK3^{T/T}$ mice, we analysed the complete telemetry data set to determine whether there were differences in overall heart rate between $SK3^{T/T}$ mice and WT control mice. We also analysed data obtained within a single epoch lasting one hour in 30-day-old mice to determine whether the PR and RR intervals were significantly different between $SK3^{T/T}$ and WT mice. The intervals were measured during the light phase of the circadian cycle as mice generally have low levels of locomotor activity during the day and this is the period when cardiovascular variability is typically assessed. As demonstrated in the plot graphs in Figure 40, the overall heart rate was slower in $SK3^{T/T}$ mice as compared to WT control mice. More specifically, while the mean heart rate in the six $SK3^{T/T}$ mice varied between 326 and 549 beats per minute, the two WT control mice had mean heart rates of 550 and 571 beats per minute respectively.
Figure 3B: Representative lead II electrogram recordings from $SK3^{T/T}$ mice at the time of sudden death. A) Normal sinus rhythm in a WT mouse demonstrating a normal PR interval and heart rate. B) Severe bradycardia with sinus arrest and atrioventricular block in $SK3^{T/T}$ mouse (number 1). C) Severe sinus bradycardia with first-degree atrioventricular block in $SK3^{T/T}$ mouse (number 2). D) Severe bradycardia with ventricular bigemini in $SK3^{T/T}$ mouse (number 3). The voltage of the QRS complexes (mV) is depicted on the y axis and time (sec) is depicted on the x axis. Abbreviations: mV, millivolts.
Figure 39: Spontaneous atrioventricular dissociation in SK3\textsuperscript{T/T} mice. A) Normal sinus rhythm in a WT mouse. B) A representative tracing demonstrating atrioventricular dissociation in an SK3\textsuperscript{T/T} mouse. The voltage of the QRS complexes (mV) is depicted on the y axis and time (sec) is depicted on the x axis. Abbreviations: mV, millivolts.
Figure 40: Plot graphs of raw data from continuous heart rate monitoring in SK3<sup>T/T</sup> and WT mice. Each data point in the graph represents a heart rate measurement. Heart rates were sampled once every 200 seconds for a period of 5 seconds. SK3<sup>T/T</sup> mice display a lower average heart rate (indicated by the dotted line) and increased heart rate variability. Abbreviations: bpm, beats per minute; HR, heart rate.
The PR interval was analysed in four $SK3^{T/T}$ mice and one of the WT control mice. Due to the quality of the telemetry recordings, we were unable to assess PR interval in $SK3^{T/T}$ (5), $SK3^{T/T}$ (6) and WT (2). As demonstrated in Figure 41 and Table 7, $SK3^{T/T}$ mice had a more prolonged median PR interval and also displayed increased PR interval variation as indicated by increased interquartile ranges when compared to the WT control mouse.

Figure 41: PR interval in conscious $SK3^{T/T}$ and WT mice. The box plots (median, interquartile range) demonstrate that $SK3^{T/T}$ mice have increased variability of the PR interval during telemetry monitoring. Abbreviations: ms, milliseconds.

The RR interval was analysed in five $SK3^{T/T}$ mice and two WT control mice. $SK3^{T/T}$ mice had a lower median heart rate compared to the WT control mice (Figure 42). $SK3^{T/T}$ mice also displayed more pronounced variability of the heart rate. The results are summarised in Table 7 and Figure 42.
Chapter 3. The Role of The SK3 Channel in Cardiac Electrophysiology

Table 7. Baseline conduction intervals in 3-month-old mice

<table>
<thead>
<tr>
<th>Mouse</th>
<th>PR Median</th>
<th>PR IQR</th>
<th>RR Median</th>
<th>RR IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SK3&lt;sup&gt;T/T&lt;/sup&gt; (1)</td>
<td>37.5</td>
<td>0.9</td>
<td>103</td>
<td>20.7</td>
</tr>
<tr>
<td>SK3&lt;sup&gt;T/T&lt;/sup&gt; (3)</td>
<td>36</td>
<td>3.7</td>
<td>140</td>
<td>37</td>
</tr>
<tr>
<td>SK3&lt;sup&gt;T/T&lt;/sup&gt; (4)</td>
<td>N.R.</td>
<td>N.R.</td>
<td>111</td>
<td>27.7</td>
</tr>
<tr>
<td>SK3&lt;sup&gt;T/T&lt;/sup&gt; (5)</td>
<td>35</td>
<td>2.4</td>
<td>141</td>
<td>55</td>
</tr>
<tr>
<td>SK3&lt;sup&gt;T/T&lt;/sup&gt; (6)</td>
<td>40.8</td>
<td>1.5</td>
<td>129</td>
<td>30</td>
</tr>
<tr>
<td>WT (1)</td>
<td>34.4</td>
<td>1.5</td>
<td>99</td>
<td>13.9</td>
</tr>
<tr>
<td>WT (2)</td>
<td>N.R.</td>
<td>N.R.</td>
<td>112</td>
<td>17</td>
</tr>
</tbody>
</table>

*No interval data was obtained from SK3<sup>T/T</sup> (2)

Abbreviations: N.R., No recording; IQR interquartile range

Figure 42: RR intervals in conscious SK3<sup>T/T</sup> and WT mice. The box plots (median, interquartile range) demonstrate that SK3<sup>T/T</sup> mice have increased heart rate variability during telemetry monitoring. Abbreviations: ms, milliseconds.
3.2.4 Ventricular conduction velocity and action potential duration in SK3$^{T/T}$ mice

Optical mapping of isolated, Langendorff-perfused hearts from SK3$^{T/T}$, SK3$^{+/T}$ and WT mice was performed at one-month of age. The one-month time point was selected due to a high incidence of sudden death amongst SK3$^{T/T}$ mice between the ages of 20 and 40 days. We mapped four SK3$^{T/T}$ and five SK3$^{+/T}$ mice and compared them to six WT control mice.

Conduction velocities were measured at a pacing cycle length of 150 ms. SK3$^{T/T}$ mice demonstrated a significantly reduced conduction velocity (CV) compared to WT mice (SK3$^{T/T}$ vs WT; CV 0.45±0.04 vs 0.60±0.09 mm/ms, p=0.001). We also observed a trend towards slower ventricular conduction in SK3$^{+/T}$ mice compared to WT mice however the difference was not statistically significant (SK3$^{+/T}$ vs WT; CV 0.57±0.04 vs 0.60±0.09 mm/ms, p=0.78). The results are summarised in Figure 43A. A representative example of the ventricular activation pattern during pacing at 150 ms from the left ventricular apex is demonstrated in Figure 43B.

As illustrated in Figure 43C, we did not observe significant differences in the APD$_{80}$ between SK3$^{T/T}$, SK3$^{+/T}$ and WT mice (SK3$^{T/T}$ vs WT; APD$_{80}$, 61±9 vs 61±7 ms, p=0.86, SK3$^{+/T}$ vs WT; APD$_{80}$, 59±5 vs 61±7 ms, p=0.62). However, we did observe a marked increase in the dispersion of the APD$_{80}$ in SK3$^{T/T}$ mice (SK3$^{T/T}$ vs WT; APD$_{80}$ dispersion, 13±4 vs 5±4 ms, p=0.02). The APD$_{80}$ dispersion in SK3$^{+/T}$ mice was not altered compared to WT mice (SK3$^{+/T}$ vs WT; APD$_{80}$ dispersion, 4±3 vs 5±4 ms, p=0.66).
Figure 43: High resolution optical mapping in isolated $SK_3^{T/T}$, $SK_3^{+/+}$ and WT mouse hearts. A) Ventricular conduction velocity at a pacing cycle length of 150 ms in one-month-old $SK_3^{T/T}$, $SK_3^{+/+}$ and WT mice. B) Representative example of ventricular activation maps on the anterior surface of the heart in a WT and $SK_3^{T/T}$ mouse. Hearts were paced from the left ventricular apex at a cycle length of 150 ms and epicardial activation was recorded as it spread from the apex to the base. C) $SK_3^{T/T}$ and $SK_3^{+/+}$ mice do not display significant differences in APD$_{80}$ as compared to WT control mice. D) $SK_3^{T/T}$ display increased dispersion of the APD$_{80}$ as compared to WT control mice. $SK_3^{+/+}$ mice do not display differences in APD$_{80}$ dispersion. APD measurements were made at a pacing cycle length of 150 ms. Abbreviations: APD$_{80}$, action potential duration at 80% repolarization.
3.2.5 Connexin-43 expression in SK3$^{T/T}$ mice

Connexin-43 is a major determinant of ventricular conduction velocity.$^{269}$ In order to exclude altered connexin-43 expression as the cause of reduced conduction velocity in SK3$^{T/T}$ mice, we performed immunohistochemical staining. Hearts from three-month-old SK3$^{T/T}$ mice were compared to WT control mice. We did not observe significant quantitative changes in connexin-43 expression or differences in connexin-43 localisation between the two groups. Consistent with previous studies,$^{269}$ connexin-43 was demonstrated to localise to the borders between myocytes (Figure 44).

Figure 44: Immunostaining of SK3$^{T/T}$ and WT mouse hearts for connexin-43. Paraffin imbedded heart sections from SK3$^{T/T}$ and WT mice were stained for connexin-43 and counterstained with Mayer’s Hematoxylin Solution. Arrows indicate connexin-43 staining at the intercalated disks of the cells. No quantitative changes in connexin-43 expression were observed between SK3$^{T/T}$ and WT hearts.
3.2.6 Programmed stimulation in $SK3^{T/T}$ and $SK3^{+/T}$ mice

Programmed electrical stimulation was performed in $SK3^{T/T}$, $SK3^{+/T}$ and WT mice to assess cardiac conduction and refractoriness as well as arrhythmia inducibility. Five parameters were measured during programmed stimulation; corrected sinus node recovery period (cSNRP), atrioventricular effective refractory period (AVERP), atrial effective refractory period (AERP), atrioventricular Wenckebach cycle length (AV WB) and ventricular effective refractory period (VERP).

The initial strategy was to characterise three-month-old mice. Three $SK3^{T/T}$ mice, thirteen $SK3^{+/T}$ mice and nine WT mice underwent programmed stimulation at three months. However, due to a high incidence of sudden death in $SK3^{T/T}$ mice between the ages of 20 - 40 days, we modified our strategy and characterised mice at a one-month time point. Six $SK3^{T/T}$ mice, five $SK3^{+/T}$ mice and eight WT mice were characterised at one month.

At one month, both $SK3^{T/T}$ and $SK3^{+/T}$ mice displayed inducible AF/atrial arrhythmias. Three of the six $SK3^{T/T}$ mice and one of the five $SK3^{+/T}$ mice had inducible atrial arrhythmias. Atrial arrhythmias were defined as arrhythmias that lasted for more than five consecutive beats. None of the eight age-matched WT mice had atrial arrhythmias. $SK3^{T/T}$ mice also displayed spontaneous atrial premature beats during ambulatory monitoring. The results are summarised in Figure 45. Of note, at a three month time point, we did not observe atrial arrhythmias in $SK3^{T/T}$, $SK3^{+/T}$ or WT mice.
One-month-old $SK3^{T/T}$ mice demonstrated an abbreviated AVERP compared to WT mice during programmed stimulation ($SK3^{T/T}$ vs WT; AVERP, 43±6 vs 52±9, \(p=0.02\)).

Further, both $SK3^{T/T}$ and $SK3^{+/T}$ mice displayed abbreviated Wenckebach cycle lengths ($SK3^{T/T}$ vs WT; AVWB, 71±4 vs 87±6, \(p=0.0001\), $SK3^{+/T}$ vs WT; AVWB, 77±5 vs 87±6, \(p=0.01\)). Among three-month-old $SK3^{T/T}$ mice on the other hand, we observed a trend towards a prolonged AVERP ($SK3^{T/T}$ vs WT; AVERP 61±1 vs 52±6 ms, \(p=0.86\)). These mice also have a trend towards a more prolonged cSNRT ($SK3^{T/T}$ vs WT; cSNRT, 124±52 vs 98±50 ms, \(p=0.86\)). The results are summarised in Table 8 and Table 9.

Table 8. Baseline Intracardiac Conduction Intervals in 1-month-old mice

<table>
<thead>
<tr>
<th></th>
<th>cSNRP (ms)</th>
<th>AVERP (ms)</th>
<th>AERP (ms)</th>
<th>AV WB (ms)</th>
<th>VERP (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT (n=8)</td>
<td>184 ± 32</td>
<td>52 ± 9</td>
<td>30 ± 3</td>
<td>87 ± 6</td>
<td>30 ± 5</td>
</tr>
<tr>
<td>$SK3^{+/T}$</td>
<td>150 ± 27</td>
<td>49 ± 2</td>
<td>30 ± 2</td>
<td>77 ± 5*</td>
<td>31 ± 5</td>
</tr>
<tr>
<td>$SK3^{T/T}$</td>
<td>180 ± 46</td>
<td>43 ± 6*</td>
<td>31 ± 5</td>
<td>71 ± 4**</td>
<td>31 ± 7</td>
</tr>
</tbody>
</table>

* \(p < 0.05\), **\(p < 0.01\)

Measurements of refractory periods were made using a basic cycle length of 100 ms. Data are expressed as mean ± standard deviation. Abbreviations: cSNRP, corrected sinus node recovery period; AVERP, atrioventricular nodal effective refractory period; AERP, atrial effective refractory period; AV WB, atrioventricular Wenckebach cycle length.

Table 9. Baseline Intracardiac Conduction Intervals in 3-month-old mice

<table>
<thead>
<tr>
<th></th>
<th>cSNRP (ms)</th>
<th>AVERP (ms)</th>
<th>AERP (ms)</th>
<th>AV WB (ms)</th>
<th>VERP (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT (n=9)</td>
<td>98 ± 50</td>
<td>52 ± 6</td>
<td>34 ± 2</td>
<td>90 ± 7</td>
<td>43 ± 9</td>
</tr>
<tr>
<td>$SK3^{+/T}$</td>
<td>88 ± 48</td>
<td>50 ± 9</td>
<td>36 ± 4</td>
<td>81 ± 10</td>
<td>38 ± 6</td>
</tr>
<tr>
<td>$SK3^{T/T}$</td>
<td>124 ± 52</td>
<td>61 ± 1</td>
<td>34 ± 6</td>
<td>88 ± 2</td>
<td>38 ± 2</td>
</tr>
</tbody>
</table>
3.3 Discussion

3.3.1 Summary of main findings

The most striking finding from this study is that homozygote ($SK3^{T/T}$) mice have a high incidence of sudden death. These mice display pronounced overexpression of $Kcnn3$ in the ventricles. The sudden death phenotype is rescued through administration of...
doxycycline, which suppresses \(Kcnn3\) expression. Continuous ambulatory monitoring demonstrated that the mechanism of sudden death in \(SK3^{T/T}\) mice is related to bradycardia and heart block. High-resolution optical mapping of the ventricle revealed a significant reduction in ventricular conduction velocity in \(SK3^{T/T}\) mice. Analysis of the ventricular action potential demonstrated that \(SK3^{T/T}\) mice have an increased dispersion of repolarisation. Programmed stimulation revealed that both \(SK3^{T/T}\) and heterozygote (\(SK3^{+/T}\)) mice have inducible AF. Further, \(SK3^{T/T}\) mice display frequent spontaneous atrial ectopics during ambulatory monitoring of the cardiac rhythm. Consistent with the findings during ambulatory monitoring, three-month-old \(SK3^{T/T}\) mice demonstrate a trend towards a more prolonged AVERP and cSNRP during programmed stimulation. Paradoxically however, \(SK3^{T/T}\) mice display an acceleration of atrioventricular conduction as evidenced by a shorter AVERP.

### 3.3.2 \(SK3^{T/T}\) mice display marked ventricular overexpression of \(Kcnn3\)

The \(SK3^T\) allele is designed to overexpress \(Kcnn3\) at baseline and eliminate gene expression upon administration of doxycycline. An interesting observation in homozygote \(SK3^{T/T}\) mice is that the degree of \(Kcnn3\) overexpression is chamber-specific. These mice demonstrate an almost five-fold increase in expression of \(Kcnn3\) in the ventricles as compared to the atria. Indeed, ventricular \(Kcnn3\) expression levels are significantly higher than those previously reported in brain tissue by Adelman et al.\cite{223}
The mechanisms underlying preferential overexpression of $Kcnn3$ in the ventricles of $SK3^{T/T}$ mice remain unclear. As discussed in Section 2.1.1, $Kcnn3$ expression in $SK3^{T}$ mice is under the transcriptional control of a tet$_{o5}$ CMV promoter. At baseline, binding of a tTa protein to the tet$_{o5}$ CMV promoter strongly activates expression of $Kcnn3$. In turn, the native promoter of the $Kcnn3$ gene drives expression of tTA. It is plausible that the native $Kcnn3$ promoter in the ventricle drives higher tTA expression as compared to the atrium. However at this stage, this mechanism is speculative.

3.3.3 $SK3^{T/T}$ mice have a high incidence of sudden death

An unexpected finding of this study is that $SK3^{T/T}$ mice display a high incidence of sudden death. Sudden death is defined as death occurring within an hour of onset of symptoms or an abrupt change in symptoms. The vast majority of instantaneous deaths arise due to primary arrhythmic events, referred to as sudden cardiac deaths. Monitoring of the cardiac rhythm in $SK3^{T/T}$ mice at the time of sudden death demonstrated severe bradycardia due to atrioventricular conduction block and sinus node dysfunction. These observations suggest that the mechanism of sudden death is related to a primary cardiac conduction defect. Severe bradycardia due to failure of cardiac impulse propagation is an important cause of sudden cardiac death. 15 - 20% of cases of sudden cardiac death are attributable to bradycardias, which include advanced atrioventricular conduction block and asystole. The following section provides a brief background on cardiac conduction disease before discussing potential mechanisms of conduction block in $SK3^{T/T}$ mice.
Sinus node dysfunction arises due to failure of impulse initiation or propagation at the SA node. The clinical manifestations of sinus node dysfunction include sinus bradycardia and episodes of sinus arrest. Atrioventricular conduction block on the other hand arises due to abnormalities of impulse propagation at the level of the atrioventricular node or the infra-His conduction system. The clinical manifestations of atrioventricular conduction block are determined by the level of block. Conduction block in the distal atrioventricular conduction system is associated with the highest risk of life-threatening bradyarrhythmias and sudden death.\textsuperscript{273}

The majority of cases of cardiac conduction disease arise due to structural abnormalities of the conduction system.\textsuperscript{273} However, conduction block has also been reported in patients with ion channel mutations. To date, multiple \textit{SCN5A} mutations have been identified in patients with isolated conduction tissue disease, which includes sick sinus syndrome and atrioventricular conduction block.\textsuperscript{141, 274, 275} Functional characterisation of these mutations revealed an attenuated $I_{Na}$ current, which is predicted to reduce conduction velocity. \textit{HCN4} mutations have also been reported to underlie isolated sinus node dysfunction.\textsuperscript{191, 276} The HCN4 channel underlies the $I_f$ current, which determines the slope of the pacemaker potential and hence sinus node automaticity. Functional analysis revealed that the mutant HCN4 channels are activated at more negative voltages resulting in a reduced $I_f$ current during diastolic depolarisation. Potassium channel mutations have not previously been implicated in isolated conduction tissue disease.
The mechanistic link between SK3 channel overexpression and sinus node dysfunction and atrioventricular conduction block is presently unclear. A plausible explanation is that consistent with its role in neural tissue, the SK3 channel causes a hyperpolarising shift in the resting membrane potential. As discussed in Section 1.4.1.3.1, SK channels contribute to classical afterhyperpolarisation in the neuronal action potential, which moves the membrane potential away from the threshold. Afterhyperpolarisation renders neurons less excitable for tens to hundreds of milliseconds. The SK3 channel subtype in particular has been reported to cause pronounced afterhyperpolarisation and conduction slowing in motor neurones. Based on these observations, overexpression of the SK3 channel would be predicted to suppress the sinus node firing rate as well as the response to depolarising inputs into the atrioventricular conduction system.

