Molecular Mechanisms in Inherited Arrhythmia Syndromes

By

Matthew James O'Neill

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Human Genetics

June 30th, 2023

Nashville, Tennessee

Approved:

Andrew M. Glazer, Ph.D. (Chair)

Nancy Cox, Ph.D.

Jonathan Brown, M.D.

Alexander Bick, M.D., Ph.D.

Dan M. Roden, M.D.

Dedicated to James Francis Gorman, MD

Acknowledgements

I am fortunate to have many current and former research mentors to thank for their support in getting me to this point. Professor Paul Helquist and Olaf Wiest were extremely generous research mentors throughout my full four years at Notre Dame. I also owe great thanks to Michael Grigalunas and Gang Liu in the Helquist lab, who taught me an incredible amount in those formative years. I am extremely grateful for additional research stays at Universität Heidelberg with Stephen Hashmi and the Scripps Research Institute with Phil Baran, all of which provided excellent preparation for my doctoral studies.

Vanderbilt kindly let me defer the beginning of my MD/PhD to pursue a Fulbright Fellowship in Germany with Josep Cornella, who I first met as a postdoc in the Baran lab at TSRI. This was an intense year with many high highs, and occasional lows. The passion that Pep brought to the lab was infectious, and I continuously aspire to approach my science with a similar sense of urgency. It also provided an opportunity to forge lifelong friendships with my roommate and fellow Fulbright Scholar Jeremy McGale, and lab partner and good friend Tim Riesebeck.

First entering the Vanderbilt MSTP with plans to study organic chemistry and drug discovery, serendipity directed me down an unexpected path into cardiac genetics. Indeed, as an applicant, I had thought I bombed my interview with Dr. Roden, and I could scarcely imagine working with him years later. Fortunately, I returned to his office at the suggestion of my first mentor at Vanderbilt, Dr. Benjamin Shoemaker. Indeed, the 12th floor of Light Hall has been a fantastic scientific home for me throughout these years. Dan has been an excellent mentor to me during this time and has been fully supportive of carving out a path in his lab. He has served as an excellent role model of a physician scientist, and an example of how to seamlessly connect work in the lab to the patient at the individual and biobank levels. I owe special thanks to another close mentor Dr. Andrew Glazer, who has taught me nearly everything I know at the lab bench, and Dr. Brett Kroncke for being another fantastic mentor and friend. Continuous clinical research time and genetic arrhythmia clinic work with Dr. Ben Shoemaker has provided a northern light to guide my efforts in lab and continuously think how research at the bench and the server connect back to patients. All of this was possible with Dan's generous support, which allowed me to fully maximize opportunities for mentorship throughout my training. Moreover, I've been further supported by a generous thesis committee in Dr. Alex Bick, Dr. Nancy Cox, and Dr. Jonathan Brown. I also owe a general thanks to all members of the Roden lab, Glazer lab, and Kroncke lab, who are too numerous to thank individually here. You know who you are.

Lastly, all of these experiences were predicated upon the unwavering support of my family. This started when I was a young, shy child with verbal apraxia who thrived in the Montessori environment with their support. It continued when they sent me to Notre Dame, and were sympathetic to the time away during certain holidays and summers of research each year. Having their support has been an incredible benefit that I will never take for granted. I was fortunate to meet my future wife, Missy, during our first year of medical school. It has been incredible moving through this process together and

complementing each other with mutual interests in academic medicine. We were married in 2022, surrounded by friends from these experiences over the last decade. I'm also very grateful for the new family I have in the Kimlingers.

Lastly, I have been fortunate to receive support for my studies through an American Heart Association Predoctoral Fellowship (AHA 907581) and an NIH F30 Predoctoral Fellowship (NIH 1F30HL163923-01). The Vanderbilt MSTP has provided longitudinal support through the NIH T32GM007347 grant. Core facilities at Vanderbilt and additional institutions, as well as excellent technical assistance, are appropriately acknowledged in the chapters to follow.

TABLE OF CONTENTS

Page

CHAPTER 1: INTRODUCTION 1
Genetic determinants of human disease. 1
Inherited arrhythmia syndromes. 5
Clinical interpretation of genetic variants. 10
Functional investigations of variant effect. 13
Chapter descriptions. 21
CHAPTER 2: DOMINANT NEGATIVE EFFECTS OF *SCN5A* VARIANTS 23
Introduction. 24
Methods. 26
Selection of Variants. 26
SCN5A Mutagenesis 26
Description of Cell Lines. 27
Generation of Variant Cell Lines. 28
Automated Patch Clamping. 29
Case-control analysis. 29
Data Analysis 30
Results 31
Homozygous and Heterozygous Measurements of LoF Variants 31
Homozygous and Heterozygous Measurements of Partial LoF Variants 33
Coupled gating in heterozygous expression 34
Elevated BrS Risk Among Dominant Negative Variants Heterozygotes 35
Structural Distribution of Dominant Negative Variants 37
Discussion 39
Dominant Negative Effect Among Most Missense LoF SCN5A Variants 39
Clinical Implication of Dominant Negative Heterozygote Status 40
High-throughput Electrophysiological Assays to Study Dominant Negative Effects . 42
Limitations 42
Conclusions 42
CHAPTER 3: BAYESIAN ANALYSIS OF MENDELIAN CHANNELOPATHY MISSENSE VARIANTS 43
Introduction 45
Methods 47
General Information 47
Description of Methodology 47
Statistical Analyses 50
Curation of Covariates and Clinical Data 51
Downstream analyses 53

Description of Code. 53
Description of Functional Covariates. 54
Results 55
Summary of Bayesian Penetrance Methodology 55
LQT1-KCNQ1 Penetrance Estimates 56
LQT3-SCN5A Penetrance Estimates 60
Bayesian Estimates Reflects Gene-specific Penetrance Among Channelopathies. 60
Distribution of Posterior Estimates within Individual ClinVar Annotations. 62
Structural Implications of Variable Penetrance. 63
Discussion 66
Bayesian Penetrance Estimates for Mendelian Arrhythmias 66
Variable Penetrance Among Genotype-Phenotype Pairs 67
Continuous Bayesian Penetrance Provides Higher Resolution Variant Interpretation. 68
Penetrance Heterogeneity within Structured Functional Domains 69
Complementary Approaches 70
Precision Medicine Applications. 71
Limitations 72
Conclusion 72
CHAPTER 4: FUNCTIONAL ASSAYS RECLASSIFY SUSPECTED SPLICE-ALTERING VARIANTS OF UNCERTAIN SIGNIFICANCE IN MENDELIAN CHANNELOPATHIES 73
Introduction 74
Methods 77
General Methods 77
Minigene Construct Synthesis and Analysis 77
Analysis of CRIPSR-Cas9-edited iPSC-CMs 79
SpliceAI Analysis 80
ACMG Reclassification 80
Results 82
Minigene Assays Reveals Aberrant Splicing 82
CRISPR/Cas9 and iPSC-CMs Reveal Aberrant Splicing 86
SpliceAI Scores of Affected Variants 88
ACMG Reclassification of Putative Splice-altering Variant. 89
Discussion 90
Splicing as a Disease Mechanism 90
Functional Assays Reveal Aberrant Splicing 90
Aberrant Splicing-associated Cardiac Morbidity 91
In silico *Prediction Concordance* 92
Therapeutic Opportunities 92
Limitations 93

Conclusion 94
CHAPTER 5: MULTICENTER CLINICAL AND FUNCTIONAL EVIDENCE RECLASSIFIES A RECURRENT NON-CANONICAL *FILAMIN C* SPLICE-ALTERING VARIANT 95
Introduction 97
Methods 99
Clinical Evaluation 99
Illegitimate Transcription 99
Generation of iPSCs 99
iPSC-CM Differentiation 100
NMD-inhibition Experiments 100
qPCR and RNA-seq 101
Patch-clamp Electrophysiology 101
Results 103
Clinical Phenotypes by Family 103
FLNC c.970-4A>G Variant Characteristics 109
Transcriptional Studies of Patient-derived iPSC-CMs 110
Electrophysiological Studies of iPSC-CMs 112
ACMG Variant Classification 113
Discussion 114
Spectrum of *FLNC* Variant Phenotypes by Genotype 115
Utility of Functional Assays to Facilitate Variant Interpretation 116
Recurrence of the intronic variant 117
Limitations 117
Conclusions 118
CHAPTER 6: PARSE-SEQ: A CALIBRATED HIGH-THROUGHPUT ASSAY TO FACILITATE CLINICAL INTERPRETATION OF PUTATIVE SPLICE-ALTERING VARIATION 119
Introduction 121
Methods 123
Selection of variants 123
Twist plasmid constructs 123
Library barcoding 124
Assembly 124
Splicing assay and Illumina library prep 124
Computational pipeline 125
Statistical analysis and data availability 126
ACMG OddsPath calculation 126
SpliceAI implementation 127
ACMG variant classification 128

High-throughput electrophysiology 128
iPSC-CM maintenance and differentiation 128
CRISPR-Cas9 gene editing 128
Results 130
Methodology overview and feasibility 130
Assay implementation 131
Functional studies and SpliceAI 133
ACMG assay calibration 134
Variant interpretation and reclassification 135
Cryptic splicing effects of exonic variants 136
Generation of cell lines to investigate effects at endogenous locus 137
Discussion 138
ParSE-seq summary 138
Incorporating splicing assays into the genetic testing pipeline 138
Comparison to previous high-throughput assays 139
Future directions 140
Limitations 140
Conclusions 141
CHAPTER 7: CONCLUSIONS, LIMITATIONS, AND FUTURE DIRECTIONS 142
Summary 142
Synthesis of Efforts 143
Limitations 145
Future Directions 146
REFERENCES 150

LIST OF TABLES

Table 4.1 Summary of variants with aggregate SpliceAI scores, assay results, case and control frequencies, and ACMG reclassifications. 89
Table 5.1 Clinical phenotypes of c.970-4A>G probands....104
Table 5.2 Genes tested by each site 107

LIST OF FIGURES

Figure 1.1 Variant frequency and effect size distributions 1
Figure 1.2 Schematic of action potentials throughout cardiac tissue. 5
Figure 1.3 Complexity of cardiac ion channel currents in cardiomyocytes 7
Figure 1.4 Overview of ACMG criteria 10
Figure 1.5 Nomenclature and regulation of splicing 16
Figure 1.6 Methods for studying splicing in cardiac disease genes 18

Figure 2.1 Stable cell lines used to study *SCN5A* variants 27
Figure 2.2 Example raw traces from homozygous and heterozygous cell lines 31
Figure 2.3 Homozygous and heterozygous sodium peak current measurements 32
Figure 2.4 Western blot of selected SCN5A variants 33
Figure 2.5 Homozygous and heterozygous sodium peak current measurements among partial loss-of-function variants 34
Figure 2.6 Raw Current-voltage traces for five variants with high shifts in voltage of half activation 35
Figure 2.7 Evidence for coupled gating among variant channels in heterozygous expression 36
Figure 2.8 Case-control analysis by variant class 37
Figure 2.9 Spatial distribution of variants by class 38

Figure 3.1 Detailed schematic of Bayesian penetrance methodology 50
Figure 3.2 Overview schematic of Bayesian penetrance methodology 52
Figure 3.3 Empirical penetrance and penetrance density by residue 57
Figure 3.4 Statistical associations of covariates and empirical penetrance 58
Figure 3.5 Bayesian prior and posterior point estimate probabilities and uncertainty by residue 59
Figure 3.6 Histogram distributions of variant penetrance for Channelopathy genotype-phenotype pairs 61
Figure 3.7 Penetrance among ClinVar-annotated variants 63
Figure 3.8 Structural basis of penetrance and ‘hot-spot’ domain heterogeneity for KCNQ1 and KCNH2 64
Figure 3.9 BrS and LQT3 Penetrance density reflected upon SCN5A 65

Figure 4.1 Aberrant splicing from *cis*-genetic variation 75
Figure 4.2 Minigene assays and studies of *SCN5A* variants 83
Figure 4.3 Minigene analysis of *KCNH2* and *KCNQ1* variants 84
Figure 4.4 Example raw gels and traces for *KCNQ1* variants 85
Figure 4.5 Variant functional analyses through CRISPR-edited iPSC-CMs 87

Figure 5.1 Pedigrees and phenotypes of *FLNC* c.970-4A>G families 103
Figure 5.2 Exercise stress test of USA proband. 105
Figure 5.3 *FLNC* c.970-4A>G variant characteristics 109
Figure 5.4 *FLNC* variant c.970-4A>G iPSC-CM transcriptional investigations. 111

Figure 5.5. Cellular electrophysiology of *FLNC* c.970-4A>G proband and healthy control iPSC-CMs. 113

Figure 6.1 Splice-altering variation and assay schematic....130
Figure 6.2 Results of ParSE-seq assay in HEK cells and iPSC-CMs 132
Figure 6.3 Comparison of ParSE-seq data and *in silico* SpliceAI scores....134
Figure 6.4 Assay calibration and variant interpretation....135
Figure 6.5 cDNA-based functional interrogation of putative splice-altering variants....136

CHAPTER 1

INTRODUCTION

A. Genetic Determinants of Human Disease

The heritability of disease has long been recognized in clinical medicine since the time of Hippocrates. In contrast to the preceding millennia of descriptive speculation of heritability, the last 70 years have witnessed phenomenal discoveries which helped build a mechanistic understanding of the heritability of disease[1]. These enabling advances include the identification of the structure and function of DNA, deciphering the molecular basis of monogenic diseases, sequencing of human genomes, and recently, editing the genome for research and therapeutic purposes. From these discoveries and technologies, we have learned that multiple types of variants can propagate disease at each step of the fundamental dogma – irrevocably altering the DNA itself, RNA transcript stability or composition, or disruption of the translated protein products. In recent years, clinical genetics has largely focused on two classes of variation to discover disease heritability[2] – 1) high effect size, rare variants, typically associated with 'monogenic disease' or Mendelian disease and 2) low effect size common variants that may themselves provoke disease in aggregate or modulate the penetrance of rare variants. To frame the work in this thesis, I will first provide a brief overview of these two classes of genetic variation. I will then describe inherited arrhythmia syndromes (the diseases on which my work has focused), their genetic underpinnings, functional assays, and the integration of this knowledge into variant interpretation frameworks.

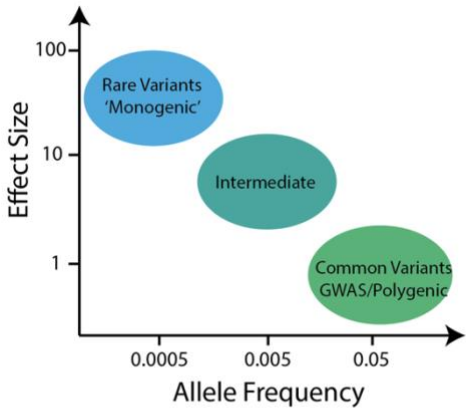


Figure 1.1. Variant frequency and effect size distributions. GWAS (Genome Wide Association Study). Adapted from Manolio et. al. 2009.

Rare variants. During the last century, recognition of the contributions of inherited and *de novo* genetic variation to human disease tremendously accelerated the pace of

basic science and medicine[3]. Scientists and physicians were able to interrogate the genome with unprecedented fidelity, with practitioners establishing a new field of genetic medicine fueled by their research productivity[4]. Illustrative examples include delineating the inheritance patterns and loci of diseases like familial hypercholesterolemia[5], sickle cell disease[6], and cystic fibrosis[7]. These studies collectively established a framework for thinking about rare variants as a major cause of monogenic disease. It also became recognized that predisposition to certain common phenotypes like hypercholesterolemia could be genetically-linked, caused by rare inherited variation in addition to environmental factors such as diet. Building upon successes of mapping through 'linkage' studies, investigators were increasingly empowered to identify causal variation through using larger cohorts of families and improved sequencing technologies, which both benefited from a 'team-science' approach engendered by the success of the Human Genome Project. Additional enabling technologies include sophisticated computational algorithms, next generation sequencing (NGS), genome editing (TALENs, ZFNs, CRISPR-Cas9), and induced-pluripotent stem cells (iPSCs).

Translational Science of Rare Variants. Rare variant, or Mendelian, human genetics has made transformative impacts on the full spectrum of translational science, spanning discoveries in genetic diagnoses to the full drug discovery and development pipeline. Rare variant investigations have inspired the development of toolkits including genome-wide knockouts with CRISPR-Cas9 machinery[8], development of model systems that highly recapitulate human disease[9], direct editing of genomic loci associated with disease in humans *ex vivo* and *in vivo*[10], and optimization of clinical trials and drug development based on participants' underlying genetic subtypes[11]. Prominent examples guiding drug discovery include human studies of patients with gain-of-function (GoF) and loss-of-function (LoF) variants in *PSCK9*[12]. These human genetic observations led to the development of multiple modalities of *PSCK9* inhibitors to recapitulate the genetic knock-out, expanding the pharmacopeia for cardiovascular disease. Dissection of the molecular

consequences of different classes of cystic fibrosis variants led to transformative medicines to treat this crippling disease[11]. Direct editing of genomic loci in a therapeutic context has quickly accelerated from dream to reality. Gene-editing therapeutics for progeria[13], sickle cell disease and beta-thalassemia[14], hypertrophic cardiomyopathy[15], and familial hypercholesterolemia[16] are in late preclinical studies or already in clinical trials. Notably, all of these therapeutic candidates resulted from a deep understanding of the underlying genetics of these diseases, often elucidated over decades of work beginning with tedious linkage analysis, to biochemical investigations, to intervenable therapeutic hypotheses. These examples showcase how rare variant genetic studies have served as inspiration and enabler of many aspects of contemporary biomedical research.

Common variants. The genetic background of common variation has been intimately tied to genetics for the last hundred years, originally deriving from seminal studies by the biometricians led by R.A. Fisher[17]. Building upon these foundational studies, understanding of common variant heritability of human disease and traits greatly advanced by Genome Wide Association Studies (GWAS)[18]. Following a framework established by the first comprehensive GWAS manuscript in 2007[19], these studies associate a binary or continuous phenotype or disease with common variants tagged by chip-based microarrays using linear (continuous trait) or logistic regression (binary trait). Building upon this framework, investigators developed 'polygenic risk scores' (PRS) or 'genome wide scores' (GWS) to associate a linear combination of risk or non-risk variants at each position with the variant coefficient to develop a 'risk' of disease[20]. For some disorders, people in very high percentiles of PRS may carry risk equivalent to a rare variant in a related disease-gene[21]. Despite great enthusiasm for their deployment, many logistic issues still exist. One complication of large genetic databases has been the legacy of inequitable sequencing of primarily European ancestry individuals; extrapolating from these data may lead to inappropriate consequences when a PRS is deployed on a non-European ancestry population. Using PRS to select for embryos with perceived 'more desirable' traits is also ethically fraught, as can be

disentangling complex social interactions when using these tools for traits such as 'educational attainment'. Work remains before these data are fully integrated into clinical decision making, and it remains to be seen how well predictive values from PRS compare with traditional clinical covariates.

Rare and Common Variant Interactions. There is great potential to further understand basic genetics by unravelling the mechanistic effects of how rare variants interact with common variants and modulate penetrance (see Chapter 3 for detailed descriptions of penetrance)[22]. While these interactions are being investigated empirically using large biobank approaches, a next step is implicating biological mechanisms. The novel technologies described above may provide the ability to explore the molecular determinants of these observations in suitable model systems. Human genetics has validated therapeutic hypotheses for drug development, and has inspired drug discovery efforts through implicating gene targets in monogenic disease[23]. It is conceivable that an understanding of the biological axis of polygenic risk could also promote therapeutic targets to encourage drug discovery and target prioritization. The main content of this thesis will focus on rare variation, rather than common variation – future directions accounting for common variant interactions are addressed in Chapter 7.

Gene-environment Interactions. Genetics is a fundamentally non-deterministic science. Oftentimes, heterozygotes for the same damaging rare variant may display a continuum of a phenotype or no phenotype at all, a phenomenon known as incomplete penetrance[24]. It is hypothesized that environmental interactions may account for this, in addition to the background of common variation described above. In the context of this thesis, environmental factors such as medications may promote arrhythmias, and exacerbate a liability created by a rare variant (see below).

In summary, the study of human genetics has provided great impetus for progress in basic science and translational medicine. Just as quickly, these investigations have provided numerous additional questions to guide the next century of translational science towards impacting patient care and refining our knowledge of the world[25].

B. Inherited Arrhythmia Syndromes

Overview. Cardiac disease remains the most prevalent cause of morbidity and mortality in the US, sometimes driven by inherited cardiac arrhythmia syndromes[26]. Arrhythmias such as Long QT Syndrome (LQTS) and Brugada Syndrome (BrS) may precipitate sudden cardiac death (SCD) as their first manifestation, while shared cardiomyopathy and arrhythmia syndromes such as arrhythmogenic cardiomyopathy (ACM) may lead to progressive heart failure or induce SCD through fatal ventricular arrythmias. A brief overview of the cardiac conduction system is provided below, followed by clinical descriptions of arrhythmia syndromes studied in this thesis. A more detailed discussion of the role of genetics in the diagnosis and management of these syndromes is provided in Chapter 1.C.

Physiology of the Cardiac Conduction System. The chief role of the heart is to pump oxygenated blood from the lungs to the rest of the body to fuel oxidative metabolism. Human hearts have evolved to link a coordinated propagation of electric signal to the pumping of the heart to maintain synchronized contraction – referred to as excitation-contraction coupling.

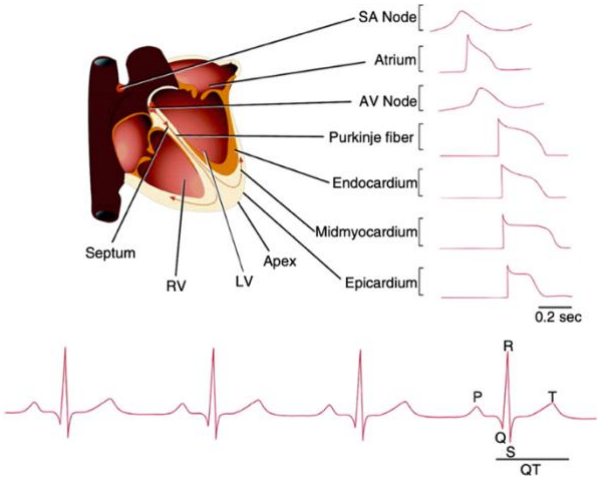


Figure 1.2. Schematic of individual cell traces summing to surface electrocardiogram. Adapted from Zipes, 2014

This electric system is comprised of specialized tissue in the atria and ventricles of the heart that use the flow of ions through channels to propagate an electric signal. The process of EC-coupling

is extremely robust, providing the millions of synchronized signals and contractions that sustain us through our lifetimes.

Cardiomyocytes are the primary cell of the heart. They are electrically excitable cells maintained at a negative resting potential to enable EC-coupling (around -80 mV in ventricular myocardia). Briefly, an electric signal is initiated by cardiac pacemaker cells, typically the sino-atrial node. It is then propagated through a conduction pathway, and then into the cardiomyocytes of the atria and ventricles (Figure 1.2). Spread of current begins with depolarization of the cell by the rapid influx of sodium ions through voltage-gated sodium channels. After activation, sodium channels rapidly inactivate, and a balance of depolarizing calcium currents and repolarizing potassium currents maintains the depolarized state of the cell. At the end of the action potential, repolarizing potassium current dominates and repolarizes the cell back to the negative resting potential. Computational models have been developed to provide *in silico* approximations of how changes in currents or genetic variants may affect this depolarization-repolarization process – the complexity of these models underscores the complex physiology associated with this system (Figure 1.3)[27]. The action potential of all cardiomyocytes sums to provide the characteristic surface electrocardiogram (Figure 1.2.).