In order to further investigate the effect of SK3 channel overexpression on cardiac conduction, we performed programmed electrical stimulation in one-month- and three-month-old SK3T/T mice. Consistent with the findings during ambulatory monitoring, three-month-old mice displayed a trend towards a more prolonged atrioventricular nodal refractory period (AVERP) and sinus node recovery period (SNRP). Paradoxically however, in one-month-old mice, we observed a shortening of the AVERP while the SNRP remained unaltered. The mechanisms underlying these discrepant findings are not clear.
In contrast to our findings, previous studies in murine models have demonstrated that overexpression of the SK2 channel, which is closely related to SK3, results in an increased firing frequency in atrioventricular nodal myocytes. Knockout of the channel has the opposite effect. Further, knockout of the SK2 channel results in a significant prolongation of the sinus node recovery period and Wenckebach cycle length. These observations suggest that SK2 and SK3 may have different roles in the cardiac conduction system.

3.3.4 SK3<sup>T/T</sup> mice have slowed ventricular conduction

SK3<sup>T/T</sup> mice display slower ventricular conduction as compared to WT control mice. Propagation of the cardiac impulse though the myocardium involves generation of action potentials in individual cardiomyocytes and conduction of the action potentials from cell-to-cell though intracellular gap junctions. The conduction velocity is determined by two main factors, the membrane characteristics and tissue structure. SK3<sup>T/T</sup> mice do not display structural changes that would be predicted to alter conduction velocity. Further, they do not have alterations in the distribution of connexin-43, the predominant intracellular gap junction. These findings suggest that abnormal ventricular conduction is due to altered membrane characteristics.

The key membrane determinant of myocardial conduction velocity is the Nav1.5 channel which conducts the depolarising $I_{Na}$ current. Nav1.5 is a voltage-gated channel with an activation threshold of approximately -55mV. It is plausible that in keeping with the proposed mechanism of conduction slowing in the cardiac conduction
system, a reduced ventricular conduction velocity in $SK3^{T/T}$ mice is attributable to pronounced hyperpolarisation in ventricular cardiomyocytes due to an augmented $I_{KCa}$ current. Membrane hyperpolarisation would be predicted to slow down action potential generation due to a prolongation of the time taken to depolarise to the activation threshold of Nav1.5 channels. Membrane hyperpolarisation is also predicted to reduce peak depolarisation during phase 0 of the action potential.\textsuperscript{278}

### 3.3.5 $SK3^{T/T}$ mice have increased repolarisation heterogeneity

$SK3^{T/T}$ mice do not display significant differences in the overall ventricular APD\textsubscript{80} as compared to wild type controls. Consistent with this observation, the ventricular ERP is not altered in $SK3^{T/T}$ mice during programmed stimulation. As discussed in Section 1.2.1.2, the ERP is related to the period of inactivation of the Nav1.5 channels following depolarisation.\textsuperscript{22} During this period, an action potential cannot be evoked by a stimulus. Recovery of Nav1.5 channels is dependent upon membrane repolarisation. Therefore, the APD and ERP are related.

Interestingly, while the overall APD\textsubscript{80} is not altered in $SK3^{T/T}$ mice, there is evidence of marked ventricular repolarisation heterogeneity. These results imply that in addition to its putative role during phase 4 hyperpolarisation, the SK3 channel may contribute to the terminal phase of ventricular repolarisation. A potential explanation for the observed repolarisation heterogeneity is that $SK3^{T/T}$ mice have non-uniform
overexpression of the SK3 channel, which may result in heterogeneous upregulation of the $I_{KCa}$ current. However, at this stage this mechanism remains speculative.

In a recent study, Chua et al demonstrated that in a rabbit heart failure model, upregulation of the $I_{KCa}$ current increases repolarisation heterogeneity as well as increasing susceptibility to recurrent ventricular arrhythmia. Similar results have also been reported in rat models with acute myocardial infarction. Interestingly however, despite marked ventricular repolarisation heterogeneity, we did not observe spontaneous ventricular arrhythmias in SK3$^{T/T}$ mice during ambulatory monitoring.

3.3.6 **SK3$^{T/T}$ mice have an increased susceptibility to AF and atrial arrhythmias**

SK3$^{T/T}$ and SK3$^{+/T}$ mice have inducible AF and atrial arrhythmias during programmed electrical stimulation. These findings suggest that overexpression of the SK3 channel creates a proarhythmogenic substrate in the atrium. Of note, SK3 channel overexpression was less pronounced in the atria of SK3$^{T/T}$ and SK3$^{+/T}$ mice relative to that observed in the ventricles. By implication therefore, relatively subtle alterations of the $I_{KCa}$ current may increase susceptibility to AF and atrial arrhythmias.

The prevailing theory regarding the mechanism of AF posits that the arrhythmia arises due to an interaction between ectopic triggers and a susceptible atrial substrate, which may foster reentry. During reentry, the action potential wavefront circulates continuously around a zone of conduction block. The zone of conduction block may arise in structurally normal tissue with functional alterations in conduction properties.
The maintenance of reentry is dependent upon the impulse propagating around the circuit slowly enough for all regions to regain excitability. Therefore, a slower conduction velocity increases the likelihood maintaining reentry. Reentrant arrhythmias are also influenced by the refractory period of the tissue. The shorter the refractory period, the greater the speed of recovery and hence the higher the chance of maintaining reentry.

The mechanism of AF in $SK3^{+/T}$ and $SK3^{+/T}$ mice remains to be elucidated. The atrial ERP is not significantly altered in $SK3^{+/T}$ and $SK3^{+/T}$ mice as compared to wild type controls. Therefore alterations in the refractory period are unlikely to underlie arrhythmia susceptibility in these mice. It is however plausible that in keeping with the findings in the ventricle, overexpression of the SK3 channel in the atrium promotes reentry by augmenting repolarisation heterogeneity and slowing conduction. The triggers for the initiation of AF are likely to originate from ectopic foci in the atrium. During ambulatory monitoring, we observed multiple spontaneous atrial ectopics in $SK3^{+/T}$ mice. It is important to note however that we did not measure repolarisation heterogeneity or conduction velocity in the atria of $SK3^{+/T}$ mice and therefore at this stage, this proposed mechanism has not been validated.

Consistent with the findings in the present study, upregulation of the $I_{KCa}$ current has been reported to increase susceptibility to AF. Ozgen et al demonstrated that upregulation of $I_{KCa}$ in response to intermittent burst pacing creates a
proarrhythmogenic substrate in the pulmonary veins in an *in vitro* model of isolated rabbit atria.\(^{214}\) Further, pharmacological blockade of SK channels has been reported to suppress AF. Diness *et al* demonstrated that SK channel inhibitors such as UCL1684, \(N\)-(pyridin-2-yl)-4-(pyridin-2-yl) thiazol-2-amine, and NS8593 suppress AF by prolonging the AERP.\(^{281}\) As discussed above, prolongation of the AERP is predicted to terminate reentrant arrhythmias as it increases the probability of the circulating wavefront encountering refractory tissue. The same group also demonstrated that the SK channel inhibitors, NS8593 and UCL1684 suppress AF in murine models with hypertension-induced atrial remodelling.\(^{282}\)

Interestingly, in contrast to the above findings, Li *et al* demonstrated that cardiac-specific knockout of the SK2 channel in a mouse model is associated with inducible AF on programmed electrical stimulation.\(^{212}\) It is likely however that the mechanism of AF in these mice is distinct to that proposed for SK3\(^{T/T}\) mice. SK2 channel knockout is associated with a prolongation of the APD with associated triggered activity. Therefore, the mechanism of AF in SK2 knockout mice can be thought of as an ´atrial torsades de pointes´. Overall, these results indicate that the SK channels play an important role in maintaining electrical stability in the atrium and both gain and loss of channel function can increase susceptibility to atrial arrhythmias.
3.3.7 Limitations

3.3.7.1 Mouse model of AF

We obtained a mouse strain ($SK3^T$) harbouring a tetracycline-regulatory cassette at the $Kcnn3$ locus which enables overexpression or inducible knockout of the gene.$^{223}$ While the $SK3^T$ mouse line demonstrated $Kcnn3$ overexpression at baseline, we were unable to knockout gene expression despite administration of adequate doses of doxycycline. The antibiotic resulted in some suppression of $Kcnn3$ expression however complete elimination of gene expression was not achieved. As a result, we were unable to characterise the functional effect of two extremes of SK3 channel expression. The cause for failure to induce gene knockout is presently unclear.

In addition to the limitations specific to the $SK3^T$ mouse strain, one of the inherent limitations of using a mouse model to study cardiac arrhythmia is that there exist significant differences between human and mouse cardiac electrophysiology. For instance, there are marked differences in repolarisation and resting heart rate. Further, the size of the mouse heart also influences susceptibility to reentrant arrhythmias. Despite these limitations, we believe that the mouse model is a valuable tool to investigate the role of the SK3 channel in cardiac electrophysiology.
3.3.7.2 Incomplete mechanistic data

An important limitation of the present study is that the mechanistic link between SK3 channel overexpression and the observed arrhythmias is not fully characterised. While it is plausible that an augmented $I_{KCa}$ current causes atrioventricular block and ventricular conduction slowing by mediating afterhyperpolarisation, the $I_{KCa}$ current and action potential were not measured directly. Further, we did not characterise how SK3 channel overexpression influences the atrial action potential. Therefore, cellular electrophysiology data is necessary to fully characterise the causal link between SK3 channel overexpression and cardiac arrhythmogenesis.

3.3.8 Next steps

3.3.8.1 Measurement of APD and $I_{KCa}$ current in SK3$^{T/T}$ mice

As discussed above, one of the limitations of the present study is that the action potential and $I_{KCa}$ current have not been measured directly in SK3$^{T/T}$ mice. Therefore, an important next step is to isolate cardiomyocytes from the conduction system, the atria and the ventricles in order to measure the action potential using patch clamp electrophysiology. The $I_{KCa}$ current can be measured by performing whole-cell recordings and calculating the difference between the total current observed at baseline and the current observed after application of apamin. As discussed previously, apamin is a selective blocker of SK channels.
3.3.8.2 Characterisation of the mechanistic link between SNPs at the 1q21 locus and KCNN3 expression

A GWAS of patients with lone AF identified an association signal in the KCNN3 gene at the 1q21 locus. In the present study, we demonstrate that overexpression of Kcnn3 in a mouse model increases susceptibility to AF. These findings support the notion that KCNN3 is the causative gene at the 1q21 locus. It is important to note however that in isolation, functional studies in animal models are insufficient to conclusively assign causation to candidate genes at GWAS loci. More comprehensive post-GWAS analysis is necessary to identify the mechanistic link between the risk SNPs at this locus and KCNN3 function.

The risk variants at the 1q21 locus do not map to alterations in the amino acid structure of KCNN3.\textsuperscript{184} It is therefore unlikely that altered KCNN3 protein function underlies the observed GWAS association. Many variants that confer a low risk of disease for complex traits like AF influence disease risk by altering the quantity of gene expression.\textsuperscript{283} Therefore, it is possible that the risk variants at the 1q21 locus mediate their effect at a transcriptional level, either through altered function of the promoter or other regulatory elements such as enhancers, silencers or insulators.

Non-coding regulatory sequences are typically highly conserved through evolution. Therefore, an effective strategy for the identification of these sequences is to use bioinformatic techniques that assess phylogenetic conservation. Potential regulatory
elements can then be characterised further using cell-based and \textit{in vivo} assays in fish and mice.\textsuperscript{284} The ultimate goal is to link the regulatory elements back to AF pathogenesis by determining whether the GWAS risk SNPs modulate \textit{KCNN3} expression by altering the function of these regulatory elements.
4.1 Introduction

GWAS have successfully identified numerous variants underlying a range of common diseases, including AF. Despite these successes however, the reported variants only account for a small fraction of the heritability of these diseases.\textsuperscript{215} In recent years, evidence has emerged to suggest that a significant proportion of the remaining or ‘missing’ heritability is likely to be accounted for by rare variants. Therefore, rare and deleterious variants in isolated pedigrees could potentially be major drivers of the heritability of AF.

Pedigree-based genetic studies may represent an effective strategy for the identification of some of the missing heritability of AF. Exome sequencing has recently emerged as a potentially powerful technique for the identification of causative mutations in Mendelian pedigrees. Exome sequencing uses next-generation sequencing technology to perform simultaneous sequencing of the entire protein coding region of the genome. Despite the fact that the exome consists of only 1\% of the genome, the majority of causative mutations underlying familial disorders are predicted to reside in the coding region of the genome.\textsuperscript{218}

Exome sequencing typically identifies in excess of 30,000 genetic variants per individual. Therefore the identification of the causative mutation can be a
Almost a third of the variants identified by exome sequencing are nonsynonymous, disrupt splice sites, or result in insertions or deletions. Therefore, the identification of a variant that appears to be pathogenic is not sufficient to assign causation. A number of filtering strategies are necessary to narrow the search for the causative mutation.

In large pedigrees with multiple affected individuals, a combinatorial approach using linkage analysis and exome sequencing represents a potentially effective strategy for filtering of the number of candidate variants. Linkage analysis is a technique designed to map the chromosomal location of mutations that underlie Mendelian diseases. The basic principle of linkage analysis involves investigating the co-segregation of a disease locus with a series of markers of known position distributed throughout the genome. Markers that segregate with the disease phenotype, and hence lie in close proximity to the disease locus, are used to define the boundaries of a disease locus. When used in combination with exome sequencing, linkage analysis has the potential to narrow the search for the causative mutation to a handful of variants.

We investigated a large, multigenerational family with autosomal dominant AF (AF-325) from the MGH-AF study by performing exome sequencing and linkage analysis. We hypothesised that using this combinatorial approach will lead to the identification a novel mutation for AF.
4.2 Results

4.2.1 Clinical characteristics

We evaluated a total of 57 individuals from pedigree AF-325, a two-generation family with an autosomal dominant pattern of inheritance of AF (Figure 46). Nine members of the pedigree were affected. The proband (III-1) has three siblings (III-3, III-9, III-11), three cousins (III-13, III-15, III-18) and two descendents (IV1 and IV2) with AF. Clinical characteristics of affected family members are summarised in Table 10. More detailed medical histories are outlined below.

The proband (III-1) is 73 years of age. She was diagnosed with AF at the age of 58. Despite multiple direct current (DC) cardioversions and treatment with antiarrhythmic drug therapy, she remained in persistent and symptomatic AF. She eventually underwent an atrioventricular node ablation and implantation of a permanent pacemaker. She has a past history of hypertension. Her baseline ECG demonstrated a paced rhythm.

Individual IV-1 was diagnosed with AF at the age of 37 years. He has a history of recurrent palpitations since the age of 25 years. He was treated with multiple antiarrhythmic agents, all of which were ineffective. He therefore underwent two pulmonary vein isolation procedures and has remained in sinus rhythm ever since. The baseline ECG demonstrated sinus rhythm with a resting heart rate of 53 beats per
Chapter 4. Identification of a Novel Locus for AF

minute. The PR interval was 172 ms, the QRS interval was 92 ms, and the corrected QT interval was 395 ms. The R wave axis was 86°. An echocardiogram demonstrated normal cardiac structure and function (EF - 65%).

Individual IV-2 died suddenly at the age of 27 years. According to her previous medical records, she developed AF at a young age. However the exact details are not available. The cause of sudden death was not ascertained. Clinical examination and toxicology screens did not reveal any abnormal findings.

Individual III-3 is 70 years of age. She developed paroxysmal AF at the age of 55 years. In 2011, she underwent a pulmonary vein isolation procedure and has remained in sinus rhythm since. Past medical history includes hypertension and hypercholesterolaemia. The baseline ECG demonstrated sinus rhythm with a rate of 64 beats per minute. The PR interval was prolonged at 272 ms. The QRS duration was 110 ms and the corrected QT interval was 472 ms. The R wave axis was 78°. An echocardiogram demonstrated an EF of 65% with mild left ventricular hypertrophy and mild diastolic dysfunction.

Individual III-9 is 63 years of age. He was diagnosed with AF at the age of 60. His baseline ECG demonstrated AF with a resting heart rate of 132 beats per minute. The QRS duration was 106 ms and the corrected QT interval was prolonged at 500 ms. The
Chapter 4. Identification of a Novel Locus for AF

R wave axis was \(-78^\circ\). His past history included severe chronic obstructive airways disease.

Individual III-11 is 58 years of age. She was diagnosed with paroxysmal AF at the age of 49. She has a history of recurrent palpitations since the age of 20 years. She is currently being treated with flecainide. Her baseline ECG demonstrated sinus rhythm with a resting heart rate of 62 beats per minute. The PR interval was prolonged at 242 ms. The QRS duration was 112 ms and the corrected QT interval was 466 ms. The R wave axis was \(-59^\circ\).

Individual III-13 is 86 years of age. He was diagnosed with AF at the age of 63 years. He also had a pacemaker implanted for bradycardia at the time. At age 18, when he attempted to enlist for the army, he was deemed unfit due to ´skipped beats´ however he was not given a formal diagnosis. He also has a history of ischaemic heart disease with multiple previous myocardial infarctions. His echocardiogram demonstrated impaired left ventricular systolic function with an EF of 20 %. His baseline ECG demonstrated a paced rhythm.

Individual III-15 is 78 years old. He was diagnosed with AF in his 60’s. He had a pulmonary vein isolation procedure for AF in 2008. His baseline ECG demonstrated sinus rhythm with a resting heart rate of 73 beats per minute. His PR interval was 214 ms, the QRS duration was 98 ms, and the corrected QT interval was 462 ms. The R
wave axis was -39°. An echocardiogram demonstrated a structurally normal heart (EF - 70 %).

Individual III-18 is 77 years of age. She was diagnosed with paroxysmal AF at the age of 70. She had been experiencing palpitations since her 20’s. She is currently being treated with sotalol. She has a background history of coronary artery disease and peripheral vascular disease. Her baseline ECG demonstrated sinus rhythm with a resting heart rate of 62 beats per minute. The PR interval was 170 ms, the QRS duration was 154 ms and the corrected QT interval was prolonged at 507 ms. She had right bundle branch block at baseline and her R wave axis was 68°.

The remaining members of family AF-325 do not have AF. Of note however, a number of the individuals from generation IV (IV-5, IV-6, IV-10 and IV-13) have a history of recurrent palpitations. IV-5, IV-6 and IV-10 have had documented regular narrow complex tachycardias on ECG. An electrophysiology study in IV-6 demonstrated an atrial tachycardia while IV-10 had an atrioventricular nodal reentrant tachycardia. IV-14 had two previous ablations however the details of the procedures are not available. IV-13 had recurrent palpitations however a tachycardia has not previously been documented on ECG. Despite the history of arrhythmias, IV-5, IV-6, IV-10 and IV-13 do not have documented episodes of AF and were therefore not considered as being definitively affected.
Figure 46: Pedigree of family AF-325. AF cases are depicted in black, obligate carriers are indicated in grey and unaffected individuals are depicted in white.
Table 10: Clinical characteristics of AF cases in family AF-325

<table>
<thead>
<tr>
<th>Pattern of AF</th>
<th>Age at onset</th>
<th>PMH</th>
<th>ECHO (EF%)</th>
<th>ECG</th>
</tr>
</thead>
<tbody>
<tr>
<td>III-1</td>
<td>Paroxysmal</td>
<td>58</td>
<td>Hypertension, PPM</td>
<td>Paced</td>
</tr>
<tr>
<td>IV-1</td>
<td>Paroxysmal</td>
<td>37</td>
<td></td>
<td>Normal</td>
</tr>
<tr>
<td>IV-2</td>
<td>Paroxysmal</td>
<td>27</td>
<td>Sudden death</td>
<td></td>
</tr>
<tr>
<td>III-3</td>
<td>Paroxysmal</td>
<td>55</td>
<td>Hypertension, PVI</td>
<td>65 %</td>
</tr>
<tr>
<td>III-9</td>
<td>Paroxysmal</td>
<td>60</td>
<td></td>
<td>Normal</td>
</tr>
<tr>
<td>III-11</td>
<td>Paroxysmal</td>
<td>49</td>
<td></td>
<td>Prolonged PR</td>
</tr>
<tr>
<td>III-13</td>
<td>Chronic</td>
<td>63</td>
<td>Hypertension, IHD, ICD</td>
<td>20 %</td>
</tr>
<tr>
<td>III-15</td>
<td>Paroxysmal</td>
<td>60</td>
<td>PVI</td>
<td>70 %</td>
</tr>
<tr>
<td>III-18</td>
<td>Paroxysmal</td>
<td>70</td>
<td>IHD</td>
<td>RBBB</td>
</tr>
</tbody>
</table>

Abbreviations; AF, atrial fibrillation; PPM, permanent pacemaker; IHD, ischaemic heart disease; ICD, implantable cardiac defibrillator; RBBB, right bundle branch block.

4.2.2 Linkage analysis

Linkage studies involve an analysis of the co-segregation of a disease phenotype with a series of markers of known position distributed throughout the genome. The identification of markers that are closely linked to the disease locus allows one to delineate a region of the genome that harbours the causative mutation. The most efficient statistical test to assess linkage between a disease locus and a marker locus is the logarithm of the odds (LOD) score. Two-point linkage analysis involves an analysis of linkage between individual markers and the disease locus. Multipoint linkage analysis on the other hand involves simultaneous analysis of numerous loci and the disease locus. Multipoint linkage analysis is a more powerful technique for the detection of linkage as compared to two-point linkage analysis.
Chapter 4. Identification of a Novel Locus for AF

Linkage analysis was performed in 19 individuals from family AF-325. Eight individuals were affected while 11 were unaffected. The individuals included in linkage analysis are circled in red in Figure 47. Linkage analysis was performed using both SNP markers and microsatellite markers. The SUPERLINK online program was used to perform the analysis. LOD scores were computed under the following assumptions; 1) an autosomal dominant pattern of inheritance, 2) complete penetrance of the causative variant, 3) a disease allele frequency of 0.0001, and 4) no phenocopies.

### 4.2.3 SNP linkage analysis

SNP genotyping for linkage analysis was performed using the Affymetrix Genome-wide Human SNP array 6.0 (Affymetrix, CA, USA) which includes a total of 906,600 SNPs. The genotype call rate for SNPs was 98%. 0.9% displayed a likely transmission error and a further 0.4% displayed non-Mendelian transmission. Due to the excessively long computation times required to perform linkage analysis with a dense set of SNP markers, the total number of markers for two-point linkage analysis was filtered to 14,844.

The peak LOD score identified using two-point linkage analysis was 2.1. An extensive region was identified on chromosome 1p21.1-q22 (Figure 48A). The locus extends from chr1: 106,124,761 - 157,005,154. The locus is bounded by marker rs10785759 at the proximal end and marker rs1750802 at the distal end. Detailed results of two-point linkage analysis are included in Figure 48B.
Figure 47: Pedigree of family AF-325 demonstrating model for linkage analysis. Individuals included in linkage analysis are circled. Black circles indicate individuals without genotyping data while red circles indicate individuals with genotyping data.
Multipoint linkage analysis was performed with 1,506 SNP markers spanning chromosome 1. The peak multipoint LOD score on chromosome 1p21.1-q22 was 3.2. The locus extends from chr1: 106,171,030-155,433,942. The locus is bounded by marker rs2689982 at the proximal end and marker rs12724079 at the distal end (Figure 49).
Figure 48B: Graphs of two-point LOD score for each autosome from a genome-wide scan. The number of SNPs along the chromosome is plotted on the x-axis while the two-point LOD score is plotted on the y-axis.
4.2.4 Microsatellite linkage analysis

In order to replicate the results obtained by SNP linkage analysis, we performed two-point and multipoint linkage analysis using microsatellite markers distributed within and around the locus on chromosome 1. Two-point analysis was performed using seven microsatellite markers (D1S2868, D1S206, D1S2726, D1S252, D1S498, D1S484 and D1S196). Multipoint linkage analysis was subsequently performed with a set of four markers (D1S206, D1S2726, D1S252 and D1S484).

Two-point linkage analysis demonstrated a peak LOD score of 3.5 around marker D1S2726 on chromosome 1p21.1-q22. The results are summarised in Table 11. LOD
scores for Θ values of 0.00, 0.01, 0.05, 0.10, 0.20, 0.30 and 0.40 are listed. Significantly positive LOD scores are depicted in green, positive LOD scores are depicted in black and negative LOD scores are depicted in red.

**Table 11:** Two-point linkage analysis using 7 microsatellite markers

<table>
<thead>
<tr>
<th>Marker information</th>
<th>Recombination fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>Marker name</td>
</tr>
<tr>
<td>1</td>
<td>D1S2868</td>
</tr>
<tr>
<td>2</td>
<td>D1S206</td>
</tr>
<tr>
<td>3</td>
<td>D1S2726</td>
</tr>
<tr>
<td>4</td>
<td>D1S252</td>
</tr>
<tr>
<td>5</td>
<td>D1S498</td>
</tr>
<tr>
<td>6</td>
<td>D1S484</td>
</tr>
<tr>
<td>7</td>
<td>D1S196</td>
</tr>
</tbody>
</table>

Multipoint linkage analysis demonstrated a peak LOD score of 4.0 at marker D1S2726 (Table 12 and Figure 50). The proximal boundary was defined by marker D1S206 (chr1: 101,685,305) while the distal boundary was defined by marker D1S484 (chr1: 160,767,708).
Chapter 4. Identification of a Novel Locus for AF

Table 12: Multipoint linkage analysis using 4 microsatellite markers

<table>
<thead>
<tr>
<th>Marker names:</th>
<th>Trait position from marker D1S206 (in CM)</th>
<th>LOD-SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-10.00000</td>
<td>2.0457</td>
</tr>
<tr>
<td></td>
<td>-5.00000</td>
<td>2.0388</td>
</tr>
<tr>
<td>D1S206</td>
<td>0.00000</td>
<td>-2.5900</td>
</tr>
<tr>
<td></td>
<td>6.15000</td>
<td>3.5078</td>
</tr>
<tr>
<td>D1S2726</td>
<td>12.30000</td>
<td>4.0077</td>
</tr>
<tr>
<td></td>
<td>14.80000</td>
<td>3.7866</td>
</tr>
<tr>
<td>D1S252</td>
<td>17.30000</td>
<td>3.5678</td>
</tr>
<tr>
<td></td>
<td>26.30000</td>
<td>2.1667</td>
</tr>
<tr>
<td>D1S484</td>
<td>35.30000</td>
<td>-4.7123</td>
</tr>
<tr>
<td></td>
<td>40.30000</td>
<td>-1.3871</td>
</tr>
<tr>
<td></td>
<td>45.30000</td>
<td>-0.7339</td>
</tr>
</tbody>
</table>

Abbreviations; cM, centimorgan; LOD, logarithm of the odds score

Figure 50: LOD score graph from multipoint analysis using 4 microsatellite markers.
A locus on chromosome 1 with a LOD score of 4 was identified at marker D1S2726. The LOD score is plotted on the y-axis and the distance in centomorgans (cM) from marker D1S206 is plotted on the x-axis.
4.2.5 Exome sequencing

A potentially powerful technique for the identification of causative mutations in large Mendelian families is to use a combination of linkage analysis and exome sequencing. As discussed in Section 2.2.6, exome sequencing involves simultaneous sequencing of the entire protein-coding region of the genome. The exome consists of approximately 1% of the genome however harbours up to 85% of rare disease causing mutations. Therefore exome sequencing represents an efficient strategy for the identification of causative mutations in Mendelian families.

Exome sequencing was performed in one affected individual from family AF-325 (individual IV-1). The Illumina HiSeq200 was used to perform exome sequencing. Approximately 5.9 Gb of sequencing data was generated as paired-end 100 base pair reads. Reads with duplicate start sites were discarded. A total of 63,693,712 reads passed quality control and of these, 95.2% were aligned to the reference genome (hg19 NCBI Build 37, 2009). A mean depth of coverage of 67x was achieved across the samples. A minimum threshold of 20x coverage was set for variant calling. 95% of bases fulfilled the coverage criterion for variant calling. 18,318 variants were identified, of which 2,535 were novel.

A number of different criteria were used to differentiate potential disease causing variants from other variants. Firstly, all synonymous variants and variants that have previously been reported in population databases (dbSNP, 1000 Genomes Project,
exome variant server) were excluded from downstream analysis on the basis of the assumption that Mendelian diseases are caused by novel nonsynonymous variants. Using this strategy, the number of candidate variants in individual IV-1 was filtered to 1,653 missense variants, 28 nonsense variants and 32 insertion-deletion variants.

Secondly, given that a strong locus for AF was identified on chromosome 1p21.1-q22, all variants that were located outside the boundaries of this locus were excluded from downstream analysis. As discussed in the previous section, the boundaries of the locus extend from chr1: 103,193,348 - 155,610,326. Using this strategy, the list of candidate variants was filtered to five (listed in Table 13).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Mutation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1orf62</td>
<td>V84I</td>
<td>Chromosome 1 open reading frame 162</td>
</tr>
<tr>
<td>CELF3</td>
<td>T240P</td>
<td>CUGBP, Elav-like family member 3</td>
</tr>
<tr>
<td>PMVK</td>
<td>V96A</td>
<td>Phosphomevalonate kinase</td>
</tr>
<tr>
<td>MRPS21</td>
<td>L75V</td>
<td>Mitochondrial ribosomal protein S21</td>
</tr>
<tr>
<td>FLG</td>
<td>H1880N</td>
<td>Filaggrin</td>
</tr>
</tbody>
</table>

Thirdly, based on the assumption that the causative mutation in pedigree AF-325 is both necessary and sufficient to cause the disease phenotype, all variants that did not segregate with disease were excluded from downstream analysis. Segregation analysis of the five candidate variants at the chromosome 1p21.1-q22 locus was performed
using direct Sanger sequencing. As demonstrated in Table 14, none of the variants at
the locus segregated with disease. Overall therefore, the causative mutation in
pedigree AF-325 remains elusive.

**Table 14**: Transmission of variants at the chromosome 1 locus in individual IV-1

<table>
<thead>
<tr>
<th>Individual</th>
<th>Genetic variant</th>
</tr>
</thead>
<tbody>
<tr>
<td>III-1</td>
<td><em>C1orf62</em> (V84I)</td>
</tr>
<tr>
<td>IV-1</td>
<td>Yes</td>
</tr>
<tr>
<td>III-3</td>
<td>No</td>
</tr>
<tr>
<td>III-9</td>
<td>No</td>
</tr>
<tr>
<td>III-11</td>
<td>Yes</td>
</tr>
<tr>
<td>III-13</td>
<td>No</td>
</tr>
<tr>
<td>III-15</td>
<td>No</td>
</tr>
<tr>
<td>III-18</td>
<td>No</td>
</tr>
</tbody>
</table>

4.3 Discussion

4.3.1 Summary of main findings

The majority of causative mutations underlying Mendelian disorders are predicted to
reside in the protein coding region of the genome. We therefore applied a two-step
approach combining linkage analysis and exome sequencing in an attempt to identify
a potential coding mutation in pedigree AF-325. Linkage analysis using an Affymetrix
SNP 6.0 array with a highly conservative analysis identified a novel AF locus on
chromosome 1p21.1-q22. Multipoint linkage mapping using 7 microsatellite markers across the region confirmed these findings. The locus spans an extensive region (approximately 30 cM) with a peak LOD score of 4.0 at marker D1S2726. Exome sequencing of one affected family member identified five candidate coding mutations within the boundaries of the locus. However, none of the variants segregated with disease. Therefore, the causative mutation in this pedigree remains elusive.

4.3.2 Potential explanations for failure to identify causative mutation in AF-325

There are two potential explanations for the failure to identify the causative mutation in pedigree AF-325; the first is that the disease phenotype is caused by a non-coding mutation at the chromosome 1 locus and is therefore not amenable to discovery using exome sequencing, and the second is that the phenotype is caused by a coding variant however the variant has not been identified due to technical limitations or the strategy used for exome sequencing and/or linkage analysis. The following section discusses these possibilities in more detail and outlines potential approaches for the identification of the causative variant.

4.3.2.1 A non-coding mutation as a potential cause of AF

More than 98 % of the genome consists of non-coding sequences of DNA. The non-coding part of the genome has traditionally been regarded as non-functional ´junk DNA´. However increasing evidence has emerged over the past few decades to
demonstrate that non-coding sequences play critical roles in the regulation of gene expression.

A number of different classes of non-coding regulatory elements have been annotated in the human genome. *Cis*-acting regulatory elements include promoters, enhancers, silencers and insulators. Promoters are located at the 5′ ends of the genes that they regulate while the other *cis*-acting regulatory elements may be located at significant distances from their target genes. miRNAs are an important group of *trans*-acting regulatory elements. miRNAs regulate the expression of remotely located genes by causing mRNA silencing.

It is plausible that a mutation in a non-coding regulatory element at the chromosome 1 locus underlies the disease phenotype in pedigree AF-325. The mutation may be a single nucleotide mutation or a structural variant. Potential structural variants include CNVs, inversions and translocations. As discussed in Section 1.3.3.2, CNVs are defined as segments of DNA (1 kb or larger) that are present in variable copy number as compared to the reference genome. Inversions arise when a chromosome breaks in two places and the broken fragment rotates 180 degrees before rejoining the end fragments. Translocations are genetic rearrangements which arise due to exchanging of DNA segments between heterologous chromosomes.
Chapter 4. Identification of a Novel Locus for AF

From a mechanistic perspective, a mutation in a non-coding regulatory element could increase AF susceptibility in pedigree AF-325 by altering the dosage of genes that are critical for maintaining electrical stability in the atrium. Examples include genes that encode ion channels and gap junction proteins.\textsuperscript{117} Potential mechanisms underlying altered gene expression include; 1) disruption of the interaction between cis-acting regulatory elements and trans-acting proteins, 2) alterations in the copy number of regulatory elements, and 3) alterations in the interaction of miRNAs with target mRNA transcripts. These mechanisms are discussed in more detail here.

The interaction between cis-acting regulatory elements and transcription factor proteins is an important step during transcription initiation. Transcription factor proteins bind at specific DNA sequences in the promoter region of genes. Single nucleotide mutations or deletions affecting these sequences may disrupt transcription factor binding which in turn could alter target gene expression.\textsuperscript{290} Larger structural variants also have the potential to alter interactions between regulatory elements and target genes.\textsuperscript{288} For instance CNVs, reciprocal translocations and inversions may cause a physical dissociation of transcriptional units from cis-acting regulatory elements by modifying the spatial organisation of the genomic interval. This phenomenon is known as a ‘position effect’.\textsuperscript{291}

As discussed above, CNVs are segments of DNA that are present in variable copy number as compared to the reference genome. CNVs may be associated with either a
loss (deletion) or gain (duplication) of genomic sequence. If a non-coding regulatory element maps to a CNV, an increased copy number of the element may lead to target gene dosage effects. The target gene may lie within or outside the boundaries of the CNV. CNVs have previously been demonstrated to affect the expression of genes up to several hundred kilobases from the genetic rearrangement.292

The interaction between miRNAs and target transcripts can be altered either by mutations in the miRNAs themselves or by mutations in the 3´ untranslated region (UTR) on target mRNA transcripts. Given that miRNAs predominantly act by repressing translation, the majority of miRNA mutations that disrupt binding would be predicted to increase target gene expression. Mutations in the 3´ UTR of mRNA have the potential to either generate de novo recognition sites for miRNA or result in the removal of recognition sites.293 Therefore, 3´ UTR mutations could be associated with gain-of-function or loss-of-function effects.

Relatively little is currently known about the role of non-coding regulatory elements in Mendelian diseases. This is primarily due to the fact that a large proportion of non-coding regulatory elements have yet to be recognised and non-coding mutations are not commonly screened for in Mendelian diseases. However increasing evidence is emerging from isolated pedigrees implicating regulatory element mutations.294 The majority of these mutations have been demonstrated to disrupt interactions between promoters and other regulatory elements.73
Single nucleotide mutations in cis-acting regulatory elements have been reported to underlie familial conditions such as Werner mesomelic syndrome, Pierre Robin syndrome and polydactyly.\textsuperscript{294-297} Gross deletions in non-coding regions that harbour regulatory elements have been reported in the familial acampomelic form of campomelic dysplasia and blepharophimosis syndrome.\textsuperscript{298,299} CNVs with either duplications or deletions have been reported to underlie Leri-Weill syndrome, Mesomelic dysplasia Kantaputra, Cooks syndrome, Pierre Robin syndrome, Pelizaeus–Merzbacher disease as well as various types of familial brachydactyly and polydactyly.\textsuperscript{300}

Single nucleotide mutations in miRNA have also been reported to underlie Mendelian diseases. For instance, a miR-96 mutation has been identified in DFNA50, an autosomal dominant form of deafness.\textsuperscript{301} Sequence variants in miRNA binding sites in the 3’ UTR of the SLITRK1 gene have been reported in familial Tourette’s syndrome.\textsuperscript{302} Further, 3’ UTR mutations in the REEP1 gene have also been identified in familial spastic paraplegia.\textsuperscript{303} These mutations are predicted to alter the target site for miR-24 and miR-140 respectively.

Based on the evidence outlined above, it is plausible that non-coding mutations that alter the regulatory landscape of the genome underlie AF of in pedigree AF.\textsuperscript{325} Potential strategies for the identification of these variants are discussed in Section 4.3.4.1.
4.3.2.2 Potential reasons for failure to identify a causative coding mutation

An alternative hypothesis to the one presented above is that the disease phenotype in pedigree AF-325 is caused by a coding mutation however the variant has not been uncovered due to the approach used for exome sequencing and/or linkage analysis. Specific reasons for the failure to uncover a potential coding variant include; 1) incorrect assumptions regarding the disease-causing variant, 2) incorrect assumptions regarding the disease model, and 3) technical errors relating to exome sequencing. These possibilities are discussed in more detail in the following sections.

4.3.2.2.1 Incorrect assumptions during analysis of exome sequencing data

Exome sequencing in the proband of pedigree AF-325 identified in excess of 18,000 genetic variants. The strategy for filtering of candidate variants was based on a number of assumptions regarding the disease-causing variant. Firstly, a single mutation was predicted to be both sufficient and necessary to cause the disease phenotype. Therefore, only variants that segregated completely with disease were deemed to be potentially pathogenic. Secondly, the mutation was predicted to be a rare variant that is private to the affected members of the pedigree. Therefore, variants that have previously been reported in publically available databases were excluded from downstream analysis. Thirdly, the mutation was predicted to cause an alteration in the amino acid sequence of the protein i.e. nonsense, insertion-deletion and missense. Therefore synonymous variants were excluded from further analysis. Finally, the mutation was predicted to reside within the AF locus identified by linkage
analysis. Therefore, only variants located within the boundaries of the locus on chromosome 1 were included in further analysis.

It is important to note that the above assumptions could result in the causal variant being incorrectly excluded from further analysis. For instance, the assumption that the disease causing mutation results in a change in amino acid sequence may be incorrect. A synonymous variant could potentially influence the disease phenotype by altering transcription, translation or splicing of a gene. The assumption that variants that are present in public databases are not pathogenic may also be incorrect. It is possible that rare variants in these databases are actually disease-causing or represent false positives. Based on our genetic model, only variants that segregated completely with disease were considered to be potentially pathogenic. It is possible however that this is too stringent a criterion and not all affected individuals in pedigree AF-325 are carriers of the same mutation due to genetic heterogeneity. Finally, it is possible that the AF locus identified on chromosome 1 is a false positive result and that the disease causing mutation is in fact located outside the boundaries of the locus. This possibility is discussed in more detail in the next section.

4.3.2.2 Incorrect assumptions during linkage analysis

Linkage analysis in pedigree AF-325 was performed under the following assumptions; 1) autosomal dominant mode of inheritance, 2) penetrance of the disease causing variant of 0.99, 3) disease allele frequency of 0.0001, 4) no phenocopies or sporadic
Chapter 4. Identification of a Novel Locus for AF

mutations, and 5) no linkage disequilibrium between the disease locus and the marker locus. It is important to note that misspecification of these parameters could potentially lead to incorrect conclusions about linkage. Additionally, misspecification of allele frequency at marker loci, pedigree errors and genotyping errors have the potential to lead to incorrect conclusions about linkage.\textsuperscript{305}

The risk of misspecification of the majority of the aforementioned parameters in pedigree AF-325 is low. Firstly, because AF-325 is a large pedigree with multiple affected individuals, the pattern of inheritance could be determined with a high degree of certainty. It was also possible to exclude pedigree errors and genotyping errors on the basis that they do not follow the laws of Mendelian inheritance. Secondly, running a conservative analysis ensured that errors due to misspecification of the penetrance of the causative variant were minimised. Thirdly, the frequencies of the marker alleles in the Affymetrix Genome-wide Human SNP array 6.0 and microsatellite repeat markers were known.