Consistent with most physiologic processes, genetic and environmental factors may affect the cardiac conduction system in potentially life-threatening ways. Most simply, if the electric signal is interrupted, the contractile apparatus will halt, and aerobic cell types will eventually die. Common causes of this form of sudden cardiac death (SCD) are sustained ventricular arrhythmias, such as monomorphic ventricular tachycardia or polymorphic ventricular tachycardia or fibrillation. Environmental causes may include re-entrant circuits in the ventricles following a myocardial infarction, while a host of genetic diseases may provide a predisposition through altered function of ion channels, as described in the syndromes immediately below.

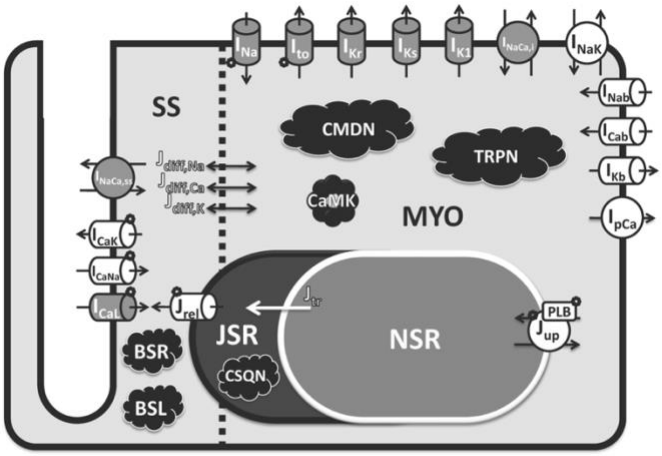


Figure 1.3. Schematic of currents in the O'Hara-Rudy model of ventricular cardiomyocytes. Adopted from O'Hara, et. al, 2011.

Brugada Syndrome. Brugada Syndrome (OMIM: 601144) is an eponymous condition named after a family of physicians that described a conspicuous constellation of ventricular fibrillation, absence of structural heart disease, and unusual ECG changes[28]. A characteristic ECG pattern was identified in the seminal report, specifically "coving" of the ST-segment, ST elevation in the precordial leads V1-V2, and right bundle branch block. This is now referred to as Brugada pattern. Since the first clinical description 30 years ago, our understanding of the disease has significantly improved following detailed laboratory investigations – spanning molecular cloning of a disease locus, patch-clamp electrophysiology of variant channels, characterization of large case cohorts, and common and rare variant analyses of those cohorts. Approximately 20-30% of patients with a monogenic form of Brugada Syndrome have rare, loss-of-function variants in the sodium ion channel gene *SCN5A*[29]. Indeed, this is the only gene definitively associated with Brugada Syndrome as adjudicated by the ClinGen (see ClinGen in section C) Brugada Syndrome gene curation group[30]. Despite these three decades of research, a complete understanding of the disease pathogenesis is far from complete, as a majority of patients have yet to receive a compelling mechanistic explanation of their presentation. Accordingly, clinical management of patients remains rather unrefined. Approaches includes avoidance of certain drugs, implantation of implantable cardioverter-defibrillators (ICDs), pharmacological therapy with quinidine or amiodarone, and increasingly catheter ablation. In Chapters 2, 3, 4, and 6 I study variants involved in BrS.

Long QT Syndrome. Long QT Syndrome (LQTS) is a disease of cardiac repolarization and is typically divided into two forms – congenital LQTS (cLQTS)[31] and acquired or drug-induced LQTS (aLQTS)[32]. The QT interval represents time between the Q and T points on the ECG, which corresponds to the time between a signal initiating ventricular depolarization to the end of ventricular repolarization. Normal QT intervals are 350-450 ms in men, and 360-460 ms in women. Prolongation of cellular action potentials may lead to premature opening of calcium channels, which may then trigger early afterdepolarizations (EADs). EADs are the principle cellular mechanism of LQTS, and may initiate fatal ventricular arrhythmias resulting in SCD through triggered activity[33]. Traditionally, individuals with prolonged QT intervals (>500 ms) without secondary cause are labeled cLQTS (typically adjudicated through the Schwartz score[34]), while those who demonstrate long QT intervals only after a drug exposure were labeled aLQTS. We will primarily consider cLQTS in this thesis, and use this term interchangeably with LQTS below. LQTS is most frequently caused by genetic variants that alter the structure and function of primary cardiac ion channels. The most frequently implicated variants are loss-of-function variants in the repolarizing potassium channels encoded by the genes *KCNQ1* (OMIM: 192500) and *KCNH2* (OMIM: 613688), or phenotypic gain-of-function (molecular loss of inactivation) variants in the depolarizing sodium channel encoded by *SCN5A* (OMIM: 603830)[*35*]. Collectively, these constitute approximately 90% of known genetic variants in patients with phenotypic LQTS, with many other more loosely disease-associated genes comprising the remaining 10%. This thesis will focus on these three main disease genes, as they are the only ones with *definitive* evidence for LQTS according to the expert ClinGen gene curation group[36]. The clinical management of LQTS is better defined than that of BrS[37] . First, QT-prolonging medicines are avoided, followed frequently by initiation of a beta blocker – commonly propranolol or nadolol. For individuals with high-risk symptoms or documented ventricular arrhythmias, an ICD may be implanted or left cardiac sympathetic denervation may be performed. In Chapters 3 and 4 I study variants involved in LQTS.

Arrhythmogenic Cardiomyopathy. Traditionally, primary arrhythmia syndromes and cardiomyopathy syndromes have been treated as two distinct entities in clinical practice. Cardiomyopathy syndromes are characterized by decreased output of the heart, typically due to decreased contractility or increased stiffness and inability to relax. Challenging this long-held distinction, recent research has shown a substantial overlap of cardiomyopathy and arrhythmias, spanning phenomena like tachycardia-induced cardiomyopathy. A syndrome termed arrhythmogenic cardiomyopathy (ACM) has been coined to describe these phenotypes[38]. As evident in the name, this phenotype comprises both ventricular dysfunction and arrhythmias. Rare variants in desmosomal and sarcomeric genes like *PKP2, DSP,* and *FLNC* are associated with forms of this condition. Truncating variants in *FLNC* are highly penetrant causes of arrhythmogenic cardiomyopathy involving the left and/or right ventricles and will be the focus of work described in Chapter 5[39]. Management of arrhythmogenic cardiomyopathy includes guideline directed medical therapy for heart failure caused by the cardiomyopathy. Suppression of arrhythmias may be undertaken with certain antiarrhythmics, and ICDs are implanted consistent with guidelines for other cardiomyopathies. In Chapter 5 I study a variant identified in arrhythmogenic cardiomyopathy.

Additional Inherited Arrhythmias. There are multiple additional inherited arrhythmia syndromes that were not treated explicitly within this thesis. These include the Short QT Syndrome (SQTS)[40], Catecholaminergic Polymorphic Ventricular Tachycardia (CPVT)[41], the newly recognized Calcium Released Deficiency Syndrome (CRDS)[42], and additional syndromic diseases such as Timothy Syndrome, Naxos disease, Andersen-Tawil Syndrome, and multiple types of muscular dystrophy. The framework I use in my studies are equally applicable to many of these conditions.

C. Clinical interpretation of genetic variations

ACMG Framework of Variant pathogenicity. The clinical consequence of genetic variants is most frequently adjudicated through the American College of Medical Genetics and Genomics (ACMG) classification system[43]. This framework weighs variant functional data, co-segregation with phenotype, *in silico* predictions, population frequency, and other factors to interpret risk associated with the variant (Figure 1.4). After integrating different criteria, the variant is assigned a classification of Benign (B), Likely Benign (LB), Variant of Uncertain Significance (VUS), Likely Pathogenic (LP), or Pathogenic (P) depending on the direction and strength of the cumulative evidence. Unfortunately, most variants fall into the nebulous category of VUS, lacking sufficient data to strongly conclude involvement (or lack of involvement) in disease[44]. This will increasingly

	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very Strong
Population	High MAF (BA1/BS1)			Absent in Population Database (PM2)	Increased prevelance in disease (PS4)	
Computational		Multiple benign computational evidence (BP4) Synonymous with no splice impact prediction (BP7)	Multiple lines of computational evidence with predicted impapct (PP3)	Variant affects protein length (PM4) Novel missense at previously P variant position (PM5)	Same AA change as established P variant (PS1)	Predicted null variant in gene with LoF mech. (PVS1)
Functional	Functional assay shows no deleterious effect (BS3)		Missense in gene w/ low rate of B missense variants (PP2)	Mutational hot spot or domain w/o benign variation (PM1)	Deleterious effect in functional studies (PS3)	
Segregation	No disease segregation (BS4)		Cosegregation in multiple families (PP1)	Increasing strength → of segregation data		
De novo				De novo - paternity maternity not confirmed (PM6)	De novo - paternity maternity confirmed (PS2)	
Other		Found in case with other cause (BP5) Reputable source w/o shared data (BP6)	Reputable source states P (PP5) Phenotype for FH highly gene specific (PP4)			

Figure 1.4. ACMG Criteria for adjudicating variant effect. Adapted from Richards et. al, 2015

be the case for new variants discovered by more comprehensive sequencing, as many new variants will be discovered that lack functional or clinical data. This distribution hinders the clinical actionability of these data, as VUS are not intended to influence clinical management, despite currently being the largest category of annotated genetic variation in the variant repository ClinVar[45].

Refinements of ACMG Criteria. The Clinical Genomic Resource[46] (ClinGen) is an interdisciplinary consortium that has contributed to pressing problems in genetic medicine - variant pathogenicity, clinical actionability of genetic findings, and gene-disease validity (see mentions for BrS and LQTS above). Towards these goals, ClinGen working groups have helped refine how the ACMG criteria are implemented for different diseases and different categories of data. Highly relevant to my work, the ClinGen Sequence Variant Interpretation working group delineated a new framework for considering functional evidence in 2020[47]. The group suggested that the weight of functional evidence should be proportional to the performance of the assay on known B/LB and P/LP controls. Often, assays are implemented for VUS without first calibrating on known controls to assess the positive and negative predictive value of the assay. This working group suggestion inspired work on assay calibration in Chapter 6, and a current follow up to Chapter 2. Another pertinent set of recommendations came from the Sequence Variant Interpretation group for non-coding variation. This provided a set of specific recommendations for putative intronic splice-altering variants which were implemented in Chapters 4-6. Finally, the field is moving towards a more quantitative implementation of the ACMG criteria using a points-based system within a Bayesian framework[48,49]. This approach reflects odds of pathogenicity in a naturally scaled manner and enable more intuitive understanding of evidence. I propose an alternative framework for variant interpretation in terms of disease penetrance estimates and uncertainties in Chapter 3 that complements this framework.

Genotype-guided Management of Inherited Arrhythmia Syndromes. For many of the diseases discussed above, there are effective therapies such as beta-blockers, antiarrhythmics, implanted cardioverter defibrillators (ICD), cardiac sympathetic denervation, and cardiac ablation. The utility of these interventions in attenuating disease sequalae is balanced against their potential risks, including additional arrhythmias, severe bleeds, pericarditis, accidental shocks, and even iatrogenic death. Genetics therefore offers an opportunity to refine the clinical diagnosis and management of patients suspected to suffer from these syndromes, and subsequently guide medical intervention. Genotype-first approaches may begin to refine diagnosis and associated risk, towards maximizing patient comfort *and* outcomes, perhaps even pre-empting the development of manifest disease[50]. This vision is predicated upon an intimate understanding of the clinical consequence of variants – one of the greatest challenges in clinical genetic medicine.