Of note however, not all potential sources of error could be minimised. For instance, it is possible that the disease allele and marker alleles are in linkage disequilibrium. False positive linkage results can arise due to incorrect assumptions regarding linkage equilibrium, especially with the dense set of markers used for SNP linkage analysis. The assumptions regarding disease allele frequency and phenocopies also represent potential sources of error.
Overall, linkage analysis identified a compelling locus for AF on chromosome 1 in pedigree AF-325. However, there remains a possibility that this is a false positive result. By implication therefore, it is plausible that the disease phenotype is caused by a coding variant which has not been detected because it lies outside the boundaries of the chromosome 1 locus.

4.3.2.2.3 Technical limitations of exome sequencing

In addition to issues relating to filtering of candidate variants, the causative mutation in pedigree AF-325 could have been missed due to technical limitations of exome sequencing. Two of the major limitations associated with exome sequencing are the inability to comprehensively detect variants such as CNVs and structural rearrangements and non-uniform coverage across the exome.\(^{306}\)

As discussed previously, CNVs are an important class of genetic variation in Mendelian diseases. CNVs have the potential to influence disease risk by altering the expression of dosage sensitive genes. A number of studies have demonstrated that transcripts that map to CNVs have significantly higher relative expression levels.\(^{292}\) While bioinformatic techniques for the identification of CNVs from exome sequencing data have been developed, these techniques are associated with significant limitations. The main reason is that captured exons are non-contiguous. The situation is further complicated by the complex nature of the relationship between read depth and copy number and technical issues such as GC content and batching of samples.\(^{307}\) On the
basis of these limitations, the variant calling strategy in the present study focused on
single nucleotide variants and small insertions and deletions. Therefore potential
structural variations affecting coding genes may have been overlooked.

The exome sequencing platform used in the present study has previously been
demonstrated to cover a very large proportion of the exome with a high level of
targeting efficiency. However, one of the inherent limitations of exome sequencing
is that coverage across the exome is not uniform and some areas have insufficient
coverage or may be missed altogether. Reduced coverage is a particular issue over
areas with high and low GC content. Therefore, if the causative mutation for
pedigree AF-325 maps to an area with high or low GC content, it is possible that it may
have been overlooked due to insufficient sequence depth. Of note however, we
demonstrated good coverage levels within the boundaries of the chromosome 1
locus.

4.3.3 Limitations

4.3.3.1 Mis-phenotyping in pedigree-based studies

Mis-phenotyping is one of the most commonly encountered problems when
performing linkage analysis and exome sequencing in monogenic families. This
problem is particularly relevant when investigating AF because the arrhythmia is often
paroxysmal and may be asymptomatic. Further, AF is associated with a relatively high
phenocopy rate in the general population. In an attempt to minimise the problems associated with mis-phenotyping, very stringent criteria for designating individuals with an affected or unaffected status were used. All other members of the family were coded as unknown for the purposes of linkage analysis. This degree of stringency would be predicted to significantly increase the chances of accurately identifying the disease locus.

4.3.3.2 Large genetic interval in pedigree AF-325

The genetic interval identified by linkage analysis in pedigree AF-325 is quite large. The extensive interval is a reflection of the conservative parameters used to perform linkage analysis. A conservative approach was employed to minimise errors due to mis-phenotyping. The large size of the locus did not significantly influence our ability to screen for potential coding variants using exome sequencing. We did not however identify a coding mutation within the boundaries of the disease locus.

As discussed above, it is plausible that the phenotype in pedigree AF-325 is caused by a non-coding mutation. A potential strategy for the identification of the non-coding variant is to perform whole-genome sequencing and to focus subsequent analysis to the chromosome 1 locus. Given the size of the locus however, this strategy is predicted to identify a very large number of non-coding variants. Therefore, identification of the causative mutation is likely to be a challenge.
Chapter 4. Identification of a Novel Locus for AF

4.3.4 Next steps

4.3.4.1 Genotyping for potential non-coding mutations

As discussed in Section 4.3.2.1, a mutation in a non-coding regulatory element at the chromosome 1 locus could underlie AF in pedigree AF-325. A potential next step for the identification of the putative non-coding mutation is to sequence the entire genomic interval within the boundaries of the chromosome 1 locus and subsequently search for mutations in functional regulatory regions. Given the falling costs of whole-genome sequencing, it would be feasible to sequence the entire genome in multiple affected individuals in pedigree AF-325 and to subsequently limit the analysis to the chromosome 1 locus.

The chromosome 1 locus spans an extensive genomic region (approximately 30 cM) and is therefore predicted to harbour a large number of non-coding variants. A number of different strategies can be used to filter the candidate variants. Firstly, variants that are predicted to be located in functional regulatory regions may be prioritised for downstream analysis. A number of online databases with comprehensive maps of cis-acting regulatory elements and miRNAs are currently available and may be used for variant prioritisation. Secondly, mutations that have not previously been reported in publically available databases, and are therefore deemed to be unique, may be prioritised for downstream analysis. Thirdly, only variants that segregate completely with disease may be included in downstream analysis. Finally,
mutations that have a high probability of affecting the function of a regulatory motif may be prioritised for further analysis. A number of predictive models are currently available to distinguish potentially deleterious mutations from silent variants. \(^{309}\)

Once a manageable list of candidate non-coding variants has been identified, the next step would be to filter variants based on their functional effects in reporter assays. Potential techniques for functional analysis include cellular assays and transgenic reporter animals. Cell-based assays involve cloning of the regulatory element of interest into a reporter construct and analysing relative expression of the reporter gene between mutant and wild type constructs. \(^{310}\) Transgenic animal studies involve cloning of the regulatory element upstream of a reporter gene and introducing the construct into a fertilised embryo e.g. from a mouse or zebrafish. The embryos are subsequently analysed for relative expression of the reporter gene between the mutant and wild type construct.

The aim of the bioinformatic and functional filtering strategies discussed above would be to identify a single compelling non-coding mutation for the disease phenotype in pedigree AF-325. If such a variant is successfully identified, the next step would be to elucidate the mechanistic link between the variant and AF pathogenesis. Chromatin confirmation capture is a valuable technique for the characterisation of the long-range interaction profiles of \(cis\)-acting regulatory elements. The technique involves chromatin fixation, digestion with restriction enzymes, and intermolecular ligation to
identify genetic regions that are likely to interact with one another in left atrial tissue.\textsuperscript{311} Chromatin confirmation is predicted to identify target genes for \textit{cis}-acting regulatory elements and may therefore provide valuable insights into the mechanistic links between regulatory element mutations and AF pathogenesis. Murine models may also be used to elucidate the potential mechanistic link between both \textit{cis}-acting regulatory element and miRNA variants and AF pathogenesis. These studies involve the characterisation of the cardiac electrophysiological phenotype of mammalian models with deletion of the regulatory element. Studies in animal models represent an important strategy for functional validation of the role of a regulatory mutation in disease pathogenesis.

It is important to note that while the approach outlined above has the potential to uncover a non-coding mutation in pedigree AF-325, it is associated with a number of pitfalls. Firstly, as discussed above, whole-genome sequencing is predicted to identify a very large number of candidate variants and current filtering strategies are associated with significant limitations.\textsuperscript{312} For instance, while the online databases of regulatory elements are highly informative, the annotation of the regulatory landscape of the genome is far from complete.\textsuperscript{313} Therefore, candidate variants may be incorrectly excluded from downstream analysis. Further, the full spectrum of genetic variation in the non-coding region of the genome has yet to be characterised. Therefore, filtering for unique variants is predicted to be less effective. Secondly, the mechanisms underlying regulatory control of gene expression are diverse and are not
clearly understood. Therefore, interpretation of functional data may be challenging. Finally, current whole genome sequencing techniques are not effective for the identification of structural variations in the genome. Overall therefore, identification of a non-coding mutation in pedigree AF-325 may represent a significant challenge.

4.3.4.2 Screening for potential structural variants at the chromosome 1 locus

As discussed in Section 4.3.2.1, structural variants at the chromosome 1 locus could underlie the disease phenotype in pedigree AF-325. CNVs are by far the most common structural variants in the genome. While bioinformatic mining of exome or whole-genome sequencing data can be used to identify CNVs, these approaches are associated with significant limitations. A more effective strategy to identify CNVs is to use the microarray-based comparative genome hybridisation array (array CGH). Array CGH is a high resolution technique for the identification of CNVs both at a genome-wide level and at a locus specific level.

The basic principle of the array CGH involves co-hybridisation of differentially labelled test and reference DNA to an array with oligonucleotide probes for specific genetic targets. Copy number analysis is derived by measuring the hybridisation intensity of the sample of interest as compared with an average value from control subjects. The array CGH has more than 24 million probes for detection of CNVs in exonic and intergenic regions. In the case of pedigree AF-325, a custom designed CGH array with probes that target both coding genes and non-coding regulatory elements at the
chromosome 1 locus represents a potentially effective strategy for the identification of potential disease causing CNVs.

If a compelling candidate CNV is identified in pedigree AF-325, the next step would be to characterise the mechanistic link between the CNV and AF pathogenesis. A potential strategy for the elucidation of the mechanistic link is to investigate the phenotypic consequences of deletions or duplications of the genomic regions of interest in murine models. Further, for CNVs that are predicted to alter the function of non-coding regulatory elements, techniques such as genome-wide CHIP-seq may be necessary to establish the interaction profiles of the regulatory variants. It is important to note however, that functional validation of the role of a CNV in disease pathogenesis may be a challenge. This point is underscored by the fact that for most CNVs implicated in disease, the underlying pathogenic mechanism remains unclear.
CHAPTER 5.
IDENTIFICATION OF A NOVEL MUTATION FOR AF

5.1 Introduction

As discussed in the previous chapter, despite the successes of GWAS, a significant proportion of the heritability of AF has yet to be uncovered. Recently, research in AF genetics has focused once again on pedigree-based studies in an attempt to uncover some of the ‘missing’ heritability of AF. The emergence of exome sequencing has significantly enhanced our ability to identify rare mutations in Mendelian pedigrees. Exome sequencing uses next-generation sequencing technology to simultaneously sequence the entire protein coding region of the genome. Of note, despite the fact that the exome consists of only 1% of the genome, it harbours more than 85% of the mutations underlying Mendelian diseases.\textsuperscript{218} Therefore, exome sequencing represents a particularly efficient approach to investigate these diseases.

While exome sequencing is a powerful technique for rapid sequencing of all coding genes, one of the key challenges associated with the technique is that it identifies thousands of genetic variants per individual.\textsuperscript{222} A significant proportion of these variants are nonsynonymous, disrupt splice sites, or result in insertions or deletions. Therefore, assigning causation to a variant solely based on the fact that it appears to be pathogenic is not sufficient. Filtering and prioritisation of candidate variants has to be performed on multiple levels in order to narrow the search for the causative mutation.
As discussed in Chapter 4, linkage analysis represents a potentially effective technique for filtering of candidate variants identified by exome sequencing. Linkage analysis effectively narrows the search space for the causative mutation to a defined subsegment of the genome. We therefore applied a two-step approach combining linkage analysis and exome sequencing in an attempt to identify a potential coding mutation in pedigree AF-325, a multigenerational pedigree with autosomal dominant AF. We identified a novel AF locus on chromosome 1p21.1-q22. Exome sequencing of one affected family member identified a number of candidate mutations within the boundaries of the locus. However, none of the variants segregated with disease. Overall therefore, the strategy failed to identify the causative mutation.

Following on from the study in pedigree AF-325, we investigated a second multigenerational pedigree with a complex phenotype of AF and atrioventricular septal defects (AF-435). In view of the limitations associated with the combinatorial approach of linkage analysis and exome sequencing in one affected individual from AF-325, we modified our strategy in pedigree AF-435 and performed exome sequencing in three affected individuals. As discussed in Section 1.4.2.1, performing exome sequencing in multiple distally related affected family members and prioritising overlapping variants is a potentially effective alternative strategy for filtering of candidate variants. The further the genetic distance between affected family members, the fewer the genetic variants that they share. Further prioritisation of candidate variants is based on their potential functional role in biological pathways,
evolutionary conservation and expression profiles. Once a small set of promising candidate variants have been identified, segregation analysis and functional analysis maybe undertaken. The following chapter discusses the results from exome sequencing analysis in pedigree AF-435.

5.2 Results

5.2.1 Clinical characteristics of pedigree AF-435

We evaluated a total of 17 individuals from pedigree AF-435, a two generation family with a complex phenotype of AF and atrioventricular septal defects (Figure 51). The trait was inherited with an autosomal dominant pattern with variable penetrance. Five members of the pedigree were affected. The proband (II-13) had an atrial septal defect (ASD). Two of her descendants (III-15 and III-16) were diagnosed with early-onset AF. The proband’s sister (II-3) also had AF and one of her descendants (III-5) had a ventricular septal defect (VSD) during childhood. More detailed medical histories for the affected individuals are outlined below.

The proband for family AF-435 (II-13) initially came to the attention of cardiologists in 2007 at the age of 52. She presented with shortness of breath, palpitations and two episodes of near syncope. Auscultation of her heart revealed a systolic murmur. Her baseline ECG revealed sinus rhythm with isolated premature ventricular ectopics. Her PR interval was 150 ms, the QRS duration was 80 ms and the corrected QT interval
was 465 ms. The R wave axis was 4°. Cardiac catheterisation revealed unobstructed coronary arteries. She did however have evidence of an ASD. She underwent cardiac surgery to correct a primum ASD. During the postoperative period she had a paroxysm of AF but has otherwise been well since the procedure.

Individual III-15 was diagnosed with paroxysmal AF at the age of 14 years. He presented with a two-year history of intermittent palpitations. His baseline ECG demonstrated a PR interval of 160 ms, a QRS duration of 106 ms and a corrected QT interval of 406 ms. His R wave axis was 77°. His echocardiogram demonstrated a left ventricular EF of 55 %. There was no evidence of atrial or ventricular septal defects. He had a trial of multiple antiarrhythmic drugs which failed to suppress the paroxysms of AF. He therefore underwent an electrophysiology study and catheter ablation. A left sided atrial tachycardia which consistently degenerated to AF was induced during the study. Radiofrequency ablation was attempted on three separate occasions however he continues to experience paroxysms of AF.

Individual III-16 developed paroxysms of AF at the age of 16 years. He had a background history of two episodes of syncope at the ages of 4 and 10 years. An echocardiogram demonstrated normal left ventricular dimensions and an EF of 49 %. There was no evidence of atrial or ventricular septal defects. He was treated pharmacologically with diltiazem and metoprolol which controlled his symptoms.
Individual III-5 was born with a VSD. The defect was not operated on and he was followed-up by paediatric cardiologists for a few years. He had a history of short-lived palpitations however no documented episodes of AF. His baseline ECG demonstrated sinus rhythm with a PR interval of 144 ms, a QRS duration of 116 ms and a corrected QT interval of 400 ms. The R wave axis was 69°. An echocardiogram at the age of 20 years demonstrated a normally sized left ventricle with an EF of 60 %. There was no evidence of a residual VSD.

Individual II-3 had a history of one documented paroxysm of AF at the age of 49 years. Her baseline ECG demonstrated sinus rhythm with a PR interval of 162 ms, a QRS duration of 72 ms, and a corrected QT interval of 412 ms. The R wave axis was 3°. Her echocardiogram demonstrated a normal left ventricular size with an EF of 72 %. There was no evidence of atrial or ventricular septal defects.

None of the other members of the family had a history of atrioventricular septal defects or AF. Of note however, individuals II-5, II-17 and III-4 had abnormal ECGs with an altered P wave axis and incomplete right bundle branch block (RBBB). Individual III-4 has not previously had an echocardiogram and therefore it is unknown whether he has atrial or ventricular septal defects.
Figure 51: Pedigree of family AF-435. The phenotypic traits of affected individuals are depicted in different colours. AF cases are depicted in black, ASD cases are depicted in blue, VSD cases are depicted in red and unaffected individuals are depicted in white. Abbreviations, ASD, atrial septal defect; AF, atrial fibrillation; VSD, ventricular septal defect.
Chapter 5. Identification of a Novel Mutation for AF

5.2.2 Mutation detection

5.2.2.1 Candidate gene screening

Transcription factors have been demonstrated to play critical roles during cardiac development. One of the developmental processes regulated by transcription factors is atrial and ventricular septation. Mutations in the transcription factor genes TBX5, TBX20, NKK-2.5 and GATA4 are commonly associated with atrial and ventricular septal defects. TBX5 mutations have been identified in patients with Holt-Oram syndrome, a condition characterised by atrial and ventricular septal defects associated with conduction system disease and AF. Mutations in the related TBX20 gene have been reported in patients with complex congenital cardiac defects including atroventricular septal defects, valvular heart defects and cardiomyopathy. Mutations in GATA4 have been reported to underlie isolated cardiac septal defects. Finally, mutations in the NKK-2.5 gene have been identified in patients with atrial and ventricular septal defects with associated conduction block.

Based on the reports linking transcription factor mutations to atroventricular septal defects, a candidate-gene approach was initially adopted in an attempt to identify the causative mutation in pedigree AF-435. Sanger sequencing of TBX5, TBX20, GATA4 and NKK-2.5 was performed in individual II-13. A single variant in NKK-2.5 and eight variants in TBX20 were identified. The NKK-2.5 variant and seven of the eight TBX20 variants were synonymous. The single nonsynonymous variant in TBX20 has
Chapter 5. Identification of a Novel Mutation for AF

previously been reported in publically available databases (dbSNP, 1000 Genomes Project, exome variant server). Overall therefore, it is very unlikely that these are pathogenic variants. The results from the candidate-gene screening approach are summarised in Table 15.

Table 15: Variants identified in \textit{NKX-2.5} and \textit{TBX20} in family AF-435

<table>
<thead>
<tr>
<th>Gene</th>
<th>Nucleotide substitution</th>
<th>Amino acid substitution</th>
<th>Exon</th>
<th>rs ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>NKX-2.5</td>
<td>63A&gt;G</td>
<td>E21E</td>
<td>1</td>
<td>rs2277923</td>
</tr>
<tr>
<td>TBX20</td>
<td>933G&gt;C</td>
<td>R311P</td>
<td>7</td>
<td>rs113201899</td>
</tr>
<tr>
<td>TBX20</td>
<td>934C&gt;T</td>
<td>R311R</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>TBX20</td>
<td>926T&gt;A</td>
<td>Y309N</td>
<td>7</td>
<td>rs111862418</td>
</tr>
<tr>
<td>TBX20</td>
<td>952C&gt;T</td>
<td>Y317Y</td>
<td>7</td>
<td>rs113335362</td>
</tr>
<tr>
<td>TBX20</td>
<td>1165A&gt;G</td>
<td>P388P</td>
<td>8</td>
<td>rs2723759</td>
</tr>
<tr>
<td>TBX20</td>
<td>1190C&gt;T</td>
<td>L397L</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>TBX20</td>
<td>1195A&gt;C</td>
<td>T398T</td>
<td>8</td>
<td>rs2532122</td>
</tr>
<tr>
<td>TBX20</td>
<td>1332 C&gt;T</td>
<td>T444M</td>
<td>8</td>
<td>rs201217462</td>
</tr>
</tbody>
</table>

5.2.2.2 Exome sequencing

Based on the results of candidate gene screening, we hypothesised that the phenotype in family AF-435 is caused by a mutation in a novel gene for congenital atrial and ventricular septal defects and AF. In order to identify the causative mutation, we performed exome sequencing.
Chapter 5. Identification of a Novel Mutation for AF

Exome sequencing was performed in three affected individuals from family AF-435 (III-5, III-15 and III-16) using the Illumina HiSeq200 platform. Approximately 5.5 Gb of sequencing data was generated per individual as paired-end, 100 base pair reads. Reads with duplicate start sites were discarded. Approximately 60 million reads per individual passed quality control and of these, 96% were aligned to the reference genome (Table 16). A mean depth of coverage of 60x was achieved across the samples (Table 17). A minimum threshold of 20x coverage was set for variant calling. 96% of bases fulfilled this criterion for variant calling. An average of 18,825 variants was identified in the three affected individuals who underwent exome sequencing. The results from variant calling are summarised in Table 18.