One recent example of using human genetic data to identify individuals at risk for inherited arrhythmia syndromes came from the eMERGE-III study (Electronic Medical Records and Genomics Phase III)[51]. The consortium obtained sequencing of 10 arrhythmia-related genes in over 21,000 participants and identified participants harboring P/LP variants. By combining electronic health record review, in-person follow up and return of results, and functional investigations, the authors were able to establish new diagnoses in a non-negligible fraction of participants. This work provided an example that could be scaled up in other biobanks; however, policy decisions and ethical frameworks will have to be further developed to allow consistent return of results and sharing of individual-level data between research groups. Further efforts could also prioritize additional VUS for reclassification based on higher-throughput approaches described below.

D. Functional Investigations of Variant Effect

The ACMG framework described above incorporates numerous classes of data to adjudicate variant pathogenicity. Functional data are heavily weighted in this framework and significantly influences the outcome of variant classification when available; however, many assays can be technically challenging and time-intensive to implement. A major goal of this thesis was to establish methods for functionally profiling missense and putative splice-altering variants, and integrating the results into ACMG schema. I will provide an overview of functional investigations of missense variants in cardiac ion channel genes, mainly patch-clamp studies, and of methods for studying splice-altering variants below. These included using automated patch-clamp (APC), minigene assays, and transcriptional studies of CRISPR-Cas9 edited and non-edited iPSC-CMs.

Missense Variation.

Functional assays such as patch-clamp electrophysiology have allowed researchers to directly test the effects of genetic variants on ion channel properties in different cell and tissue types[52]. Patch-clamp involves measuring the current through ion channels. Voltage-gated channels can be tested in voltage clamp protocols which probe channel responses to changes in transmembrane voltage. These methods enable investigators to study ion channel function in detail, and to study specific ion currents in isolation from other currents. Experiments may involve *ex vivo* disease models such as mice, dog, or guinea pig preparations, induced pluripotent stem cell cardiomyocytes (iPSC-CMs), or heterologous expression of channels in robust cell models such as HEK cells or CHO cells. More recently, the advent of automated patch-clamp has allowed the interrogation of variant channel properties at unprecedented scale[53]. Within the arrhythmia space, investigators have focused significant attention on studying the surface membrane proteins encoded by *SCN5A, KCNQ1,* and *KCNH2* due to their association with arrhythmia syndromes[54,55].

Patch-clamp Electrophysiology Covariates. Patch-clamp studies provide information on several different aspects of channel function. As voltage-gated ion channels switch, or gate, through different conformations – closed, open, inactive – the kinetics of transitions and amplitude of current can be derived from mathematical modeling of raw data. I will describe some of the key electrophysiologic properties that were used in Chapters 2 and Chapter 3 below – these general properties apply to SCN5A, KCNQ1, and KCNH2. Clinical correlates are described for KCNQ1 channels.

Peak Current. This property describes the maximum amount of current conducted through the channel at a given membrane voltage. Peak current is the most commonly investigated functional property of these channels. In the case of the cardiac potassium channels KCNQ1 and KCNH2, decreased peak repolarizing current prolongs the action potential duration, which at a tissue-level, extends the QT-interval and susceptibility to fatal ventricular arrhythmias.

Voltage of Half Activation and Inactivation. This parameter describes the voltage at which half of the channels are activated and able to conduct current, or inactivated and unable to conduct current. Genetic variants may affect these parameters by affecting the protein conformation at different membrane voltages, and thereby altering gating. For KCNQ1, increased voltage of half activation leads to fewer channels being activated, and therefore decreases net repolarizing current resulting in action potential and QT prolongation.

Tau Activation and Inactivation. Channel activation and inactivation kinetics can be modeled using the time constant τ. This parameter describes the rate of rise or fall of membrane voltage. The larger a time constant, the slower the activation/inactivation of channel current in response to changes in membrane voltage. KCNQ1 variants with an increased time constant of activation would therefore have a slower rise of current, meaning less repolarization, and longer action

potentials. Variants that decrease the potassium channel inactivation time constant would reflect more rapid channel inactivation, decreasing the total repolarizing currents available and therefore extending the action potential.

Additional missense functional assays in arrhythmia genes have been explored that capture channel trafficking to the cell membrane[56], often using flow cytometry, or general functional effect within a selection assay[55]. An advantage of these methods is that they allow higher throughput interrogation than more time-intensive manual patch-clamp studies. Certain examples implementing these assays at scale are described in Chapter 7 as part of future directions.

Splice-altering Variation.

General mechanism. Splicing takes place in the cytoplasm of eukaryotic cells through a complex assembly of RNA and proteins known as the spliceosome. This machinery consists of highly evolutionarily conserved RNA binding proteins (RBPs), effector proteins, small nuclear RNAs (snRNA), and small nuclear ribonucleoproteins (snRNP) that co-transcriptionally recognize exons and catalyze the splicing out of intronic sequences during RNA processing. The major spliceosome snRNAs recognize the exon border, typically flanked by AG-GT nucleotides, polypyrimidine tracts, and branch points to orchestrate an evolving complex to catalyze the transesterification reaction (Figure 1.5A). Beyond the largely conserved 2-nucleotide canonical splice sites, there are other short *cis*-acting sequences that modulate splicing outcomes in concert with *trans*-acting factors. These include Exonic Splice Enhancers (ESEs) and Intronic Splice Enhancers (ISEs) which recruit SR-rich proteins to promote exon recognition, and Exonic Splice Silencers (ESSs) and Intronic Splice Silencers (ISSs) which interact with hnRNPs to sterically repress splicing (Figure 1.5B). These *cis*- and *trans*-elements coordinate, but may conversely perturb each other in disease states.

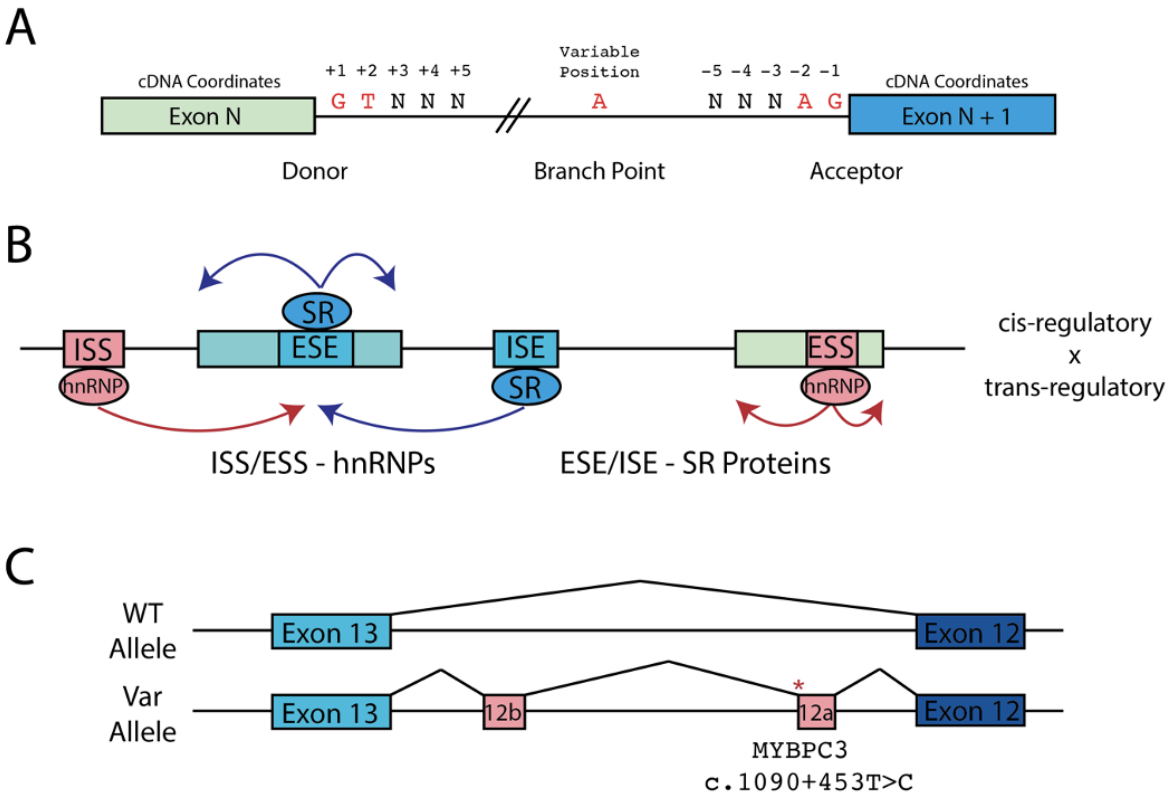


Figure 1.5. Nomenclature and regulation of splicing.
A) Canonical positions and elements of a typical exon splicing reaction. The intronic position is assigned based on closest cDNA boundaries.
B) Regulation of splicing by *cis*- and *trans*-regulatory elements. ISS and ESS interact with heterogenous nuclear ribonuclear proteins to suppress exon recognition. ISE and ESE interact with serine, arginine-rich proteins to recruit splicing factors to recognize an exon and promote productive splicing. These sequences may be disrupted by point mutations, leading to aberrant splicing.
C) Example of aberrant splicing arising from a 'deep' intronic variant, 453 nucleotides from the splice junction in *MYBPC3*, a common hypertrophic cardiomyopathy disease gene. The aberrant transcript results from the activation of multiple cryptic splice sites, creating 2 pseudoexons in the processed read.

Variant-Induced Aberrant Splicing. Germline variants may be an inherited or *de novo* substrate for Mendelian disease. Variation within and outside the 2-bp canonical splice site may disrupt splicing by ablating the canonical site, introducing competing splice acceptor/donor motifs, decreasing the ability of spliceosome components to recognize the canonical splice sites, or by disrupting regulatory elements. These molecular aberrations may introduce frameshifts or premature termination codons (PTCs) into the mature mRNA, which then may disrupt translated protein or cause nonsense-mediated decay (NMD) of the incorrectly spliced transcript. An

example in the cardiomyopathy-related gene *MYBPC3* is shown in Figure 1.5C. In this case, a 'deep intronic' variant, defined as >100 nucleotides from the exon/intron border, activates cryptic splice sites and creates 2 intron retention events.

Tissue-specific Regulation of Splicing. One difficulty of studying splice-altering variation is access to disease relevant primary tissue, due to differences in transcript composition across cell types[57]. An alternative to primary tissue acuiqisition utilizes the process of illegitimate transcription to study low abundance 'tissue-specific' transcripts in nearly any cell type[58]; however, these analyses may be confounded by cell-type specific nonsense-mediated decay (NMD) or other processes such as relative expression of *trans*-acting RBPs. In Chapter 5 we show than an aberrantly spliced transcript was only unambiguously identifiable after treating disease-relevant pseudo-primary tissue with the small molecule NMD-inhibitor cycloheximide. Furthermore, a native cell may be more resilient to splice-perturbing variation discovered in an *in vitro* system due to alternative splicing events in the primary transcript. For example, variants effect in the large gene *TTN* is often ambiguous due to additional gene regulation and widespread alternative splicing of *TTN* in the adult cardiomyocyte.

Functional Investigations of Splice-altering Variation

Compared to protein-coding variation, potential splice-altering variants have been less frequently investigated, owing to a historical paucity of strong *in silico* tools and time-intensive functional assays. I outline below different approaches for studying splicing from clinical tissues and *in vitro* models.