Table 16: Number of paired-end reads aligned to the reference genome

<table>
<thead>
<tr>
<th></th>
<th>Total read pairs</th>
<th>Unique read pairs</th>
<th>% mapped reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>III5</td>
<td>59,952,850</td>
<td>57,499,424</td>
<td>95.9</td>
</tr>
<tr>
<td>III15</td>
<td>60,122,024</td>
<td>57,662,318</td>
<td>95.9</td>
</tr>
<tr>
<td>III16</td>
<td>58,645,600</td>
<td>56,114,632</td>
<td>95.6</td>
</tr>
</tbody>
</table>

Table 17: Exome sequencing coverage data

<table>
<thead>
<tr>
<th>Sample</th>
<th>Mean coverage</th>
<th>% 15x coverage</th>
<th>% 30x coverage</th>
<th>% 40x coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>III5</td>
<td>63</td>
<td>84</td>
<td>71</td>
<td>61</td>
</tr>
<tr>
<td>III15</td>
<td>63</td>
<td>83</td>
<td>70</td>
<td>60</td>
</tr>
<tr>
<td>III16</td>
<td>56</td>
<td>83</td>
<td>74</td>
<td>63</td>
</tr>
</tbody>
</table>
Chapter 5. Identification of a Novel Mutation for AF

Table 18: Number of candidate variants indentified by exome sequencing in AF-435

<table>
<thead>
<tr>
<th></th>
<th>III5</th>
<th>III15</th>
<th>III16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total no. of variants</td>
<td>18,529</td>
<td>18,019</td>
<td>19,928</td>
</tr>
<tr>
<td>No. of novel variants</td>
<td>803</td>
<td>771</td>
<td>592</td>
</tr>
<tr>
<td>No. of novel missense variants</td>
<td>548</td>
<td>521</td>
<td>385</td>
</tr>
<tr>
<td>No. of novel nonsense variants</td>
<td>15</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>No. of splice site variants</td>
<td>0</td>
<td>1</td>
<td>7</td>
</tr>
</tbody>
</table>

A number of different criteria were used to differentiate potential disease causing variants from other variants. Firstly, only novel nonsynonymous variants were included in downstream analysis. The majority of the variants were filtered out as they were synonymous and/or have previously been reported in population databases (dbSNP, 1000 Genomes Project, exome variant server) and are therefore highly unlikely to be causative. The number of novel missense variants varied between 385 and 548, nonsense variants varied between 7 and 17, and splice site variants varied between 0 and 7 (Table 18). Secondly, based on the fact that all three affected individuals are predicted to carry the causative mutation, all variants that did not segregate with disease were excluded from downstream analysis. Using this strategy, the number of candidate variants was reduced to 18 missense variants. The variants are summarised in Table 19.

The majority of variants underlying rare monogenic diseases affect sequences that are highly conserved through evolution. We therefore prioritised the 18 candidate missense variants based on evolutionary conservation. Interspecies conservation
across 44 vertebrates at every base position was calculated using the phylop tool.

Four variants were located at bases that are highly conserved through evolution (as evidenced by strongly positive phylop scores) Conservation scores are summarised in Table 19.

**Table 19:** List of potentially causative genetic mutations in family AF-435

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>Function</th>
<th>Conservation score</th>
</tr>
</thead>
<tbody>
<tr>
<td>PABPC1</td>
<td>Poly(A) binding protein, cytoplasmic 1</td>
<td>R374C</td>
<td>5.95</td>
</tr>
<tr>
<td><strong>GATA6</strong></td>
<td>GATA binding protein 6</td>
<td>R585L</td>
<td>4.94</td>
</tr>
<tr>
<td>SLC35F3</td>
<td>Solute carrier family 35, member F3</td>
<td>H297P</td>
<td>3.624</td>
</tr>
<tr>
<td>MUC4</td>
<td>Mucin 4, cell surface associated</td>
<td>P1168L</td>
<td>2.798</td>
</tr>
<tr>
<td>STOX2</td>
<td>Storkhead box 2</td>
<td>R480S</td>
<td>0.882</td>
</tr>
<tr>
<td>HHATL</td>
<td>Hedgehog acyltransferase-like</td>
<td>K484N</td>
<td>0.555</td>
</tr>
<tr>
<td>MUC4</td>
<td>Mucin 4, cell surface associated</td>
<td>D3317E</td>
<td>0.428</td>
</tr>
<tr>
<td>MUC4</td>
<td>Mucin 4, cell surface associated</td>
<td>G1148A</td>
<td>0.162</td>
</tr>
<tr>
<td>LOC440563</td>
<td>Heterogeneous nuclear ribonucleoprotein C-like</td>
<td>E245D</td>
<td>-0.185</td>
</tr>
<tr>
<td>MUC4</td>
<td>Mucin 4, cell surface associated</td>
<td>V1177A</td>
<td>-0.259</td>
</tr>
<tr>
<td>C3orf77</td>
<td>Chromosome 3 open reading frame 77</td>
<td>N430T</td>
<td>-0.272</td>
</tr>
<tr>
<td>MUC4</td>
<td>Mucin 4, cell surface associated</td>
<td>D2309G</td>
<td>-0.279</td>
</tr>
<tr>
<td>MUC4</td>
<td>Mucin 4, cell surface associated</td>
<td>D2309N</td>
<td>-0.416</td>
</tr>
<tr>
<td>ZNF208</td>
<td>Zinc finger protein 208</td>
<td>V685I</td>
<td>-0.485</td>
</tr>
<tr>
<td>PRUNE2</td>
<td>Prune homolog 2 (Drosophila) who</td>
<td>G1946D</td>
<td>-0.628</td>
</tr>
<tr>
<td>LOC440563</td>
<td>Heterogeneous nuclear ribonucleoprotein C-like</td>
<td>A248V</td>
<td>-0.673</td>
</tr>
<tr>
<td>MUC4</td>
<td>Mucin 4, cell surface associated</td>
<td>A4062T</td>
<td>-0.783</td>
</tr>
<tr>
<td>ACAN</td>
<td>Aggrecan</td>
<td>D1390E</td>
<td>-1.164</td>
</tr>
</tbody>
</table>
Further prioritisation of candidate variants was performed on the basis of cardiac expression and a potential functional effect of the gene in cardiac development and function. Using this strategy, we identified a single compelling candidate mutation (R585L) in the GATA6 gene. GATA6 is a member of a family of zinc finger transcription factors that are highly expressed in the heart.\textsuperscript{319} Mutations in GATA6 have been reported to cause congenital heart defects such as Tetralogy of Fallot and ASD.\textsuperscript{320}

The R585L mutation is located at the C terminal of the GATA6 peptide (Figure 52). As mentioned above, the affected residue is highly conserved throughout evolution (Figure 53). Based on these observations, the GATA6 R585L mutation was prioritised for segregation and functional analysis.

\textbf{Figure 52: Schematic representation of GATA4 and GATA6 proteins.} The R585L mutation is located at the C terminus of the GATA6 protein (red cross).
Chapter 5. Identification of a Novel Mutation for AF

5.2.3 Segregation analysis of the GATA6 R585L mutation

In order to confirm the presence of the GATA6 R585L mutation in the three individuals who underwent exome sequencing (III-5, III-15 and III-16) and to determine whether other members of pedigree AF-435 are mutation carriers, we performed direct Sanger sequencing. 13 members of the family were screened for the GATA6 R585L mutation. Four additional individuals were found to be heterozygous carriers of the mutation (II-3, II-5, II-17, III-4). Example electrophoretograms from Sanger sequencing are illustrated in Figure 54.

Despite harbouring the GATA6 R585L mutation, individuals II-5, II-17 and III-4 did not display structural cardiac abnormalities or significant arrhythmias. However, as discussed in Section 5.2.1, these individuals had an abnormal ECG with altered P wave axis and incomplete RBBB. Of note, the individuals who did not harbour the GATA6

AF-435  EVTSSVLPDSWCA
Human   EVTSSVRPDSWCA
Mouse   EVTSSVRQDSWCA
Cow     EVTSSVRQDSWCA
Dog     EVTSSVRQDSWCA
Chick   EVTASVRQDHWCA
Lizard  EVTASVRQDSWCA

Figure 53: Multi-species alignment of GATA6 amino acid sequence. Arginine 585 (highlighted in the red box) is highly conserved across many species.
R585L mutation had a normal baseline ECG. It is plausible therefore, that the ECG changes represent a more subtle form of the phenotype. The clinical characteristics of the mutation carriers are summarised in Table 20. Based on the findings from segregation analysis, functional analysis of the mutation was undertaken.

Figure 54: Electrophoretogram of GATA6 R585L mutation. Direct Sanger sequencing of GATA6 revealed a heterozygous mutation (R585L). The upper electrophoretogram is from a control subject while the lower electrophoretogram is from a family member with the mutation (III-5).
Table 20: Clinical characteristics of family AF-435

<table>
<thead>
<tr>
<th>ID</th>
<th>Sex</th>
<th>Age</th>
<th>Phenotype</th>
<th>PR</th>
<th>QRS</th>
<th>QT</th>
<th>ECG</th>
<th>LVEF (%)</th>
<th>LA (mm)</th>
<th>LVID (mm)</th>
<th>RV base</th>
<th>RS85L mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>II-1</td>
<td>F</td>
<td>66</td>
<td>-</td>
<td>122</td>
<td>68</td>
<td>436</td>
<td>RBBB, abnormal P axis</td>
<td>66</td>
<td>31</td>
<td>36</td>
<td>38</td>
<td>N</td>
</tr>
<tr>
<td>II-3</td>
<td>F</td>
<td>65</td>
<td>PAF</td>
<td>162</td>
<td>72</td>
<td>421</td>
<td>RBBB, abnormal P axis</td>
<td>72</td>
<td>37</td>
<td>42</td>
<td>29</td>
<td>Y</td>
</tr>
<tr>
<td>II-5</td>
<td>F</td>
<td>64</td>
<td>-</td>
<td>130</td>
<td>84</td>
<td>426</td>
<td>RBBB, abnormal P axis</td>
<td>77</td>
<td>33</td>
<td>40</td>
<td>27</td>
<td>Y</td>
</tr>
<tr>
<td>II-7</td>
<td>M</td>
<td>62</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>N</td>
</tr>
<tr>
<td>II-9</td>
<td>F</td>
<td>61</td>
<td>-</td>
<td>148</td>
<td>86</td>
<td>447</td>
<td>RBBB, abnormal P axis</td>
<td>66</td>
<td>34</td>
<td>45</td>
<td>32</td>
<td>N</td>
</tr>
<tr>
<td>II-11</td>
<td>M</td>
<td>59</td>
<td>-</td>
<td>154</td>
<td>88</td>
<td>422</td>
<td>-</td>
<td>78</td>
<td>37</td>
<td>44</td>
<td>28</td>
<td>N</td>
</tr>
<tr>
<td>II-13</td>
<td>F</td>
<td>57</td>
<td>ASD</td>
<td>150</td>
<td>80</td>
<td>465</td>
<td>-</td>
<td>55</td>
<td>53</td>
<td>43</td>
<td>-</td>
<td>Y</td>
</tr>
<tr>
<td>II-15</td>
<td>M</td>
<td>55</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>N</td>
</tr>
<tr>
<td>II-17</td>
<td>M</td>
<td>51</td>
<td>-</td>
<td>168</td>
<td>98</td>
<td>429</td>
<td>RBBB, abnormal P axis</td>
<td>69</td>
<td>36</td>
<td>50</td>
<td>43</td>
<td>Y</td>
</tr>
<tr>
<td>III-4</td>
<td>M</td>
<td>32</td>
<td>-</td>
<td>122</td>
<td>92</td>
<td>416</td>
<td>RBBB, abnormal P axis</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Y</td>
</tr>
<tr>
<td>III-5</td>
<td>M</td>
<td>30</td>
<td>VSD</td>
<td>144</td>
<td>116</td>
<td>400</td>
<td>-</td>
<td>60</td>
<td>37</td>
<td>48</td>
<td>27</td>
<td>Y</td>
</tr>
<tr>
<td>III-9</td>
<td>F</td>
<td>26</td>
<td>-</td>
<td>130</td>
<td>82</td>
<td>444</td>
<td>-</td>
<td>73</td>
<td>31</td>
<td>46</td>
<td>33</td>
<td>N</td>
</tr>
<tr>
<td>III-13</td>
<td>M</td>
<td>27</td>
<td>-</td>
<td>144</td>
<td>104</td>
<td>434</td>
<td>-</td>
<td>64</td>
<td>37</td>
<td>50</td>
<td>37</td>
<td>N</td>
</tr>
<tr>
<td>III-14</td>
<td>F</td>
<td>24</td>
<td>-</td>
<td>154</td>
<td>76</td>
<td>394</td>
<td>-</td>
<td>68</td>
<td>30</td>
<td>43</td>
<td>31</td>
<td>N</td>
</tr>
<tr>
<td>III-15</td>
<td>M</td>
<td>24</td>
<td>PAF</td>
<td>152</td>
<td>104</td>
<td>405</td>
<td>-</td>
<td>55</td>
<td>40</td>
<td>54</td>
<td>-</td>
<td>Y</td>
</tr>
<tr>
<td>III-16</td>
<td>M</td>
<td>21</td>
<td>PAF</td>
<td>148</td>
<td>94</td>
<td>401</td>
<td>-</td>
<td>49</td>
<td>33</td>
<td>53</td>
<td>-</td>
<td>Y</td>
</tr>
</tbody>
</table>

PAF, paroxysmal atrial fibrillation; HR, heart rate; LVEF, left ventricular ejection fraction; LA, left atrium; RBBB, right bundle branch block; P axis, P wave axis.
4.1.4 Functional analysis of the GATA6 R585L mutation

The GATA6 transcription factor is a member of a family of zinc finger transcription factors. The GATA transcription factors regulate the function of a variety of genes by binding to a specific consensus DNA sequence (A/T)GATA(A/G) in the promoter or enhancer region of the genes. Examples of genes regulated by GATA transcription factors include the cardiac-specific genes NPPA, BNP and αMHC. Therefore, an effective strategy to measure functional activity of GATA transcription factors is to measure their effect on the promoters of these target genes.

Luciferase reporter assays represent a sensitive technique for measuring promoter activity. The reporter constructs for luciferase assays consist of regulatory elements (promoters or enhancers) fused to the luciferase gene. The luciferase gene encodes the luciferase enzyme which oxidises a substrate (luciferin) to emit light. The expression of the luciferase gene is under the direct control of the fused regulatory element. Therefore, the amount of luminescence is a reflection of the activity of the regulatory element. In turn, the activity of the regulatory element is correlated with the function of the transcription factor of interest.

A Firefly and Renilla luciferase reporter assay was used to measure GATA6 activity in the present study. Three separate luciferase reporter constructs were used to compare transcriptional activity of the wild type and mutant GATA6. The NPPA promoter-luc construct consists of a 500 bp fragment of the rat ANF promoter.
subcloned into a luciferase reporter construct.\textsuperscript{259} The $\alpha$MHC promoter-luc consists of 2.8 kb of the $\alpha$MHC promoter in a luciferase reporter construct.\textsuperscript{260} The BNP promoter-luc consists of 2.5 kb of the rat BNP promoter in a luciferase reporter construct.\textsuperscript{261} The reporter constructs were co-transfected with expression plasmids encoding wild type and mutant GATA6 into a human embryonic kidney cell line (HEK293). Transcriptional activity was measured as relative luciferase units.

The mutant GATA6 (mGATA6) expression plasmid resulted in a significant and consistent increase in transcription activity compared to the wild-type (NPPA promoter-luc, 1.7 fold, p=<0.05); $\alpha$MHC promoter-luc, 1.7 fold, p=<0.05; BNP promoter-luc, 1.9 fold, p=<0.05). These results indicate that the GATA6 mutation is associated with gain-of-function type modulation of transcription factor function (Figure 55).

![Figure 55: Luciferase reporter assays of mutant GATA6.](image_url)

**Figure 55: Luciferase reporter assays of mutant GATA6.** Luciferase assays were performed using NPPA promoter-luc, $\alpha$MHC promoter-luc and BNP-promoter-luc. The mutant GATA6 (mGATA6) has enhanced transcriptional activity compared with the wild type control. These results indicate that R585L is a gain-of-function mutation.
5.3 Discussion

5.3.1 Summary of main findings

In a large pedigree with a complex phenotype of AF, atrial septal defects and ventricular septal defects (AF-435), exome sequencing of three affected family members identified a compelling candidate mutation in the GATA6 gene (R585L). GATA6 is a member of a family of zinc finger transcription factors that are highly expressed in the heart. The GATA transcription factors are known to play critical roles during cardiac development and morphogenesis. Sanger sequencing demonstrated that the R585L mutation segregates with disease in AF-435. The R585L mutation is located at the C-terminus of the GATA6 peptide in a residue that is highly conserved throughout evolution. Functional analysis using luciferase reporter assays demonstrated that the R585L GATA6 mutation is associated with gain-of-function type modulation. Taken together, these findings suggest that R585L GATA6 is the causative mutation in pedigree AF-435.

The following section contains a background on the role of transcription factors in cardiac development followed by a discussion on the potential mechanistic link between the R585L GATA6 mutation and congenital heart defects.
5.3.2 Cardiac transcription factors

Transcription factors are proteins that bind to specific DNA sequences in the promoter regions of genes and regulate gene expression. The human genome encodes more than 500 transcription factors. These transcription factors may be ubiquitous or tissue-specific. Based on their structure and DNA binding domains, transcription factors are classified into different families.

A range of cardiac-specific transcription factors are involved in regulation of cardiac development and morphogenesis. Core cardiac transcription factors include the GATA family of zinc finger proteins (GATA4, GATA5, and GATA6), the homeodomain protein NKX-2.5, and T-box factors (TBX1, TBX2, TBX3, TBX5, TBX18, and TBX20). NKX2-5, TBX5 and GATA4 are the most extensively characterised subtypes from these transcription factor families. The following discussion will focus specifically on the GATA family of transcription factors.

5.3.2.1 The GATA family of transcription factors

GATA transcription factors preferentially bind to the consensus DNA binding sequence (A/T)GATA(A/G) through two zinc fingers (CysX2-CysX17-CysX2-Cys). The GATA family consists of six transcription factors. Based on their tissue distribution, GATA transcription factors are classified into two subgroups; GATA1, GATA2 and GATA3 are expressed in hematopoietic cells while GATA4, GATA5 and GATA6 are expressed in the heart. The GATA4 and GATA6 subtypes are expressed in the developing and postnatal
myocardium. In contrast, GATA5 expression is restricted to the developing endocardium.\textsuperscript{325}

GATA4, GATA5 and GATA6 constitute proteins of 48, 42 and 45 kDa respectively. In addition to the zinc finger domains, GATA4 contains N and C terminal transactivation domains, a nuclear localisation signal and multiple phosphorylation sites. The other two GATA subtypes share significant structural homology with GATA4 and are therefore predicted to harbour similar transactivation and localisation domains (Figure 56). Recent studies have demonstrated that GATA6 has two isoforms. The second isoform has an N-terminal extension and is expressed in multiple tissues including the heart.\textsuperscript{324}

GATA4 is the most extensively characterised of the cardiac GATA transcription factors. Current evidence relating to the GATA6 subtype is relatively limited. The following discussion focuses on the GATA4 and GATA6 subtypes.

GATA4 plays a critical role in a range of developmental processes including survival, migration and differentiation of cardiac precursor cells.\textsuperscript{326} In a murine model, knockout of GATA4 results in early embryonic lethality due to failure of primitive heart tube development.\textsuperscript{327} In addition to its role in development, a number of studies have indicated that GATA4 plays an important role in the postnatal myocardium. For instance, GATA4 has been implicated as a mediator of cardiac hypertrophy in the adult
heart.\textsuperscript{328} More recently, GATA6 has also been demonstrated to be critical for cardiac development. In keeping with the findings relating to GATA4, knockout of GATA6 in a mouse model results in early embryonic lethality due to a failure of cardiac development.\textsuperscript{329}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{gata_structural_organisation.png}
\end{figure}

The GATA transcription factors are key regulators of the expression of a range of cardiac genes. GATA4 is a potent transactivator of sarcomeric proteins, transcription factors and enzymes (Table 21).\textsuperscript{321} The target genes for GATA6 are just beginning to be uncovered. Consistent with the findings relating to GATA4, GATA6 has been
demonstrated to activate the promoter regions of *ANP*, *BNP*, *α-MHC* and *cardiac troponin C*. Of note, the (A/T)GATA(A/G) binding sequence is found in the promoter elements of most of the genes expressed in the heart. Therefore the aforementioned genes are likely to represent a fraction of the targets for the GATA transcription factors.

GATA4 and GATA6 are coexpressed in cardiac myocytes. Previous studies have demonstrated that these transcription factors interact physically and functionally to regulate the expression of specific subsets of genes. Physical interaction between GATA4 and GATA6 occurs via the zinc finger binding domain. Functional interaction requires both the zinc finger domain and the C-terminal activation domain of GATA4. It is important to note that while GATA4 and GATA6 are predicted to act in concert, these transcription factors individually fulfil essential and nonredundant roles during cardiac development.

GATA transcription factors also interact with a range of other transcription factors and DNA binding proteins. GATA4 interacts with NKK-2.5, TBX5, NFATc4, MEF-2, dHAND, FOG-2, and the transcriptional co-activator protein p300. The interaction profile of GATA6 is less well characterised. GATA6 interacts with the transcriptional co-activator p300, NFATc1 and the peroxisome proliferator activator (PPAR) binding protein.
Table 21. Cardiac proteins regulated by GATA transcription factors

<table>
<thead>
<tr>
<th>Category</th>
<th>Proteins/Proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiac troponin C</td>
<td></td>
</tr>
<tr>
<td>Cardiac troponin I</td>
<td></td>
</tr>
<tr>
<td>Myosin heavy chain</td>
<td></td>
</tr>
<tr>
<td>Cardiac actin</td>
<td></td>
</tr>
<tr>
<td>Myosin light chain-3</td>
<td></td>
</tr>
<tr>
<td>β Myosin heavy chain</td>
<td></td>
</tr>
<tr>
<td>Slow myosin heavy chain</td>
<td></td>
</tr>
<tr>
<td>Myosin light chain-2</td>
<td></td>
</tr>
<tr>
<td>NNX-2.5</td>
<td></td>
</tr>
<tr>
<td>dHAND</td>
<td></td>
</tr>
<tr>
<td>Cardiac-restricted ankyrin repeat protein (CARP)</td>
<td></td>
</tr>
<tr>
<td>MEF2C</td>
<td></td>
</tr>
<tr>
<td><strong>Signalling peptides and receptors</strong></td>
<td></td>
</tr>
<tr>
<td>Atrial natriuretic peptide (ANP)</td>
<td></td>
</tr>
<tr>
<td>Brain natriuretic peptide (BNP)</td>
<td></td>
</tr>
<tr>
<td>Wnt2</td>
<td></td>
</tr>
<tr>
<td>Cardiac m2 muscarinic acetylcholine receptor</td>
<td></td>
</tr>
<tr>
<td>A1 adenosine receptor</td>
<td></td>
</tr>
<tr>
<td>Platelet-derived growth factor receptor β (PDGFRβ)</td>
<td></td>
</tr>
<tr>
<td>Angiotensin II type Ia receptor</td>
<td></td>
</tr>
<tr>
<td>BMP-4</td>
<td></td>
</tr>
<tr>
<td>Endothelin 1</td>
<td></td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
</tr>
<tr>
<td>Na^+–Ca^+ exchanger gene</td>
<td></td>
</tr>
<tr>
<td>Corin</td>
<td></td>
</tr>
<tr>
<td>Carnitine palmitoyltransferase 1β</td>
<td></td>
</tr>
<tr>
<td>Adenylosuccinate synthetase I</td>
<td></td>
</tr>
<tr>
<td>N-Cadherin</td>
<td></td>
</tr>
<tr>
<td>α T-Catenin</td>
<td></td>
</tr>
<tr>
<td>Bcl-X</td>
<td></td>
</tr>
<tr>
<td>Bcl-2</td>
<td></td>
</tr>
</tbody>
</table>

5.3.3 Cardiac development and congenital heart disease

Congenital heart diseases are defined as abnormalities of the heart that occur before birth. Congenital heart diseases can be broadly classified into two groups; structural cardiac malformations and functional cardiac abnormalities, which include cardiac arrhythmias. A major type of structural heart defects are septation defects, which include atrial and ventricular septal defects.\textsuperscript{238}

In order to understand congenital heart diseases, it is important to gain an appreciation of normal cardiac development. During mammalian cardiac development, cardiac progenitor cells within the so-called first heart field (anterior lateral plate mesoderm) migrate to the midline and fuse to form a primitive heart tube. The linear heart tube elongates by recruiting precursor cells from the second heart field (the pharyngeal mesoderm) and subsequently undergoes looping. Endocardial cushions form and septate the heart and the outflow tracts. Finally, the heart undergoes extensive remodelling, both structural and electrical.\textsuperscript{323} The stages of cardiac development are summarised in Figure 57.

The formation and compartmentalisation of the heart is reliant on the orchestrated deployment and differentiation of cardiac progenitor pools. In turn, large numbers of intracellular transcriptional factors are necessary for the regulation of these processes. Mutations in transcription factor genes can result in a disruption of these processes resulting in structural and/or functional cardiac defects.
Figure 57: Normal cardiac development. a) Depicts early stages of cardiac development. At the initial stages of development, two pools of precursor cells form a cardiac crescent. The first heart field (FHF) eventually forms the left ventricle (LV) while the second heart field eventually forms the right ventricle (RV), the left and right atria (LA and RA), the sinus venosus (SV) and the outflow tract (OT). b) Depicts cardiac maturation stages. The atrioventricular valves arise from the cardiac cushions (CC). The ventricular septum (VS) arises from the right and left ventricular myocardium. The atrial septum (AS) forms from the growth of two septa; the primary septum (depicted in green) and the secondary septum (depicted in pink). The common outflow tract is separated into the aorta (AO) and pulmonary artery (PA) through outflow tract septation. From: Bruneau BG. Nature. 2008;451(7181):943-8. 238
5.3.3.1 Transcription factor mutations and congenital heart disease

The first transcription factor gene to be implicated in congenital heart disease was TBX5. Mutations in TBX5 cause Holt-Oram syndrome, a condition characterised by ASD, VSD, cardiac conduction abnormalities and arrhythmias. Since the discovery of TBX5 mutations in Holt-Oram syndrome, a range of transcription factor mutations have been reported in congenital heart diseases (summarised in Figure 58). Important amongst these are GATA gene mutations.

![Figure 58: Origin and genetic mutations underlying congenital heart diseases.](image)

Genes which cause isolated congenital heart defects are indicated in black while genes that form part of wider syndromes are indicated in blue. From: Bruneau BG. Nature. 2008;451(7181):943-8.
Multiple \textit{GATA4} mutations have been reported to underlie congenital heart disease. In 2003, Garg \textit{et al} \cite{317} identified two \textit{GATA4} mutations (G296S and E359del) in pedigrees with isolated cardiac septal defects. Rajagopal \textit{et al} \cite{336} subsequently screened a cohort of patients with a number of congenital heart defects and identified four \textit{GATA4} mutations. The first (G296C) was identified in a pedigree with ASD and pulmonary stenosis, the second (L403M) in a patient with an ASD and a hypoplastic right ventricle, and the third and fourth (P163S and A346V) in patients with endocardial cushion defects. Tomita-Mitchell \textit{et al} \cite{337} performed candidate-gene screening in a relatively large cohort of congenital heart disease patients and also identified four \textit{GATA4} mutations (G93A, Q316E, A411V and D425N). Two mutation carriers had ASD, two had VSD, and one had Tetralogy of Fallot. All of the reported mutations were associated with loss-of-function effects.

More recently, \textit{GATA6} mutations have been reported in congenital heart disease. In a study by Kodo \textit{et al} \cite{338} two \textit{GATA6} mutations (E486del and N466H) were identified after screening a cohort of 21 patients with persistent truncus arteriosus. Huang \textit{et al} \cite{339} reported two \textit{GATA6} mutations (G367X and G394C) after screening 52 unrelated patients who underwent surgical repair of Tetralogy of Fallot. Finally, Lin \textit{et al} \cite{320} identified the same heterozygous missense mutation in three patients (S184N) after screening 270 individuals with congenital heart disease. Two patients had ASDs while the third had Tetralogy of Fallot. All the mutations were associated with loss-of-function type modulation.
5.3.3.2 Transcription factor variants and AF

In addition to their role in structural congenital heart diseases, transcription factors are emerging as potential contributors to the pathogenesis of AF. Both TBX5 and GATA mutations have been implicated in monogenic forms of AF. Postma et al.\textsuperscript{340} previously reported on a pedigree with a TBX5 mutation causing an atypical form of Holt-Oram syndrome with musculoskeletal defects and a premature onset of AF. Of note, the majority of AF patients did not have structural cardiac abnormalities.

Multiple candidate-gene screening studies have identified isolated GATA4 mutations in AF patients.\textsuperscript{341-343} The overall prevalence of such mutations is low (approximately 0.1\%). GATA6 mutations have also been reported to underlie AF. Yang et al.\textsuperscript{344} identified a heterozygous GATA6 mutation (Y235S) after screening a cohort of 138 AF patients. In a separate study, 140 patients with AF were screened and a second GATA6 heterozygous mutation (G469V) was reported. Both GATA6 mutations resulted in a reduction of transcriptional factor activity.\textsuperscript{345}

In addition to the reports in monogenic forms of AF, transcription factors have also been implicated in population-based studies of AF. GWAS have reported susceptibility loci for AF at two cardiac transcription factor genes, \textit{PITX2} and \textit{PRRX1}. \textit{PITX2} encodes a paired homeodomain transcription factor that plays a major role in establishing atrial identity, and is also a major determinant of normal connection of the pulmonary veins.\textsuperscript{346} \textit{PRRX1} encodes a homeodomain transcription factor which is involved in the
Chapter 5. Identification of a Novel Mutation for AF

development of the pulmonary vasculature.\textsuperscript{347} As discussed previously, the most frequent site of origin of the focal triggers that provoke and potentially drive AF is the pulmonary veins.\textsuperscript{52} Therefore, altered function of these transcription factors could potentially increase susceptibility to AF through defective morphogenesis of the pulmonary vein myocardium.

5.3.4 R585L expands the spectrum of GATA6 mutations causing AF and structural cardiac defects

The identification of the R585L GATA6 mutation in the present study further expands on the potential role of the GATA6 transcription factor in the pathogenesis of AF and structural congenital heart diseases. As discussed above, GATA6 mutations have previously been identified in patients with either structural heart defects or lone AF. Interestingly, the R585L GATA6 mutation appears to cause discordant clinical phenotypes in members of the same pedigree; one mutation carrier has an ASD, another has a VSD, two individuals have AF, while a further four have more subtle ECG phenotypes. These findings may be a reflection of the fact that genotype-phenotype correlations in patients with transcription factor mutations are not straightforward.

In keeping with the findings in the present study, transcription factor mutations have been reported to produce considerable intrafamilial variation in cardiac malformations.\textsuperscript{348} A typical example is Holt-Oram syndrome, which is characterised by
atrial and ventricular septal defects, Tetralogy of Fallot, left heart hypoplasia, AF, and cardiac conduction abnormalities.\textsuperscript{349}

The molecular mechanisms underlying the variable clinical phenotypes in patients with transcription factor mutations remain largely unknown. Potential explanations include the presence of modifier genetic variants and/or environmental factors. Modifier genetic variants could influence the expressivity of transcription factor mutations in a number of different ways. Firstly, these variants could potentially be located in the promoter sequences of target genes and could therefore alter the effect of the mutant transcription factor. Secondly, variants within other transcription factors or cofactor proteins which interact with the mutant transcription factor could influence the effect of the primary mutation.\textsuperscript{348} Environmental factors are predicted to play less of a role in the expression of structural congenital heart defects. However, they may exert a significant influence on the development of arrhythmias, which typically develop after birth. A useful analogy to explain the observed variation in expression is that the transcription factor mutation serves as a ‘gross tune’ while other factors such as modifier genetic variants serve as ‘finer tuning’ of the phenotype.\textsuperscript{350}

The presence of mild ECG phenotypes in R585L GATA6 mutation carriers is potentially a reflection of incomplete penetrance of the mutation. The proposed mechanisms underlying incomplete penetrance of monogenic mutations are the same as those for
variation in clinical phenotypes discussed above i.e. additional variants that modify the effect of the primary mutation and environmental factors.

5.3.5 Gain-of-function mutation of GATA6 causes AF and structural cardiac defects

The vast majority of transcription factor mutations identified in patients with congenital heart diseases and AF have been associated with loss-of-function type modulation. In the specific case of the GATA transcription factors, all mutations reported to date are loss-of-function mutations. Further, the majority of mutations are located in the zinc finger domains which are responsible for DNA binding.

The R585L GATA6 mutation displays a number of distinguishing properties when compared to previously reported transcription factor mutations. R585L is located in a highly conserved residue in the C-terminal of the peptide. Relatively little is known about the function of the C-terminal protein sequences of GATA transcription factors. Deletion analysis of the GATA4 protein with C-terminal truncations of 20 and 35 amino acids has not been reported to result in a significant alteration of transcription factor activity. On the other hand, deletion of 50 amino acids results in an 80% reduction of transcriptional activity. The C-terminus of GATA6 shares significant homology with GATA4. Based on the findings relating to GATA4, the R585L GATA6 mutation would not be predicted to significantly alter transcriptional activity as the amino acid substitution is located within the terminal 20 amino acids at the C-
Chapter 5. Identification of a Novel Mutation for AF

terminus. Interestingly however, this mutation results in a gain-of-function effect. By implication therefore, the C-terminal portion of GATA6 may harbour important functional domains which are not present in the GATA4 peptide.

Despite the fact that the R585L GATA6 mutation is associated with a gain-of-function effect, it causes very similar phenotypic abnormalities to previously reported loss-of-function mutations. While these results are apparently conflicting, they compare favourably to previous reports relating to the TBX5 transcription factor. While the majority of TBX5 mutations underlying Holt-Oram Syndrome result in a loss-of-function effect, chromosome 12q2 duplications, which increase TBX5 dosage, and gain-of-function TBX5 mutations have also been reported to underlie this condition.\(^3\)\(^4\) Further, in chicken and Xenopus models, overexpression of Tbx5 causes a thinned and hypoproliferative myocardium which is reminiscent of the phenotype observed in Tbx5 knockout models.\(^3\)\(^5\)\(^2\), \(^3\)\(^5\)\(^3\) On the basis of these observations, it could be speculated that a normal cardiac phenotype is dependent upon the correct level of transcription factor function or correct gene dosage.\(^3\)\(^5\)\(^4\) Therefore, gain-of-function and loss-of-function GATA6 mutations could underlie the same phenotypic defects.

GATA6 plays an important role in regulating cardiac differentiation. Interestingly, previous experiments in Xenopus embryos demonstrated that artificially elevating levels of GATA6 results in a delay in terminal differentiation of cardiomyocytes. However, when GATA6 levels dropped due to mRNA degradation, cardiomyocyte
maturation proceeded and the differentiation programme was completed. These findings suggest that GATA6 may be responsible for the maintenance of cardiac precursor cells in an undifferentiated and proliferative state. These observations lend further support to our finding that gain-of-function of GATA6 can cause significant developmental cardiac defects.\textsuperscript{321, 355}

5.3.6 Mechanistic link between the R585L GATA6 mutation and AF

The current paradigm for AF posits that the arrhythmia arises due to a complex interplay between focal triggers as well as a susceptible atrial substrate.\textsuperscript{356} Transcription factor variants could potentially play an important role in the formation of a substrate for atrial reentry as well as in the generation of ectopic triggers.

The GATA6 transcription factor activates transcription of a number of downstream target genes. It is plausible that the R585L GATA6 mutation increases susceptibility to AF by altering the expression of genes that influence atrial conduction velocity and/or refractory period. As discussed in Section 1.2.2.3, alterations in the atrial conduction velocity and refractory period promote stabilisation of atrial reentry circuits.

The R585L GATA6 mutation causes an upregulation of expression of the NPPA gene, which encodes ANP. Previous studies have reported that when exposed to pathophysiological levels of ANP, atrial myocytes display altered electrophysiological properties.\textsuperscript{154} More specifically, excessive ANP causes a shortening of the atrial action
potential duration. Further, pathophysiological levels of ANP have the potential to cause structural atrial remodelling due to the pro-apoptotic effect of the peptide. Therefore a potential mechanistic link between the R585L GATA6 mutation and AF could be related to upregulation of NPPA expression. Interestingly, Hodgson-Zingman et al previously identified a gain-of-function mutation in the NPPA gene in a large pedigree with AF. Of note however, the upregulation of NPPA expression caused by the R585L GATA6 mutation is relatively subtle compared to that in the aforementioned study by Hodgson-Zingman et al.

As discussed previously, the GATA DNA binding sequence is found in the promoter region of most of the genes expressed in the heart. Therefore, in addition to the NPPA gene, the R585L GATA6 mutation could potentially increase activation of a number of genes that create a substrate for reentry. Examples include ion channel genes and connexins. A similar mechanism has previously been reported in patients with Holt-Oram syndrome. Posch et al identified a gain-of-function mutation in TBX5 in a pedigree with Holt-Oram syndrome and demonstrated that the mutation results in an upregulation of expression of NPPA, connexin-40 and KCNJ2. Upregulation of the expression of potassium channel genes such as KCNJ2 would be predicted to shorten the action potential, which facilitates reentry.

The most frequent site of origin of the focal triggers that provoke and potentially drive AF is the pulmonary veins. Previous studies have demonstrated that the NKX-2.5
transcription factor, which is highly expressed in the pulmonary venous myocardium, plays an important role in pulmonary vein development.\textsuperscript{346} The GATA transcription factors interact with NKX-2.5.\textsuperscript{357} It is plausible therefore that the R585L \textit{GATA6} mutation alters the interaction profile of NKX-2.5, which in turn may promote pulmonary vein ectopy by causing defective morphogenesis of the pulmonary vein myocardium. A number of studies have demonstrated that patients with AF have anatomical differences in the pulmonary veins including longer muscle sleeves, thicker pulmonary vein myocardial tissue and dilatation of the superior pulmonary veins.\textsuperscript{358} Further, conduction delays in pulmonary veins have been reported to be caused by altered myocyte fibre direction.\textsuperscript{359} Of note however, the R585L \textit{GATA6} mutation has not been demonstrated to alter the interaction profile with NKX-2.5. Therefore, at this stage, this proposed mechanism remains speculative.

\subsection{5.3.7 Limitations}

\subsubsection{Complex discordant phenotypes}

A major assumption in the present study is that the discordant clinical phenotypes in pedigree AF-435 arise due to the R585L \textit{GATA6} mutation. One mutation carrier has an ASD, another has a VSD and three have AF. Further, four mutation carriers only have subtle ECG phenotypes. These discordant phenotypes are potentially a reflection of incomplete and variable penetrance of the R585L \textit{GATA6} mutation. As discussed previously, transcription factor mutations are commonly associated with complex
overlapping phenotypes. It is important to note however that it is also plausible that the discordant phenotypes are unrelated and arise due to distinct genetic variants. Therefore, while the R585L GATA6 mutation is a compelling candidate variant, due to the complex phenotypes in pedigree AF-435, the results would have to be interpreted with caution.

### 5.3.7.2 Incomplete segregation analysis

In the present study, three levels of filtering of candidate variants identified by exome sequencing were used to narrow the search for the causative variant in pedigree AF-435; 1) only nonsynonymous variants were included in downstream analysis, 2) variants that have previously been reported in population databases were excluded from downstream analysis and, 3) variants that were not present in all three affected individuals who underwent exome sequencing were excluded from downstream analysis. Using this strategy, the number of candidate variants was reduced to 18 missense variants. Further prioritisation of the 18 candidate variants was based on evolutionary conservation and a potential role in cardiac function. Using this strategy, the R585L GATA6 variant emerged as a strong candidate variant and was prioritised for segregation and functional analysis. It is important to note however, that we did not perform segregation and/or functional analysis of the other 17 candidate variants. Therefore, while R585L GATA6 is a compelling candidate variant, it may not be the causative variant and the actual causative variant may have been overlooked.
5.3.8 Next steps

5.3.8.1 Screening AF cohorts for GATA6 variants

Based on our finding of a novel GATA6 mutation in pedigree AF-435, it is plausible that genetic variation in the GATA6 gene increases susceptibility to AF in the general population. As discussed previously, candidate-gene studies in Chinese AF cohorts have identified isolated GATA6 mutations.344, 345 Of note however, these studies were performed in relatively small cohorts of patients. A potential next step is to screen a large cohort of patients with early-onset AF for mutations in GATA6. The Massachusetts General Hospital (MGH) AF Study has more than 1200 individuals of European descent with early-onset AF. This represents an ideal cohort for a candidate-gene study to determine the prevalence of GATA6 mutations in the general population.