Primary Human Tissue. The most direct approach to studying aberrant splicing is to investigate the mature mRNA composition from affected individuals in primary tissue. In cardiac disease, primary tissue analysis requires invasive and potentially dangerous biopsies for prospective

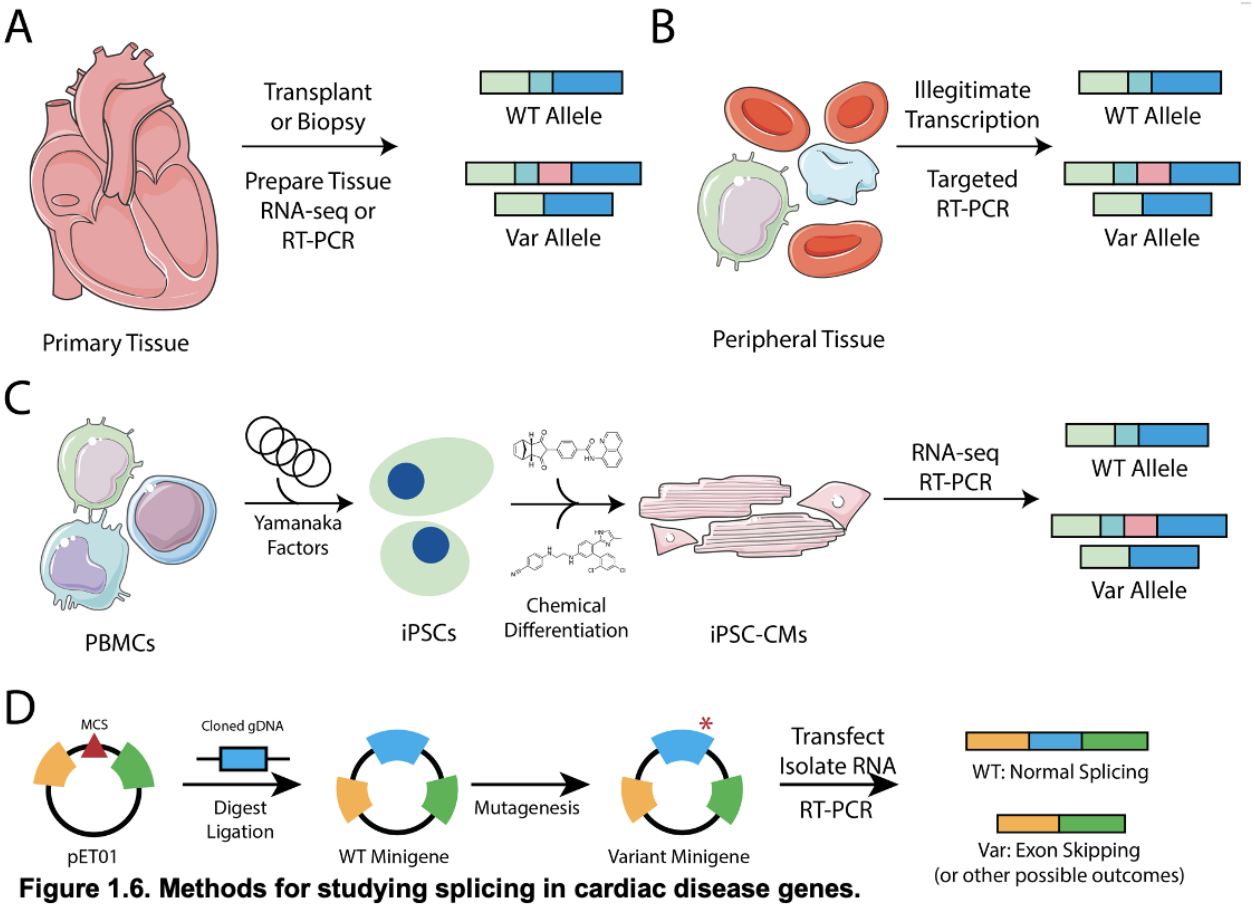


Figure 1.6. Methods for studying splicing in cardiac disease genes.
A) Primary tissue can be obtained for the highest fidelity readout following transplant or biopsy. RT-PCR (reverse transcription, polymerase chain reaction).
B) Peripheral tissue may serve as a surrogate for primary tissue through the phenomenon of illegitimate transcription.
C) Pseudo-primary tissue can be obtained following reprogramming of terminally differentiated cells into induced pluripotent stem cells (iPSCs), followed by chemical differentiation into iPSC-CMs (iPSC-cardiomyocytes). CRISPR-editing may create isogenic controls at the iPSC stage.
D) The *in vitro* minigene model may allow the cloning of relevant DNA and assess transcriptional consequences after transfection into different cell types of interest.

diagnostic value (Figure 1.6A). Successful examples of such analyses include those studying variants linked to hypertrophic cardiomyopathy, as extracting relevant tissue can be both diagnostic and therapeutic. As an example, Horie and colleagues presented elegant RNA studies from tissue obtained during open heart surgery of a patient with a splice-altering exonic *LMNA* variant[59].

Peripheral Tissue. Surrogate tissue has been used in splicing assays by taking advantage of illegitimate transcription. Easily accessible peripheral blood can be used to study splicing by analyzing low levels of gene expression in peripheral blood mononuclear cells (Figure 1.6B). This provides a reasonable approximation of splicing patterns in a non-disease relevant, yet clinically accessible tissue. For example, illegitimate transcription allowed Kremer et. al. to identify aberrant splicing events in 10% of patients with a clinical diagnosis of mitochondriopathy without an identified genetic cause by using RNA-seq of patient-derived fibroblasts[60]. Confounding may arise as discussed above, with variable splicing outcomes obtained in different systems. For example, inferring splicing events from clinically accessible tissues may nonetheless lead to discrepancies when inferring splicing in non-clinically accessible tissues[61]. A team led by Bohj quantified these differences by investigating splicing patterns among different tissues in the Genotype-Tissue Expression (GTEx) project[57], cautioning against assumptions of identical splicing patterns.

Pseudo-primary Tissue. Patient-derived induced pluripotent stem cell-derived cardiomyocytes (iPSC-CMs) have transformed the ability to investigate aberrant splicing in cardiac disease (Figure 1.6C). These model systems enable molecular investigations in *in vitro* derived cells approximating native cells. An advantage lies in the *trans*-acting milieu of proteins and whole-cell phenotypes that may be captured and probed using 'pseudo' primary tissue. Moreover, patient-derived models also capture the polygenic background that may influence large effect rare variant consequences. With the advent of CRISPR-Cas9 genome editing, isogenic controls can be created, either starting from healthy controls or patient-derived cells[62]. Critically, these have allowed me and many other investigators to follow up a variant-induced transcriptional consequence (out-of-frame aberrant splicing), with cellular-level consequence such as altered electrophysiology or contractility (see Chapters 4 and 6).

Minigene Approaches. Minigene assays have been used for more than two decades to investigate the effect of genetic variation on splicing outcomes. These assays rely on overexpression of a plasmid bearing known exons and introns intervened by a relevant stretch of genomic sequence in a variant- to wild-type-plasmid comparison (Figure 1.6D). Details of the assay implementation are described in Chapters 4 and 6. A major strength of this approach is that they are disease agnostic, allowing a quick assessment of variant effects with a consistent methodology. Indeed, nearly any genomic region of interest can be subcloned into commercially available minigene vectors for studies. Conversely, this intrinsic flexibility creates a limitation, as the assays capture transcript-level consequence but not downstream consequences in disease relevant tissue – i.e., disease relevance must be assumed from mRNA composition. Indeed, slight differences have been observed between readouts from minigene studies and those from direct analysis of primary or pseudo-primary tissue[63].

Complementary In silico *Tools*. Predicting variant-induced splicing alterations has historically been difficult – multiple molecular motifs often overlay each other, few experimental observations were available for calibration, and limited training data were available. The current availability of large sets of RNA-seq data and improved computational approaches has catalyzed the development of increasingly accurate *in silico* predictors of variant-induced splicing changes. Recently developed tools such as SpliceAI[64] have greatly improved upon these previous tools, leveraging findings from vast amounts of training data in combination with sophisticated neural networks[65].

E. Chapter Descriptions

Chapter 2: Dominant negative effects of SCN5A variants. The principal cardiac voltage-gated sodium channel, $Na_V1.5$, is a protein encoded by *SCN5A* that has largely been thought of as a monomer. Recent small-scale studies have shown compelling evidence that $Na_V1.5$ proteins nonetheless interact at the cell surface, and that certain variants may exhibit dominant negative effects. Leveraging high-throughput automated patch-clamping, I show that many loss-of-function variants in $Na_V1.5$ indeed display dominant negative behavior when studied in heterozygous configurations. We show the extent of peak current abrogation, the primary molecular correlate of disease, correlates with more severe penetrance of Brugada Syndrome in a large case-control comparisons.

Chapter 3: Bayesian interpretations of variant effect. Rare variants frequently demonstrate incomplete penetrance. Some variant heterozygotes display fulminant disease, some attenuated disease, and some no phenotype at all. To model this phenomenon, we elaborated a Bayesian framework for understanding variant effect, developing a prior that is comprised by variant specific features, and a posterior from clinical phenotyping. Combining these, we generate a point estimate of penetrance and its associated uncertainty for each single nucleotide variant in the *KCNQ1*-LQT1 and *SCN5A*-LQT3 genotype-phenotype pairs. We show that this is a well-calibrated approach to studying variant effect, and that it may provide helpful insights into underlying protein biology.

Chapter 4: Functional assays reclassify suspected splice-altering variants of uncertain significance in mendelian channelopathies. Putative splice-altering variants are difficult to interpret by conventional methods. We employed two functional assays to interpret the transcriptional effect of 12 VUS with a high likelihood of driving Brugada Syndrome or Long QT

Syndrome through aberrant splicing. Using minigene studies or CRISPR-Cas9 editing of iPSC-CMs, we were able to reclassify 9 VUS - 8 to LP and 1 to LB.

Chapter 5: Multicenter clinical and functional evidence reclassifies a recurrent non-canonical Filamin C splice-altering variant. Refining the clinical interpretations of VUS often requires multidisciplinary approaches, best undertaken by a team of investigators. Here, we lead an investigation with international collaborators to use clinical phenotyping, RNA studies, and electrophysiology of patient-derived iPSC-CMs to reclassify a recurrent *FLNC* intronic variant associated with highly penetrant arrhythmogenic cardiomyopathy.

Chapter 6: Development of a platform methodology to adjudicate putative splice-altering variation. Despite an estimated 10% of genetic variants precipitating disease through non-canonical aberrant splicing, most methods to investigate disease-relevant splicing variants are low-throughput and labor-intensive (see Chapters 4 and 5). Here, we leveraged long- and short-read next generation sequencing, rapid oligonucleotide synthesis, and bespoke computational pipelines to develop Parallel Splice Effect sequencing (ParSE-seq), a method to simultaneously measure the effect of hundreds of putative splice-altering variants. We validate the method by studying select variants in cDNA-based automated patch-clamping and in CRISPR-Cas9 edited iPSC-CMs.

Chapter 7: Conclusions, Limitations, and Future Directions. I provide a synthesis of the work described in this thesis, discussing the accomplishments and insights from the research, its limitations, and next steps to continue investigations.

Chapter 2

Dominant Negative Effects of *SCN5A* Missense Variants

Portions of this chapter are published under the same title in *Genetics in Medicine*

Purpose: Up to 30% of patients with Brugada Syndrome (BrS) carry loss-of-function (LoF) variants in the cardiac sodium channel gene *SCN5A*. Recent studies suggested that the *SCN5A* protein product $Na_V1.5$ can dimerize and exert dominant negative effects. Here we sought to explore the generality of $Na_V1.5$ dominant negative effects and their clinical severity.

Methods: We identified 35 LoF variants (<10% of wildtype [WT] peak current) and 15 partial LoF variants (10-50% of WT peak current) that we assessed for dominant negative effects. *SCN5A* variants were studied in HEK293T cells alone or in heterozygous co-expression with WT *SCN5A* using automated patch clamp. To assess clinical risk, we compared the prevalence of dominant negative vs. putative haploinsufficient (frameshift/splice/nonsense) variants in a BrS consortium and the gnomAD population database.

Results: In heterozygous expression with WT, 32/35 LoF and 6/15 partial LoF variants showed reduction to <75% of WT-alone peak current, demonstrating a dominant negative effect. Individuals with dominant negative LoF variants had an elevated disease burden compared to individuals with putative haploinsufficient variants (2.7-fold enrichment in BrS cases, p=0.019).