5.3.8.2 Analysis of interaction of mutant GATA6 with other transcription factors

The GATA4 and GATA6 transcription factors interact physically and functionally to regulate the expression of specific subsets of genes.331 Interaction between GATA transcription factors has been demonstrated to be critical for normal cardiac development.360 Functional interaction between GATA4 and GATA6 requires both the zinc finger domain and the C-terminal activation domain of GATA4. GATA6 shares significant structural homology with GATA4 and is therefore predicted to harbour similar activation domains.324 It is plausible therefore, that the R585L GATA6
mutation, which is located in the C-terminal of the peptide, alters the functional interaction between GATA4 and GATA6. Therefore, a potential next step to more comprehensively characterise the functional effect of the R585L GATA6 mutation is to perform \textit{in vitro} assays investigating the effect of the mutation on cooperative regulation of myocardial gene expression by GATA4 and GATA6.
CHAPTER 6. DISCUSSION

AF is the most common cardiac arrhythmia in clinical practice. AF is associated with significant morbidity and mortality and therefore represents a major burden to healthcare systems. Over the past two decades, we have come to appreciate that AF is a condition with a significant heritable component. The identification of the genetic substrate underlying AF is likely to provide important insights into the molecular mechanisms underlying the arrhythmia.

GWAS have emerged as valuable tools for the identification of common variants underlying complex traits like AF. Since 2007, GWAS have identified nine susceptibility loci for AF. In one of the GWAS, which focused on lone AF, a susceptibility locus was identified at the KCNN3 gene. KCNN3 encodes a calcium-activated potassium channel (SK3) which is abundantly expressed in the heart. In the first part of this study, we sought to investigate whether altered expression of the SK3 channel increases susceptibility to AF by characterising a mouse model with variable expression of the channel. We demonstrated that overexpression of the SK3 channel results in an increased susceptibility to AF. Interestingly, these mice also display a high incidence of sudden death due to heart block.

Despite the successes of GWAS, the common variants reported to date only account for a fraction of the heritability of AF. In recent years, increasing evidence has emerged to suggest that a significant proportion of the ‘missing heritability’ of complex traits like AF is accounted for by rare variants. Therefore, pedigree-based
Chapter 6. Discussion

genetic studies represent a potentially effective strategy for the identification of some of the missing heritability of AF. With the emergence of enrichment techniques and next-generation sequencing technology, it has become possible to simultaneously sequence the entire protein coding region of the genome, or the exome. In the second part of this study, we took advantage to two large, multigenerational families with AF from the MGH-AF study in an attempt to identify novel causative mutations for the arrhythmia. In the first AF pedigree, we used a two-step strategy combining linkage analysis and exome sequencing. We identified a novel locus for AF on chromosome 1p21.1-q22. However the causative mutation at this locus remains elusive. Exome sequencing of multiple affected individuals in the second AF pedigree identified a compelling mutation in the transcription factor gene GATA6.

The findings of the present study provide further insights into the genetic substrate underlying AF, however much work remains. The following section discusses these findings in the context of existing evidence and outlines future areas of research in the field.

6.1 Moving from association to mechanism in the post-GWAS era

Functional studies aimed at characterising the mechanistic links between GWAS discoveries and disease pathogenesis represent the cornerstone of successful translation. However, functional validation of GWAS findings poses a significant challenge. This point is underscored by the fact that to date, more than a thousand...
GWAS have identified close to 4,000 risk variants for a range of different traits, and yet only a handful of these variants have been convincingly functionally validated.\textsuperscript{361, 362} Despite the challenges, or indeed because of the challenges, there is a need for greater investment into post-GWAS functional studies.\textsuperscript{363, 364}

The mechanistic links between the risk variants identified in AF GWAS and arrhythmia pathogenesis are only beginning to be elucidated. The most comprehensively characterised AF locus in this context is 4q25, which was first identified in 2007. The risk variants at the 4q25 locus lie in proximity to the \textit{PITX2} gene. \textit{PITX2} plays a role in asymmetric morphogenesis of the heart, suppression of sinus node formation in the left atrium and, possibly most importantly, development of the pulmonary vein myocardium.\textsuperscript{346} Therefore, \textit{PITX2} represents a strong candidate gene for AF.

A number of studies in murine models have investigated the effect of variable expression of \textit{PITX2} on arrhythmia susceptibility. Homozygous knockout of \textit{PITX2} is lethal \textit{in utero} and is associated with multiple structural cardiac defects.\textsuperscript{346} Heterozygote \textit{PITX2} knockout on the other hand is associated with an increased susceptibility to AF.\textsuperscript{365, 366} Further, haploinsufficiency of \textit{PITX2} results in an abbreviated action potential, a reduction in the action potential amplitude, and a more depolarised resting membrane potential, changes that would be predicted to create a proarrhythmogenic substrate in the atrium.\textsuperscript{365, 367} Interestingly, \textit{PITX2} expression is also downregulated in atrial tissue from human subjects with AF.\textsuperscript{367}
Chapter 6. Discussion

These studies strengthen the argument that PITX2 is the causative gene at the 4q25 locus.

It is important to note that while the aforementioned studies have demonstrated functional links between the PITX2 gene and arrhythmia susceptibility, a fundamental question remains unanswered; what is the mechanistic link between the risk SNPs at the 4q25 locus and the function and/or expression of PITX2? The SNPs at the 4q25 locus do not map to alterations in amino acid structure. Therefore altered protein function is very unlikely to underlie the observed GWAS association. A number of previous studies have demonstrated that common variants that underlie complex traits like AF influence disease susceptibility by altering the quantity of gene expression.\(^\text{283}\) Therefore, it is possible that the causal variants at the 4q25 locus mediate their effect at a transcriptional level, through altered function of regulatory elements. Potential post-GWAS analysis strategies for the identification of the causal variants, regulatory elements and the putative target genes are discussed here.

The initial step in post-GWAS analysis commonly involves fine mapping of a GWAS locus. As discussed in Section 1.3.4.2.2, GWAS are designed to identify associations between marker SNPs, which are chosen to capture LD structure, and disease pathogenesis. Therefore, GWAS rarely identify causal variants directly.\(^\text{361}\) Fine mapping is a strategy designed to refine the GWAS signal in an attempt to identify the causal variant. The technique involves genotyping all known SNPs in an LD block which
contains the marker SNP and associating them with disease. Fine mapping allows one to identify a sub segment of the LD block which is more strongly associated to the disease phenotype, and is therefore likely to harbour the causal variant.\textsuperscript{113}

Lubitz \textit{et al} recently performed fine mapping of the 4q25 locus. In addition to the initially reported SNP (rs2200733), they identified two additional signals that were independently associated with AF. The variants were associated with a graded risk of AF which corresponded to the number of risk variants present.\textsuperscript{368} Interestingly, the variants were located in a region that is highly conserved though evolution, an observation that suggests that the region may harbour non-coding regulatory elements that alter the expression of nearby genes.

An important step in understanding how candidate causal variants identified by fine mapping affect gene function involves characterising the regulatory landscape of the GWAS locus.\textsuperscript{361} A number of techniques can be used to identify potential regulatory elements at these loci. A simple strategy is to search for regions of phylogenetic conservation as regulatory elements, which include enhancers, promoters, silencers and insulators, are typically highly conserved through evolution. Other techniques include DNase hypersensitivity assays and chromatin immunoprecipitation sequencing.\textsuperscript{369, 370}
Chapter 6. Discussion

Once potential regulatory elements have been identified, the next step involves determining whether candidate causal variants alter the function of these elements. Cell-based assays represent a useful tool for initial high-throughput screening for regulatory element function. More detailed characterisation can subsequently be performed using *in vivo* models such as fish and mice. Of note, investigators in Dr. Ellinor’s laboratory have recently screened the entire 4q25 locus for potential regulatory elements. A number of compelling regulatory elements have been identified and experiments are currently ongoing to determine whether the risk-associated SNPs at the 4q25 locus alter the function of these elements.

Once SNPs which alter the function of regulatory elements have been identified, the next step involves identifying the target genes for these regulatory elements. Potential strategies for the identification of target genes include; 1) knockout of the regulatory elements in murine models followed by genome-wide expression profiling, 2) performing chromatin conformation capture, with the candidate regulatory sequence as bait, and 3) expression quantitative trait loci (eQTL) analysis, which involves correlating different genotypes at the disease-associated locus with abundance of candidate gene transcripts. The ultimate goal of these experiments is to link the regulatory elements back to AF pathogenesis.

The focus of the present study was the AF locus on chromosome 1q21. The most significantly associated GWAS signal at this locus lies within the KCNN3 gene. We
demonstrate that in a murine model, overexpression of KCNN3 results in an increased susceptibility to AF. Our findings support the notion that KCNN3 is the causative gene at the 1q21 locus. However, as highlighted in the discussion above, functional analysis of a candidate gene in an animal model represents only one aspect of post-GWAS analysis. More comprehensive analysis is necessary in order to assign causation to the KCNN3 gene at the 1q21 locus. The studies relating to the 4q25 (PITX2) locus above provide a useful template to guide further functional characterisation of the 1q21 (KCNN3) locus. In the longer term, a similar strategy can also be adopted to characterise the remaining seven AF GWAS loci.

6.2 The question of missing heritability

While GWAS have successfully identified multiple risk variants associated with AF, they are associated with a number of important limitations. Firstly, GWAS are not designed to identify rare variants. Secondly, due to stringent significance thresholds applied to GWAS, a number of common variants that could potentially play a role in disease pathogenesis maybe overlooked. Finally, current GWAS arrays are not effective at detecting structural variation in the genome. As a result, a significant proportion of the heritability of AF remains unaccounted for.

The role of rare variants in the pathogenesis of complex traits like AF has recently received much attention. With the emergence of exome sequencing technology, it has become possible to simultaneously sequence the entire protein-coding region of the
Chapter 6. Discussion

genome. These developments have fundamentally accelerated our ability to identify rare disease-causing variants. In the present study, we focused on Mendelian forms of AF and used exome sequencing to identify a novel mutation in the transcription factor gene, \textit{GATA6}. Further, efforts are currently ongoing in Dr. Ellinor’s laboratory to determine the prevalence of \textit{GATA6} mutations in an AF population. In a separate pedigree, we identified a novel locus for AF however we did not identify a coding mutation. This observation may point to a potential role for non-coding mutations in the pathogenesis of AF.

Recently, in an attempt to identify variants of intermediate penetrance that may underlie AF, exome sequencing has been extended to population-based studies. The population-based exome sequencing projects have also facilitated the development of the exome chip, which is an array designed to capture common and rare coding variants identified by exome sequencing. The exome chip is a cost-effective way to detect low frequency coding variants in larger-scale population studies. In the longer term, as sequencing costs fall, it will be possible to perform whole-genome sequencing in AF populations to identify additional coding as well as non-coding variants.\textsuperscript{217}

Some of the missing heritability of AF is likely to be accounted for by larger numbers of common variants that have thus far been overlooked in GWAS.\textsuperscript{215} It is important to note that to date, GWAS in AF have largely been restricted to populations of European
descent. Due to the differences in the linkage disequilibrium structure of the genome between different races, GWAS in other ethnic populations have the potential to uncover a significant proportion of the missing heritability of AF. Further, performing GWAS in larger populations is likely to uncover variants that may have been overlooked due to stringent significance thresholds.215

Structural variants in the genome, which include tandem repeat sequences, insertions and deletions, CNV, translocations, and inversions, may account for a proportion of the unexplained heritability of complex traits.215 Structural genomic research is currently in its early stages and therefore identification of disease-causing structural variants represents a challenge. In recent years, researchers have attempted to integrate analysis for structural variation into analysis of GWAS data and exome sequencing data.215 However, these approaches are associated with significant limitations. In the future, investigation of the role of structural variation at a population level will require the development of novel technologies.

Overall, rare variants are likely to be important drivers of the heritability of AF. Familial forms of AF represent an important resource for the identification of rare variants. In the future, the identification of these variants will allow for more comprehensive characterisation of the genetic substrate underlying AF.
6.3 Clinical relevance and future therapeutic perspectives

6.3.1 Novel therapeutic targets for AF

The majority of the traditional pharmacological agents used for the treatment of AF target ion channels. Examples include sodium channel blockers, calcium channel blockers and potassium channel blockers. Most of these drugs exhibit only moderate efficacy and are associated with significant side effects, which may limit their use. Therefore, there exists a need for novel, more efficacious therapeutic agents for AF with fewer adverse side effects.

SK channels have recently emerged as potential therapeutic targets for AF. Based on the atrial-specific expression profile of the SK1 and SK2 channel subtypes, it has been proposed that SK channel modulators may suppress AF without ventricular proarrhythmogenic side effects. SK channel blockers such as UCL1684, \( N-(pyridin-2-yl)-4-(pyridin-2-yl) \) thiazol-2-amine, and NS8593 have been demonstrated to suppress AF in animal models. The findings of the present study provide further evidence to suggest that SK channels, and in particular the SK3 channel, is a promising drug target for AF. It is important to note however that our findings also suggest that altered SK3 channel function has the potential to influence conduction in the ventricle and in the atrioventricular conduction system. Therefore, SK channel modulators may be associated with more side effects than previously anticipated. Further research is
Chapter 6. Discussion

necessary to fully characterise the efficacy and side effect profile of SK channel modulators.

Based on the findings of the present study and a number of recent reports, transcription factors are emerging as promising therapeutic targets for AF. Drugs that target cardiac-specific transcription factors could potentially allow manipulation of multiple genes involved in the formation of a proarrhythmogenic substrate in the atrium. Of note, transcription factors have been demonstrated to play important roles during structural and electrical remodelling of the atrium, which is known to increase susceptibility to AF. Therefore, innovative therapies aimed at transcription factors may protect the atrium from adverse remodelling. However, further research is necessary to elucidate the complex interplay between transcription factors and target genes before drugs aimed at altering atrial gene regulation can be developed.

In the future, next-generation sequencing techniques such as exome sequencing and whole-genome sequencing are predicted to uncover multiple additional genetic variants and biological pathways underlying AF. For instance, whole-genome sequencing in pedigree AF-325 may uncover a non-coding mutation at the AF locus on chromosome 1. The identification of these pathways may lead to the emergence of additional therapeutic targets for the arrhythmia.
6.3.2 Risk prediction of AF

In recent years, the potential use of variants identified by GWAS to predict risk of developing disease among healthy individuals has received much attention. For some phenotypes, initial attempts have been made to use the most significantly associated SNPs from GWAS as predictors of risk.\(^{373}\) However, these studies have had limited success. A potential reason is that for most complex traits, the risk of developing disease is influenced by multiple genes, each with a relatively small effect size.\(^{374}\) By implication therefore, large numbers of SNPs are required to significantly enhance risk prediction.

AF GWAS to date have identified risk SNPs at nine loci and most of these SNPs are associated with modest effect sizes. These variants account for a relatively small proportion of the heritability of AF.\(^{215}\) Therefore, before genotype-based risk prediction algorithms can be developed for AF, a larger proportion of the ´missing heritability´ would have to be identified.

As discussed in Section 6.2, for most common diseases, including AF, a significant proportion of the unexplained heritability is likely to be accounted for by rare and low frequency genetic variants.\(^{215}\) Therefore, pedigree-based studies, such as the ones described in the present study are predicted to uncover additional variants that may contribute to risk prediction. Further, population-based exome sequencing projects are predicted to uncover low frequency variants with intermediate penetrance.
Chapter 6. Discussion

The ultimate goal is to develop clinically applicable AF risk algorithms that incorporate multiple-risk alleles with established risk factors such as hypertension and diabetes to accurately determine the individual risk of developing AF.

6.3.3 Risk prediction of congenital heart disease

The diagnosis and management of congenital heart diseases has advanced significantly in the past few decades. Despite these advances however, these conditions are associated with significant morbidity and mortality. Identification of the genetic basis of congenital heart disease is of great importance in determining clinical prognosis of affected individuals, genetic screening in family members of affected individuals, and for prenatal diagnosis.375

Over the past decade, multiple genetic mutations have been identified in pedigrees with congenital heart disease. The majority of these mutations reside in transcription factor genes.376 The identification of a GATA6 mutation in the present study further expands on the spectrum of transcription factor mutations in congenital heart disease. Genetic testing for single gene mutations in congenital heart disease is still currently in the research phase.375 In the future, rapid advances in genotyping technology are likely to uncover multiple additional variants which will not only enhance risk prediction but may also inform therapeutic decisions.

6.4 Conclusions
Despite the high prevalence of AF, the molecular basis of the arrhythmia remains incompletely understood. As a consequence, current therapeutic interventions are associated with significant limitations. AF is a condition with a significant genetic component. In recent years, the emergence of next-generation sequencing technology has spawned an exciting new era of research in AF genetics. Large-scale GWAS have identified multiple common variants that confer increased susceptibility to AF. These variants are located in the vicinity of previously unsuspected genes which could represent novel therapeutic targets. However, before GWAS findings can be translated into more effective therapies, a more detailed understanding of the molecular basis of the associations is necessary.

*KCNN3* is a compelling candidate gene identified in a GWAS of lone AF. In the present study, we demonstrate that overexpression of *KCNN3* creates a proarrhythmogenic substrate in the atrium. These results provide a rationale for considering this channel as an important potential cause of AF. An unexpected additional finding of this study was that variable expression of *KCNN3* also has a significant effect on impulse propagation in the atrioventricular conduction system and the ventricular myocardium.

Despite the successes of GWAS, the variants indentified to date only account for a proportion of the heritability of AF. A large proportion of the ‘missing heritability’ is likely to be accounted for by rare variants. Therefore, the focus in AF genetics
Chapter 6. Discussion

research has once again shifted to familial forms of the arrhythmia. In the present study, we identified a compelling mutation for AF and abnormal cardiac septation in the transcription factor gene GATA6 using pedigree-based exome sequencing. 

Transcription factors play critical roles in regulating the expression of genes involved in cardiac development and have also been implicated in AF. Using classical Mendelian genetics, we also identified a novel locus for AF in a second pedigree. However exome sequencing failed to identify a causative coding mutation at this locus. This finding may suggest a potential role for non-coding variants and/or structural variants in the pathogenesis of AF.

The long-term goal of AF genetics research is to translate knowledge of genotype to improve clinical outcomes in AF patients. Two major areas of development in this context are risk prediction and drug development. The goal of risk prediction algorithms is to identify individuals who are at high risk of developing AF or arrhythmia related complications. Enhanced risk prediction may help target preventative efforts. Efforts are already currently underway to develop genotype-based risk stratification algorithms. As discussed above, the majority of the pharmacological agents currently used for the treatment of AF exhibit only moderate efficacy and are associated with significant side effects. The identification of novel genes for AF may be exploited to develop novel, more efficacious pharmaceutical agents for the arrhythmia. However, much work remains before personalised genotype-based medicine becomes a reality.
APPENDIX

Table A.1 - *Kcnn3* genotyping primer sequences

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence 5' → 3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutant reverse</td>
<td>GGA GTA CTC ACC CCA ACA GC</td>
</tr>
<tr>
<td>Common</td>
<td>AGA GGA CTG GAG CCA AGA CA</td>
</tr>
<tr>
<td>Wild type reverse</td>
<td>ACA GAC CAG GAT GGA CAA GC</td>
</tr>
</tbody>
</table>

Table A.2 – *Kcnn3* genotyping PCR reaction components

<table>
<thead>
<tr>
<th>Reaction component</th>
<th>Volume (μl)</th>
<th>Final concentration</th>
<th>Total volume (μl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 x AB PCR buffer II</td>
<td>1.20</td>
<td>1.00 x</td>
<td>1.20</td>
</tr>
<tr>
<td>ddH₂O</td>
<td>3.90</td>
<td>-</td>
<td>3.90</td>
</tr>
<tr>
<td>5 U/μl Taq DNA polymerase</td>
<td>0.12</td>
<td>0.05 U/μl</td>
<td>0.12</td>
</tr>
<tr>
<td>2.5 mM Dntp</td>
<td>0.96</td>
<td>0.20 mM</td>
<td>0.96</td>
</tr>
<tr>
<td>5 mM DNA loading dye</td>
<td>1.66</td>
<td>0.69 mM</td>
<td>1.66</td>
</tr>
<tr>
<td>25 mM MgCl₂</td>
<td>0.96</td>
<td>2.00 mM</td>
<td>0.96</td>
</tr>
<tr>
<td>DNA (20 ng/μl)</td>
<td>2.00</td>
<td>-</td>
<td>2.00</td>
</tr>
<tr>
<td>20 μM mutant reverse primer</td>
<td>0.20</td>
<td>0.33 μM</td>
<td>0.20</td>
</tr>
<tr>
<td>20 μM common primer</td>
<td>0.40</td>
<td>0.67 μM</td>
<td>0.40</td>
</tr>
<tr>
<td>20 μM wild type reverse primer</td>
<td>0.60</td>
<td>1.00 μM</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Table A.3 – *Kcnn3* genotyping PCR cycling protocol

<table>
<thead>
<tr>
<th>Step</th>
<th>Temp °C</th>
<th>Time</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>94</td>
<td>3 min</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>94</td>
<td>30 sec</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>68</td>
<td>1 min</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>72</td>
<td>1 min</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-</td>
<td>Step 2-4, 35 times</td>
</tr>
<tr>
<td>6</td>
<td>72</td>
<td>2 min</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>-</td>
<td>Hold</td>
</tr>
</tbody>
</table>

276
### Table A.4: RT-PCR - Reaction mixture for first step of cDNA synthesis

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume (µl)</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total RNA</td>
<td>8</td>
<td>5 µg</td>
</tr>
<tr>
<td>Random hexamers</td>
<td>1</td>
<td>50 ng</td>
</tr>
<tr>
<td>dNTP mix</td>
<td>1</td>
<td>10 mM</td>
</tr>
</tbody>
</table>

### Table A.5: RT-PCR - Reaction mixture for second step of cDNA synthesis

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume (µl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 x RT buffer</td>
<td>2</td>
</tr>
<tr>
<td>25 mM MgCl2</td>
<td>4</td>
</tr>
<tr>
<td>0.1 M DTT</td>
<td>2</td>
</tr>
<tr>
<td>RNaseOUT (40 U/µl)</td>
<td>1</td>
</tr>
<tr>
<td>SuperScript III RT (200 U/µl)</td>
<td>1</td>
</tr>
<tr>
<td>Reaction mixture from step 1</td>
<td>10</td>
</tr>
</tbody>
</table>

### Table A.6: RT-PCR - Thermal cycling parameters cDNA synthesis

<table>
<thead>
<tr>
<th>Temp (°C)</th>
<th>Time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>25.0</td>
<td>10</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>85</td>
<td>5</td>
</tr>
</tbody>
</table>

### Table A.7 – RT-PCR - Primer sequences

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence 5’ → 3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>KCNN3 F</td>
<td>GGC ATC ATG GGT GCA GGA TGC</td>
</tr>
<tr>
<td>KCNN3 R</td>
<td>GCG GCA TTC TTG ATC CTG TTG GTG</td>
</tr>
<tr>
<td>B actin F</td>
<td>TCT ACG AGG GCT ATG CTC TCC</td>
</tr>
<tr>
<td>B actin R</td>
<td>TCT TTG ATG TCA CGC ACG ATT TC</td>
</tr>
</tbody>
</table>
### Table A.8: RT-PCR - Two-step mastermix

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
<th>Final Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>2x SYBR Green PCR Master Mix</td>
<td>25</td>
<td>1X</td>
</tr>
<tr>
<td>Forward Primer</td>
<td>1</td>
<td>100 nM</td>
</tr>
<tr>
<td>Reverse Primer</td>
<td>1</td>
<td>100 nM</td>
</tr>
<tr>
<td>Template</td>
<td>6 (50 ng/μl)</td>
<td>300 ng</td>
</tr>
<tr>
<td>Water</td>
<td>17</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>

### Table A.9: RT-PCR - Thermal Cycling Parameters

<table>
<thead>
<tr>
<th>Cycle (40 cycles)</th>
<th>Denature</th>
<th>Anneal/Extend</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temp</td>
<td>95.0 °C</td>
<td>60.0 °C</td>
</tr>
<tr>
<td>Time</td>
<td>15 sec</td>
<td>1 min</td>
</tr>
</tbody>
</table>
## Table A.10. Primer sequences for transcription factor candidate gene screening

<table>
<thead>
<tr>
<th>Gene</th>
<th>Exon</th>
<th>F/R</th>
<th>Sequence 5′ → 3′</th>
</tr>
</thead>
<tbody>
<tr>
<td>NKX-2.5</td>
<td>1</td>
<td>F</td>
<td>CACCATGCAGGGAAGCTG</td>
</tr>
<tr>
<td>NKX-2.5</td>
<td>1</td>
<td>R</td>
<td>AGTTCTTTGGGACGAAGACG</td>
</tr>
<tr>
<td>NKX-2.5</td>
<td>2 (amplicon 1)</td>
<td>F</td>
<td>CCCTACCATTACTGTGGCG</td>
</tr>
<tr>
<td>NKX-2.5</td>
<td>2 (amplicon 1)</td>
<td>R</td>
<td>CGAGTCCCCCTAGGCATGG</td>
</tr>
<tr>
<td>NKX-2.5</td>
<td>2 (amplicon 2)</td>
<td>F</td>
<td>AGACTCTGGAGCTGGTG</td>
</tr>
<tr>
<td>NKX-2.5</td>
<td>2 (amplicon 2)</td>
<td>R</td>
<td>CGAGAGTCAGGGAGCTGGT</td>
</tr>
<tr>
<td>TBX20</td>
<td>1</td>
<td>F</td>
<td>TGTTTGGGTCTTTGTCTCC</td>
</tr>
<tr>
<td>TBX20</td>
<td>1</td>
<td>R</td>
<td>AGGGGCACAGACGGAT</td>
</tr>
<tr>
<td>TBX20</td>
<td>2</td>
<td>F</td>
<td>TCCAGTGACATTGTTGTTATG</td>
</tr>
<tr>
<td>TBX20</td>
<td>2</td>
<td>R</td>
<td>ACTACCCAGGGAGTGTCTGTCC</td>
</tr>
<tr>
<td>TBX20</td>
<td>3</td>
<td>F</td>
<td>GAGTCAGACCTTCCCTTCCTCC</td>
</tr>
<tr>
<td>TBX20</td>
<td>3</td>
<td>R</td>
<td>AGGCTTGGAATGCTCTCTTG</td>
</tr>
<tr>
<td>TBX20</td>
<td>4</td>
<td>F</td>
<td>CCCACTTATATAGTTATGTTGTTCC</td>
</tr>
<tr>
<td>TBX20</td>
<td>4</td>
<td>R</td>
<td>AGACAGTTTTGTGGACCTCG</td>
</tr>
<tr>
<td>TBX20</td>
<td>5</td>
<td>F</td>
<td>CCATTGGGTCTTTCCTGCC</td>
</tr>
<tr>
<td>TBX20</td>
<td>5</td>
<td>R</td>
<td>TGAACATCCATTGGTACTGG</td>
</tr>
<tr>
<td>TBX20</td>
<td>6</td>
<td>F</td>
<td>TGCTATGGAGAAATAAAAC</td>
</tr>
<tr>
<td>TBX20</td>
<td>6</td>
<td>R</td>
<td>GATTCTCTGGTGCAAAGG</td>
</tr>
<tr>
<td>TBX20</td>
<td>7</td>
<td>F</td>
<td>CCTAAGGCTCTCTGTATCTCTG</td>
</tr>
<tr>
<td>TBX20</td>
<td>7</td>
<td>R</td>
<td>CTCATCAACGGACCTGCC</td>
</tr>
<tr>
<td>TBX20</td>
<td>8</td>
<td>F</td>
<td>AGAAAAGTAGATTGGCTCGTG</td>
</tr>
<tr>
<td>TBX20</td>
<td>8</td>
<td>R</td>
<td>TCCTGTGCTCTGGAAAGC</td>
</tr>
</tbody>
</table>
### Table A.10. Primer sequences for transcription factor candidate gene screening

<table>
<thead>
<tr>
<th>Gene</th>
<th>Exon</th>
<th>F/R</th>
<th>Sequence 5’ → 3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBX5</td>
<td>1</td>
<td>F</td>
<td>TCTCTCTGTCCCTCCCCACC</td>
</tr>
<tr>
<td>TBX5</td>
<td>1</td>
<td>R</td>
<td>GAGCAGGAAAGCCAGACTC</td>
</tr>
<tr>
<td>TBX5</td>
<td>2</td>
<td>F</td>
<td>GGGAGGGAATGCCACTAC</td>
</tr>
<tr>
<td>TBX5</td>
<td>2</td>
<td>R</td>
<td>TTCAAGGCACTTTTCTTC</td>
</tr>
<tr>
<td>TBX5</td>
<td>3</td>
<td>F</td>
<td>TAAAATGGATGGAGGCTGC</td>
</tr>
<tr>
<td>TBX5</td>
<td>3</td>
<td>R</td>
<td>TTTTGGAAGAGTTCCAC</td>
</tr>
<tr>
<td>TBX5</td>
<td>4</td>
<td>F</td>
<td>GTGCACTGCGCTACCTCC</td>
</tr>
<tr>
<td>TBX5</td>
<td>4</td>
<td>R</td>
<td>CCAAGAAGGGAGAAACCC</td>
</tr>
<tr>
<td>TBX5</td>
<td>5</td>
<td>F</td>
<td>GGGGAGGATGTTTATCTGG</td>
</tr>
<tr>
<td>TBX5</td>
<td>5</td>
<td>R</td>
<td>GCAGGAAACCTTTGAGATTC</td>
</tr>
<tr>
<td>TBX5</td>
<td>6</td>
<td>F</td>
<td>GCTCATGTCTGAGGCTGTC</td>
</tr>
<tr>
<td>TBX5</td>
<td>6</td>
<td>R</td>
<td>GTTGCTGTGCTACCTG</td>
</tr>
<tr>
<td>TBX5</td>
<td>7</td>
<td>F</td>
<td>CTCTCACACTGCTACCC</td>
</tr>
<tr>
<td>TBX5</td>
<td>7</td>
<td>R</td>
<td>GGGTAGGAACATGTAAGGGG</td>
</tr>
<tr>
<td>TBX5</td>
<td>8 (amplicon 1)</td>
<td>F</td>
<td>CTCCATTTAGCTGCTGG</td>
</tr>
<tr>
<td>TBX5</td>
<td>8 (amplicon 1)</td>
<td>R</td>
<td>AACATCCCTCTCCCCAGC</td>
</tr>
<tr>
<td>TBX5</td>
<td>8 (amplicon 2)</td>
<td>F</td>
<td>AGCAGCTGCAAGGTCCAC</td>
</tr>
<tr>
<td>GBX5</td>
<td>8 (amplicon 2)</td>
<td>R</td>
<td>CTTGGCTACTGCTCTCTCTTC</td>
</tr>
<tr>
<td>GATA4</td>
<td>1 (amplicon 1)</td>
<td>F</td>
<td>TGGCTCTTGACCTGCG</td>
</tr>
<tr>
<td>GATA4</td>
<td>1 (amplicon 1)</td>
<td>R</td>
<td>CCACGTGCTGAGGCGC</td>
</tr>
<tr>
<td>GATA4</td>
<td>1 (amplicon 2)</td>
<td>F</td>
<td>CTGCTACCTCCAGGC</td>
</tr>
<tr>
<td>GATA4</td>
<td>1 (amplicon 2)</td>
<td>R</td>
<td>CTCAAGAGGGCCCTTCG</td>
</tr>
<tr>
<td>GATA4</td>
<td>2</td>
<td>F</td>
<td>TCAGATGTGAGGCTGCGG</td>
</tr>
<tr>
<td>GATA4</td>
<td>2</td>
<td>R</td>
<td>AACCAGAGGATGTCACC</td>
</tr>
<tr>
<td>GATA4</td>
<td>3</td>
<td>F</td>
<td>CACACGGAGGTGGAAG</td>
</tr>
<tr>
<td>GATA4</td>
<td>3</td>
<td>R</td>
<td>GACAAGGGGAGAGACTGAGAGATG</td>
</tr>
<tr>
<td>GATA4</td>
<td>4</td>
<td>F</td>
<td>TTCAATGTGTGAGACTACG</td>
</tr>
<tr>
<td>GATA4</td>
<td>4</td>
<td>R</td>
<td>TGCCCTAAGGAGGATATG</td>
</tr>
<tr>
<td>GATA4</td>
<td>5</td>
<td>F</td>
<td>CGGCTGCTGGTTGTCC</td>
</tr>
<tr>
<td>GATA4</td>
<td>5</td>
<td>R</td>
<td>GTACTAGGC TGCCCTCTGGG</td>
</tr>
<tr>
<td>GATA4</td>
<td>6</td>
<td>F</td>
<td>CAGCTAGACCTCCAGGC</td>
</tr>
<tr>
<td>GATA4</td>
<td>6</td>
<td>R</td>
<td>CCTCCTCTTGTGCTATCC</td>
</tr>
</tbody>
</table>

Abbreviations; F, forward; R, reverse
Table A.11: PCR cycling protocol for candidate gene screening

<table>
<thead>
<tr>
<th>Step</th>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denaturing step</td>
<td>94°C</td>
<td>45 seconds</td>
</tr>
<tr>
<td>Annealing step</td>
<td>50-60°C</td>
<td>30 seconds</td>
</tr>
<tr>
<td>Extension step</td>
<td>72°C</td>
<td>1 minute</td>
</tr>
<tr>
<td>Termination step</td>
<td>4°C</td>
<td></td>
</tr>
</tbody>
</table>

Table A.12 – Primer sequences for confirmation of GATA6 R585L variant

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence 5´ → 3´</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward</td>
<td>ACCAGATTTGTAACCGCTTC</td>
</tr>
<tr>
<td>Reverse</td>
<td>GCAGACACGAGTGGAGTGAG</td>
</tr>
</tbody>
</table>

Table A.13: Comparison of wild type and mutagenic primers for site-directed mutagenesis

<table>
<thead>
<tr>
<th>Primer sequence</th>
<th>Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT GATA6 forward</td>
<td>TCACGTCCCTCCGTCGACCGGATTCCTGGTGCGCC</td>
</tr>
<tr>
<td>mGATA6 forward</td>
<td>TCACGTCCCTCCGTCGATTCCGATTCCTGGTGCGCC</td>
</tr>
<tr>
<td>WT GATA6 reverse</td>
<td>GGCGCACCAGGAAATCCCGTGCACGGAGGACGTGA</td>
</tr>
<tr>
<td>mGATA6 reverse</td>
<td>GGCGCACCAGGAAATCCCGTAGACGACGAGGACGTGA</td>
</tr>
</tbody>
</table>

Abbreviations; WT GATA6, wild type GATA6 sequence; mGATA6, mutant GATA6 sequence. Base substitutions are indicated in red.

Table A.14: Reaction mixture for GATA6 mutant strand synthesis

<table>
<thead>
<tr>
<th>Reaction component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 × reaction buffer</td>
<td>5 µl</td>
</tr>
<tr>
<td>dsDNA template</td>
<td>1 µl (125 ng/µl)</td>
</tr>
<tr>
<td>mGATA6 forward primer</td>
<td>1 µl (125 ng/µl)</td>
</tr>
<tr>
<td>mGATA6 reverse primer</td>
<td>1 µl (125 ng/µl)</td>
</tr>
<tr>
<td>dNTP mix</td>
<td>1 µl</td>
</tr>
<tr>
<td>PfuTurbo DNA polymerase</td>
<td>1 µl (2.5 U/µl)</td>
</tr>
<tr>
<td>ddH2O</td>
<td>40 µl</td>
</tr>
</tbody>
</table>
Table A.15: Thermal cycling conditions for GATA6 mutant strand synthesis

<table>
<thead>
<tr>
<th>Segment</th>
<th>Cycles</th>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>95°C</td>
<td>30 seconds</td>
</tr>
<tr>
<td>2</td>
<td>12–18</td>
<td>95°C</td>
<td>30 seconds</td>
</tr>
<tr>
<td>2</td>
<td>12–18</td>
<td>55°C</td>
<td>1 minute</td>
</tr>
<tr>
<td>2</td>
<td>12–18</td>
<td>68°C</td>
<td>1 minute/kb of plasmid length</td>
</tr>
</tbody>
</table>

Table A.16: Sequencing primers for GATA6 clone

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence 5’→3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vector Primer v1.5</td>
<td>GGACTT TCCAAA ATG TCG</td>
</tr>
<tr>
<td>GATA6 sequencing F primer 1</td>
<td>CTGACGGCCGGCTGGTGCTTG</td>
</tr>
<tr>
<td>GATA6 sequencing F primer 2</td>
<td>CACTCTGCGCCGGCCGGG</td>
</tr>
<tr>
<td>GATA6 sequencing F primer 3</td>
<td>CTCGCCCTACGTGGGGGCC</td>
</tr>
<tr>
<td>GATA6 sequencing F primer 4</td>
<td>CCCATGAATCCAAACTTCACCTT</td>
</tr>
<tr>
<td>Primer XL39</td>
<td>ATTAGGACAAAGGCTGTG</td>
</tr>
</tbody>
</table>

Table A.17: Primers for confirmation of mutations identified by exome sequencing in AF-325

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1orf62 forward</td>
<td>TTT TCT TCC GCC CTG CC</td>
</tr>
<tr>
<td>C1orf62 reverse</td>
<td>GAC GGATGC CCT TCA GTG TC</td>
</tr>
<tr>
<td>CELF3 forward</td>
<td>TGA CTG GTT CTC TGC CTG TG</td>
</tr>
<tr>
<td>CELF3 reverse</td>
<td>GGG AGT GGT GGT TTA GAT GG</td>
</tr>
<tr>
<td>PMVK forward</td>
<td>AAA GGC CGG TGA CGT AGA C</td>
</tr>
<tr>
<td>PMVK reverse</td>
<td>CGA AGA GCT AGG GAA ACA GG</td>
</tr>
<tr>
<td>MRPS21 forward</td>
<td>ACC AAT TGT TTG GTC AAT CG</td>
</tr>
<tr>
<td>MRPS21 reverse</td>
<td>GTT TAT TGC AGA GAA TGG TAA TAGG</td>
</tr>
<tr>
<td>FLG forward</td>
<td>GAT CCC ACT ATG AGC AAT CG</td>
</tr>
<tr>
<td>FLG reverse</td>
<td>AGA CCC AGA CCA CCT CTC AG</td>
</tr>
</tbody>
</table>
Registry of 15,400 Emergency Department Patients in 46 Countries: The RE-LY AF Registry. Circulation 2014.


Bibliography


44. Al-Khatib SM, LaPointe NM, Kramer JM, Califf RM: What clinicians should know about the QT interval. JAMA 2003; 289:2120-2127.


58. Lewis T: Observations upon the state of rapid re-excitation of the auricles. Heart 1921; 8:311-340.


Bibliography


Bibliography


Bibliography


al.: Paired-related homeobox gene Prx1 is required for pulmonary vascular

dilated cardiomyopathy and pulmonary hypertension in knockout mice. Proc

cardiac potassium channel (HERG) by caveolin-1. Biochem Cell Biol 2008;
86:405-415.

Neurol 2011; 101:155-166.

newly identified genomic mutations in a Japanese female patient with
fructose-1,6-bisphosphatase (FBPase) deficiency. Mol Genet Metab 2002;

bradycardia associated with a mutation in the cardiac pacemaker channel. N

192. Frey N, Olson EN: Calsarcin-3, a novel skeletal muscle-specific member of the
277:13998-14004.

lacking calsarcin-1 are sensitized to calcineurin signaling and show accelerated
 cardiomyopathy in response to pathological biomechanical stress. Nat Med

Cardiovasc Electrophysiol 2012; 23:1400-1406.


potassium channels, introductory chapter. IUPHAR database (IUPHAR-DB)
http://www.iuphar-
db.org/DATABASE/FamilyIntroductionForward?familyId=69. Accessed 16/02,
2014.

197. Adelman JP, Maylie J, Sah P: Small-conductance ca(2+)-activated k(+) channels:

intermediate conductance calcium-activated potassium channel. Proc Natl
Acad Sci U S A 1997; 94:11651-11656.


267. Cox DH: Ca(2+)-regulated ion channels. BMB Rep 2011; 44:635-646.


Bibliography