Conclusions: Most *SCN5A* missense LoF variants exert a dominant negative effect. This class of variant confers an especially high burden of BrS.

Introduction

Brugada Syndrome (BrS) is a clinical arrhythmia syndrome with characteristic EKG changes in the absence of underlying structural heart abnormalities[66]. While often asymptomatic or clinically unrecognized, sudden cardiac death (SCD) due to ventricular tachyarrhythmia can be the sentinel manifestation. Up to 30% of BrS patients have heterozygous loss-of-function (LoF) variants in the cardiac sodium channel gene *SCN5A*, which encodes the channel protein $Na_V1.5$[29]. A recent evaluation by ClinGen asserted that *SCN5A* was the only gene with strong evidence for Mendelian associations with BrS[30]. LoF *SCN5A* variants are also associated with other arrhythmias including sick sinus syndrome[67] and progressive cardiac conduction disease[68].

Hundreds of LoF variants within *SCN5A* have been reported in the literature and in ClinVar across multiple variant classes including missense, nonsense, splice-altering, and frameshift/premature truncation[29,45,69,70]. *SCN5A* encodes a channel with 4 transmembrane domains, each consisting of 6 transmembrane segments[71]. $Na_V1.5$ has traditionally been thought to function as a monomer; however, a recent study indicated that $Na_V1.5$ can form dimers with coupled intracellular trafficking and/or gating at the plasma membrane[72]. Similar to variants in established multimeric proteins that can generate dominant negative effects, several missense *SCN5A* variants with dominant negative effects on trafficking or coupled gating at the cell surface have been reported *in vitro* and *in vivo*[73-75]. However, the dominant negative behavior of most of the approximately 40 functionally-validated LoF missense variants in *SCN5A* has not been tested[76]. Moreover, the degree of dominant negative effects among partial LoF missense variants has not been evaluated.

Variable penetrance is a hallmark of pathogenic BrS variants, and the extent to which distinct pathogenic mechanisms (e.g., dominant negative vs haploinsufficiency) contribute to this effect is unknown. Large cohort studies and variant curation efforts provide datasets of *SCN5A* variants associated with BrS cases[29,69,70]. In addition, large population cohorts such as gnomAD

provide sets of individuals more likely representing putative controls[77]. Together, these datasets enable the comparison of BrS disease risk among different variant classes.

Here, we study the prevalence of the dominant negative effect among *SCN5A* LoF and partial LoF missense variants. We use case and control cohorts to test the relative BrS disease risk of dominant negative missense variants compared with other variant classes.

Methods

Selection of Variants. We selected variants from previously published functionally characterized variants[69,76]. Variants with peak currents <10% compared to WT were considered LoF, and variants with peak currents between 10-50% compared to WT were considered partial LoF.

SCN5A Mutagenesis. The *SCN5A* variant plasmids were mutagenized using a previously described "zone" system[76]. Briefly, *SCN5A* individual zones on small plasmids were mutagenized using the QuikChange Lightning Multi kit (Agilent) with primers designed using the online QuikChange Primer Design tool. Primers used in this study are listed in the Appendix. The variant-containing zone was then subcloned by restriction digestion into a plasmid containing an AttB:SCN5A:IRES:mCherry-blasticidinR plasmid[76,78,79]. The entire sequence of the zone containing the variant was confirmed by Sanger sequencing. In a previous study of 82 variants generated by this approach, 0/82 plasmids had any additional *SCN5A* variants outside the target zone[76]. All analyses used the most common *SCN5A* transcript in the adult heart, including the adult isoform of exon 6 and a deletion of the alternatively spliced Gln1077 residue (ENST00000443581). As per convention, all variants are named in accordance with the full 2,016 amino acid form (ENST00000333535).

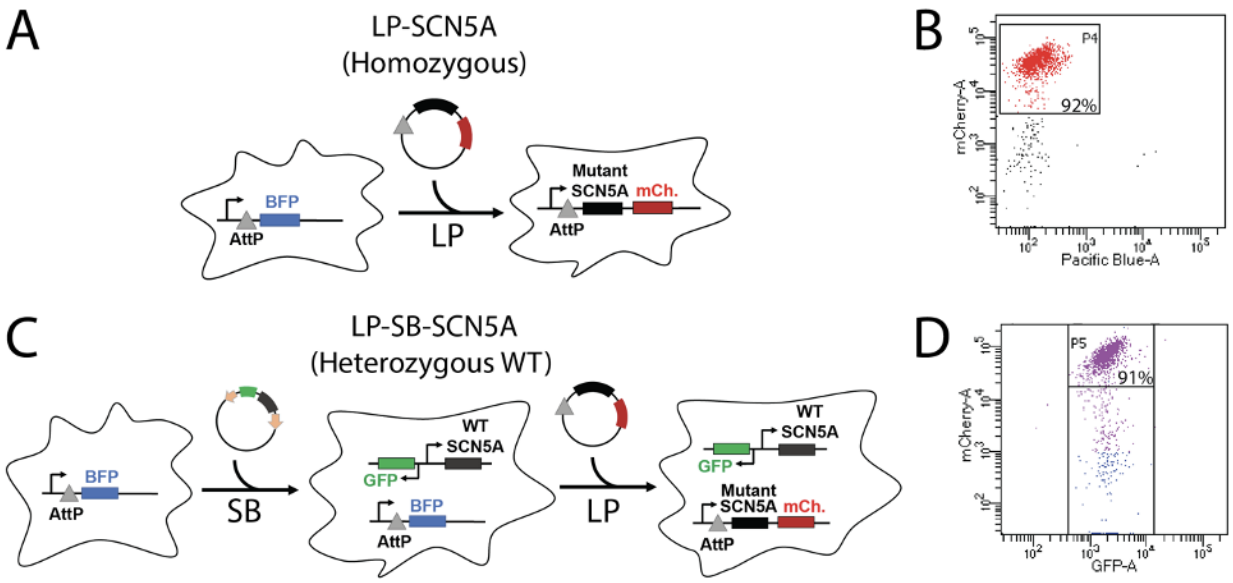


Figure 2.1. Stable cell lines used to study dominant negative *SCN5A* variants. 1 or 2 copies of *SCN5A* were inserted into engineered HEK293 LP cells. The Landing Pad (LP) comprises an AttP and BFP locus. and allows insertion of a single insert per cell. A second Sleeping Beauty (SB) transposon system was used to introduce a second copy of the gene for heterozygous experiments.
A) Design of homozygous LP-SCN5A cell line with LP integration.
B) Analytical flow cytometry after incorporation of plasmid into the LP. Cells that do not have BFP expression and highly express mCherry (P4 gate) have a successful integration.
C) For heterozygous experiments, we used a combination of LP and SB systems. First, a SB plasmid bearing a WT copy of *SCN5A* was randomly inserted into the genome. A clone of these cells was identified that has an equal level of $Na_V1.5$ in patch clamp experiments to typical LP expression (Figure 2). Next, a second copy of *SCN5A* bearing WT or variant was incorporated through the LP system.
D) Results of flow cytometry after SP and LP integration. Cells express GFP associated with SB integration, and mCherry after LP integration (P5 gate).

Description of Cell Lines: All experiments used Human Embryonic Kidney HEK293T "negative selection" landing pad (LP) cells as previously described (gift of Kenneth Matreyek)[76,78,79]. The AttB/AttP LP allows a single integration event per cell and a consistent level of target gene expression (Figure 2.1). Homozygous experiments were carried out in LP cells (Figure 2.1A). Plasmids carrying *SCN5A* variants were transfected along with transposase and integrated into the LP site to allow stable expression. Transfection efficiency was monitored by the percentage of cells expressing mCherry (Figure 2.1B). We termed these lines LP-SCN5A.

For heterozygous expression, we first generated LP cells stably expressing WT *SCN5A* in a non-LP site using the Sleeping Beauty (SB) transposon system and identified a clone of this cell line

with peak sodium current (I_{Na}) equivalent to that observed with WT *SCN5A* in the LP site (Figure 2.1C). We then generated cell lines with *SCN5A* variants transfected into the LP site, thereby allowing us to express the two SCN5A alleles at equivalent levels and assess dominant negative effects. We quantified the expression using mCherry from the LP locus, and GFP from the SP integration (Figure 2.1D). These lines are referred to as LP-SB-SCN5A.

Generation of Variant Cell Lines: Cells were cultured at 37°C in humidified 95% air/5% CO_2 incubator in "HEK media": Dulbecco's Eagle's medium supplemented with 10% fetal bovine serum, 1% non-essential amino acids, and 1% penicillin/streptomycin. Stable integration of a WT *SCN5A* into LP-cells was achieved using an optimized SB transposon system[80] using the pSBbi-GN plasmid (a gift from Eric Kowarz, Addgene #60517), which contains SB transposon sequences for genomic integration flanking a promoter upstream of GFP and a second promoter upstream of a multiple cloning site (MCS) for expression of a gene of interest. A NotI restriction site was first cloned into the multiple cloning site using Gibson assembly (New England Biolabs). Then, WT *SCN5A* was cloned into the MCS by NotI digestion (New England Biolabs). Next, 1 ug of pSBbiGN-SCN5A and 100 ng of pCMV(CAT)T7-SB100, a plasmid expressing SB transposase (a gift from Zsuzsanna Izsvak, Addgene #34879), were cotransfected into the cells[81], using FuGENE 6 (Promega) following manufacturer's instructions. At day 7 post-transfection, GFP+ cells were sorted by fluorescence-activated cell sorting (FACS), and individual colonies were picked and re-analyzed by analytical flow cytometry to identify clones expressing varying levels of GFP (and thus varying levels of $Na_V1.5$). Clones were then tested by SyncroPatch automated patch clamping (see below) to identify a clone expressing an equal peak sodium current as results from typical integration of a single copy of wild-type Nav1.5 into the AttB/AttP landing pad.

For homozygous patch clamp experiments, LP cells were transfected with an AttB-SCN5A variant:IRES:mCherry-BlasticidinR plasmid and studied as previously described[76]. For heterozygous patch clamp experiments, LP-SB-SCN5A cells were transfected using similar

methods. For all cell lines, cells were transfected with FuGENE 6 or Lipofectamine 2000 following manufacturer's suggested protocols using an AttB-containing SCN5A:IRES:mCherry:blasiticidinR plasmid and a plasmid bearing Bxb1 recombinase; cells underwent negative selection for 6 days with 1 ug/mL doxycycline (to induce promoter expression; Sigma), 100 ug/mL blasticidin S (to kill cells not expressing the blasticidin-resistant plasmid; Sigma), and 10 nM AP1903 (to kill un-integrated cells expressing the AP1903-sensitive caspase gene; MedChemExpress) in HEK media. At the end of selection, cells were assessed by analytical flow cytometry to assess percentage of mCherry-positive, BFP-negative cells (LP integration of *SCN5A* variant) and GFP-positive cells (SB integration of *SCN5A*).

Automated Patch Clamping. Electrophysiology data were collected with the SyncroPatch 384PE automated patch clamping device (Nanion) using the same cell preparation and solutions as previously reported[76]. Peak currents are reported at -20 mV after a 200 msec pulse from a resting potential of -120 mV; peak sodium current is presented as the mean of data obtained in ≥8 cells/variant (homozygous experiments) or ≥27 cells/variant (heterozygous experiments). Voltage of half activation, voltage of half inactivation, time of 50% recovery from inactivation, and late current at 200 ms were obtained using previously published protocols[76]. As previously described, cells with values greater than 2.5 standard deviations from the mean were removed in an automated process[76]. For these additional parameters, only variants with data collected from >10 cells were included.

Case-control analysis. We performed case-control analyses to test the penetrance of different classes of variants. We used BrS case counts from a recent International BrS Genetics Consortium and putative controls from gnomAD[70,77]. We use gnomAD as putative controls; although phenotypes are not available for gnomAD participants, the vast majority of these individuals should not have Brugada Syndrome. All gnomAD counts were taken from gnomAD

v2.1.1 transcript ENST00000333535.4. A cut-off minor allele frequency of 2.5e-5 was used to designate ultra-rare variants, as previously suggested[82]. To test the severity of each disease class (i.e., missense vs. indel vs. splice/frameshift/nonsense), we compared the relative number of cases versus controls by variant, drawing from the BrS consortium and gnomAD. Frameshift, splice, and nonsense variants at amino acid position > 1800 (post-transmembrane domain IV) were excluded due to the possibility that these variants may not be full LoF. We calculated the odds ratio associated with each variant class according to the formula (a/b)/(c/d), where a = BrS cases with variant, b = BrS cases without variant, c = gnomAD controls with variant, and d = gnomAD controls without variant. Since the allele number varied for different variants in gnomAD, the average allele number was calculated over all relevant variant types (missense, frameshift, nonsense, and splice site) and divided by 2 to obtain a count of sequenced gnomAD participants to use in odds ratio calculations, following a previously published approach[70].

Data Analysis. SyncroPatch 384PE data were analyzed as previously reported[76]. Peak current densities were calculated by dividing peak current at -20 mV by cell capacitance. For homozygous experiments, peak current densities were normalized to peak current densities observed in cells expressing WT plasmid. For heterozygous experiments, peak current densities were normalized to that observed in LP-SB-SCN5A cells, *i.e.,* those expressing a single WT allele. As described below, WT+WT cells displayed ~200% peak I_{Na} compared to LP-SB-SCN5A cells. Heterozygote (WT+variant) cells displaying <75% of peak I_{Na} compared to LP-SB-SCN5A cells were designated as exerting a dominant negative effect. Statistical comparisons were made using two-tailed Fisher's exact tests, implemented in R Studio (version 1.3.1093). R code used for all analyses is available upon request.

Results

Homozygous and Heterozygous Measurements of LoF Variants

We generated 37 LP-SCN5A stable lines (1 *SCN5A* allele expressed/line), each expressing LoF variants or the nonsense variant p.Trp822Ter[78,79]. Representative traces for WT and p.Ala735Glu are shown in Figure 2.2A. We recorded peak I_{Na} at -20 mV: 35/37 missense variants exhibited a peak current density <10% compared to WT (Figure 2.3A – only LoF shown). The remaining 2 variants (previously reported to be LoF) showed >10% peak current when compared to WT and were studied separately with other partial LoF variants. One LoF variant, p.Arg893Cys, was previously detected in patients with BrS but has not been previously assessed by patch clamping[29].

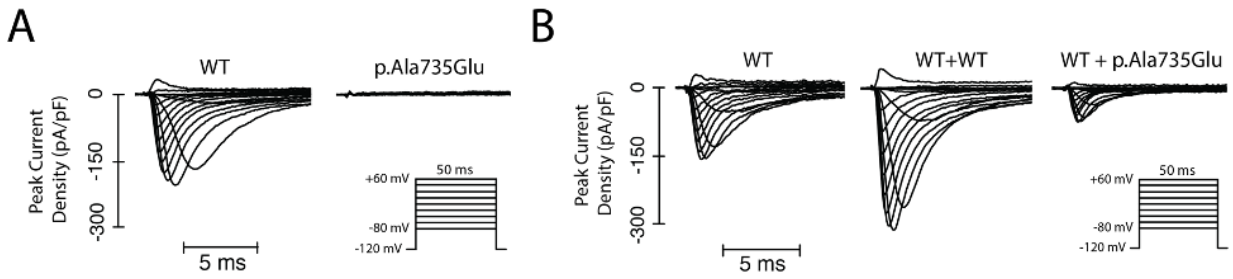


Figure 2.2. Example raw traces from homozygous and heterozygous cell lines.
A) Representative raw peak current densities in a WT and p.Ala735Glu cell. Inset: voltage protocol used.
B) Representative raw peak current densities in a single transfected WT, dually integrated WT+WT, and WT+p.Ala735Glu cell.

We then tested each LoF variant in heterozygous expression using engineered cell lines (WT+variant). Figure 2.2B shows representative traces of cells expressing WT, WT+WT, and WT+p.Ala735Glu (an example dominant negative variant). Figure 2.3B presents peak I_{Na} for the same 35 LoF variants studied in homozygous configuration. WT+WT cells expressed peak I_{Na} of 218.4±7.7% relative to WT alone in LP-SB-SCN5A cells, i.e., those expressing a single WT allele. By contrast, 32/35 of the WT+variant cell lines showed <75% peak I_{Na} compared to LP-SB-SCN5A cells, indicating a dominant negative effect. The heterozygous dominant negative variants displayed a gradient of effect, from 13.9±3.3% to 74.4±5.4% of WT alone. Two previously studied dominant negative variants, p.Arg104Trp and p.Arg121Trp[83], both also exhibited dominant

negative effects in this study (69.6±7.3% and 52.7±8.4% of WT, respectively). While p.Trp822Ter, p.Gly1661Arg, p.Ser1672Tyr, and p.Arg893Cys had LoF peak currents in homozygous experiments, they did not exhibit a dominant negative effect.

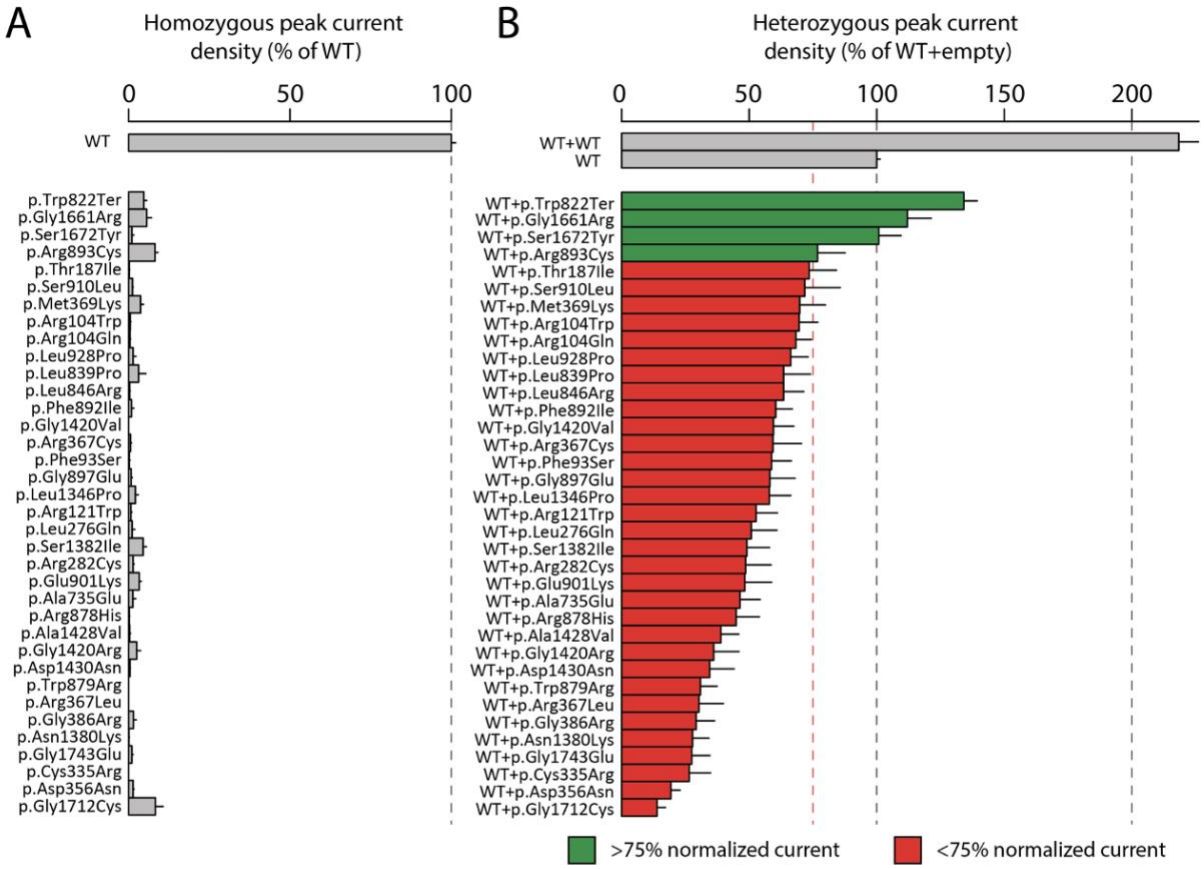


Figure 2.3. Homozygote and Heterozygote Peak Current Measurements.
A) Measurement of homozygous peak current density in 35 *SCN5A* missense variants and one nonsense variant (normalized to WT). Mean ± standard errors. 11-67 cells were studied per variant.
B) Peak current density measurements for 35 *SCN5A* missense variants and one nonsense variant in expression with WT *SCN5A* (normalized to single WT). Mean ± standard errors. 27-164 cells were studied per variant.

Prior to each SyncroPatch experiment, we assessed expression of each variant using flow cytometry. We observed consistently high expression of our reporter gene mCherry sharing the same promoter as the *SCN5A* variants in homozygous and heterozygous cells. In an orthogonal experiment, we tested the expression of selected variants in homozygous expression by Western blotting (Figure 2.4). Consistent with the flow cytometry measurements, we observed uniform expression of mCherry by Western blot; however, we observed somewhat variable expression of

$Na_V1.5$ across 8 tested variants. This result could indicate a diversity of mechanisms of LoF, including a subset arising from variable rates of protein degradation.

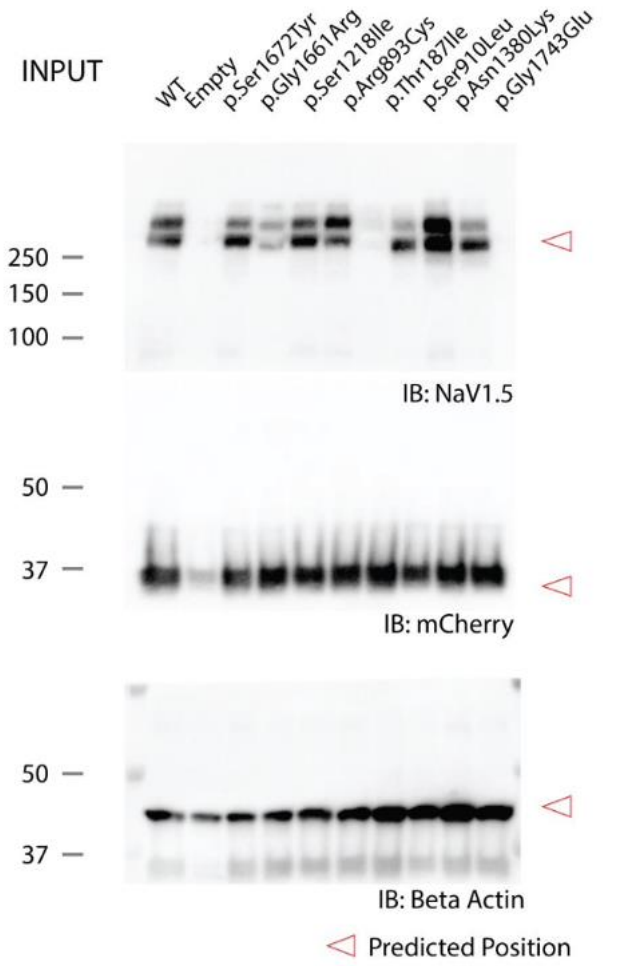


Figure 2.4. Western blot of selected SCN5A variants. Expression of variants was assessed by Western blot for both NaV1.5 and the mCherry reporter. We studied variants with no homozygous current and no dominant negative effect (p.Ser1672Tyr, p.Gly1661Arg, and p.Ser1218I), variants with no homozygous current and a weak dominant negative effect (p.Arg893Cys, p.Thr187Ile, and p.Ser910Leu) and variants with no homozygous current and a strong dominant negative effect (p.Asn1380Lys, p.Gly1743Glu). Predicted positions of bands are shown with red triangles.

Homozygous and Heterozygous Measurements of Partial LoF Variants

We also studied the prevalence of dominant negative effects in 15 partial LoF variants using LP-SB-SCN5A lines. We first confirmed that variant peak currents were 10%-50% compared to WT in homozygous expression with LP-SCN5A cells. (Figure 2.5A). The set of 15 variants included two variants (p.Arg282His and p.Gly1740Arg) previously reported to be LoF but measured as >10% peak I_{Na} in our system[84,85]. Figure 2.5B shows a gradient of I_{Na}, with partial LoF variants showing a greater range of effect in heterozygous expression than those of LoF variants (24.7±5.6% to 231.6±10.8%). 6/15 partial LoF variants had a dominant negative effect whereas the remaining 9 variants all exceeded normalized WT peak current.

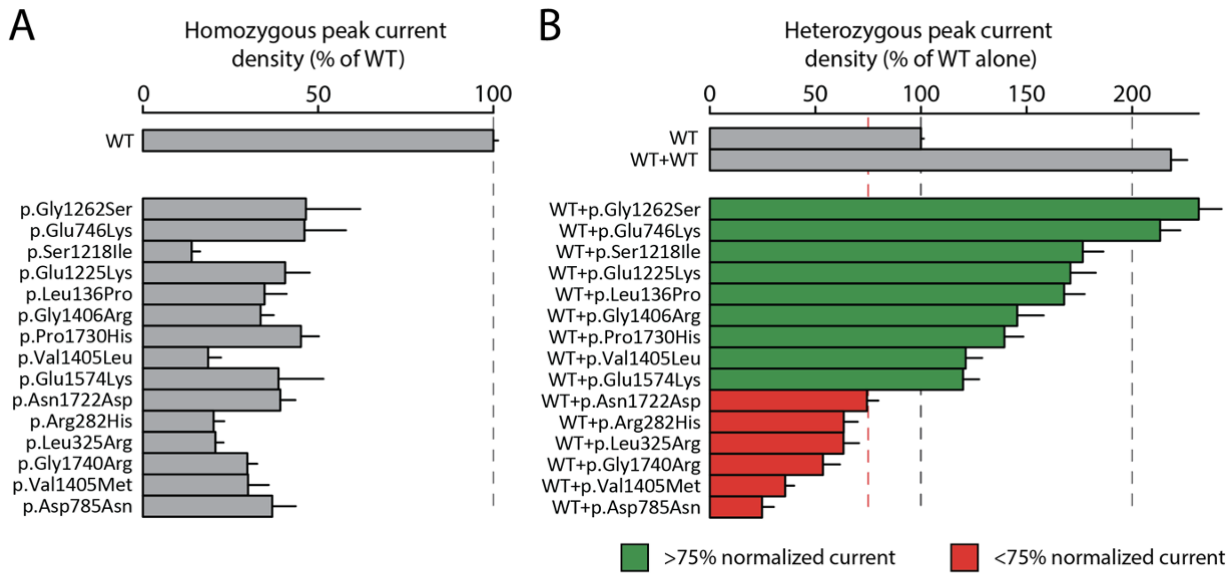


Figure 2.5. Homozygous and heterozygous sodium peak current measurements among partial loss-of-function variants
A) Measurement of homozygous peak current density in 15 partial LoF *SCN5A* variants (normalized to WT). Mean ± standard errors. 8-36 cells were studied per variant.
B) Measurement of heterozygous peak current density in 15 partial LoF *SCN5A* variants (normalized to WT). Mean ± standard errors. 27-53 cells were studied per variant.

Coupled gating in heterozygous expression

In addition to assessing peak sodium current, we also examined other parameters of channel function to measure the extent of coupled gating, a phenomenon where the LoF allele alters the gating properties of the WT allele. These parameters required additional experimental protocols and quality control filters, so these parameters were not comprehensively obtained in all variants studied; only variants with data from >10 qualifying cells are presented. We examined voltage of half activation among the missense variants investigated above (representative raw data shown in Figure 2.6A and 2.6B).

16/50 variants (14 LoF and 2 partial LoF) showed a >10 mV shift in the voltage of half activation, suggesting widespread coupled gating affecting this parameter (Figure 2.7A). We did not observe widespread changes for other parameters beyond voltage of half activation. No variants were shown to induce a shift in voltage of half inactivation >10 mV (Figure 2.7B). One variant (p.Gly1406Arg) had a 1.71-fold change in recovery from inactivation when compared to

WT; the other 34 qualifying variants had <50% shifts in recovery from inactivation (Figure 2.7C). No variants induced late current >1% when co-expressed with WT (Figure 2.7D). Due to the very low or absent peak currents in homozygous LoF variants, it was not feasible to assess parameters other than peak current in homozygous expression.

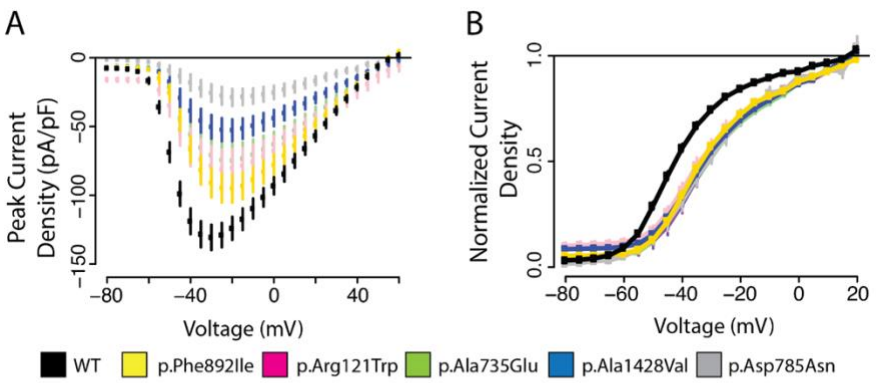


Figure 2.6. Raw Current-voltage traces for five variants with high shifts in voltage of half activation.
A) Current-voltage plot of WT (black) and 5 missense SCN5A variants with large shifts in voltage of half activation: p.Ala735Glu (light green), p.Arg121Trp (pink), p.Asp785Asn (grey), p.Ala1428Val (blue), p.Phe892Ile (gold). The voltage protocol is identical to that used in Figure 2.3.
B) Voltage of half activation curve for WT and 5 missense SCN5A variants (variants and color same as in A).

Elevated BrS Risk Among Dominant Negative Variants Heterozygotes

Having identified dominant negative missense variants as a variant subclass with functional features predictive of increased clinical penetrance, we aimed to characterize the clinical impact of these variants. We therefore compared the relative frequencies of BrS cases to controls among discrete variant classes (missense, indel, haploinsufficient, dominant negative). This approach was inspired by similar studies comparing variant frequencies in cases from disease consortia to those in large population databases as putative controls[70]. We restricted our study to rare variants meeting a population frequency of less than 2.5e-05 as previously described[82]. From these comparisons we could derive odds ratios to quantify the relationship between an exposure (rare *SCN5A* variant) and BrS.

Case and control counts of heterozygotes of the dominant negative variants described above were interrogated using a published consortia of BrS cases[70] and gnomAD, a database of population variation that we considered to contain largely putative controls[77] (Figure 2.8A). In

Figure 2.8B we present the odds ratios (ratio of odds in BrS cohort:gnomAD). The LoF missense dominant negative variants had an odds ratio of 323 compared to 11.0 for missense, 24.2 for indel, and 118 for putative haploinsufficient variants (nonsense, splice, frameshift). Thus, the relative risk of dominant negative missense variants compared to haploinsufficient variants is 2.7 (Fisher's exact test, p = 0.019). All categories were significantly enriched compared to all missense variants (Fisher's exact test, p<0.05).

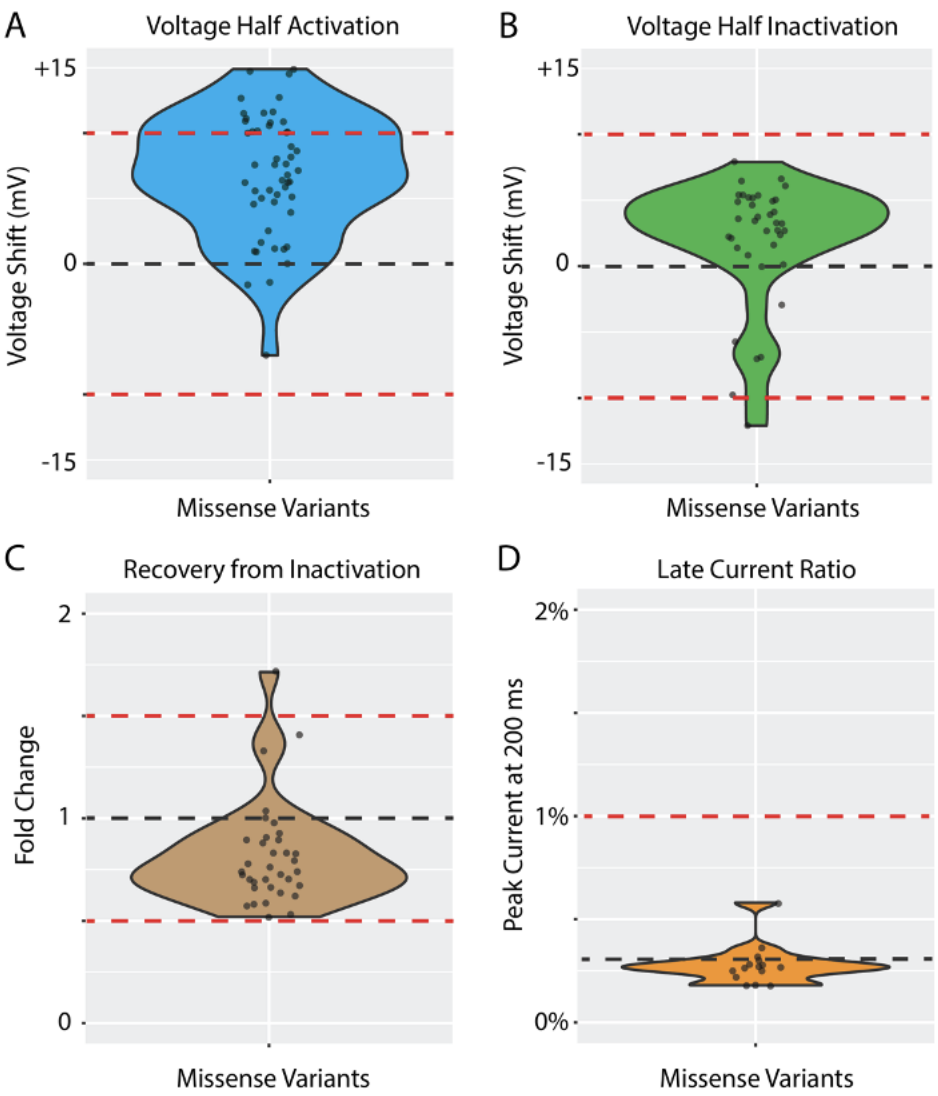


Figure 2.7. Evidence for coupled gating among variant channels in heterozygous expression.
A) Voltage of half activation shift of all missense variants compared to WT. Thresholds correspond to +10 mV and -10 mV shifts from WT to indicate dominant negative effect.
B) Voltage of half inactivation shift of all missense variants compared to WT. Thresholds correspond to +10 mV and -10 mV shifts from WT to indicate dominant negative effect.
C) Time of 50% recovery from inactivation measured in fold change for all missense variants normalized to WT. Thresholds of fold change of 1.5 and 0.5 represent strong deviations from WT.
D) Late current percentage (% of peak current) measured at 200 ms for all missense variants compared to WT. Threshold at 1% indicates threshold of potential dominant negative effect.

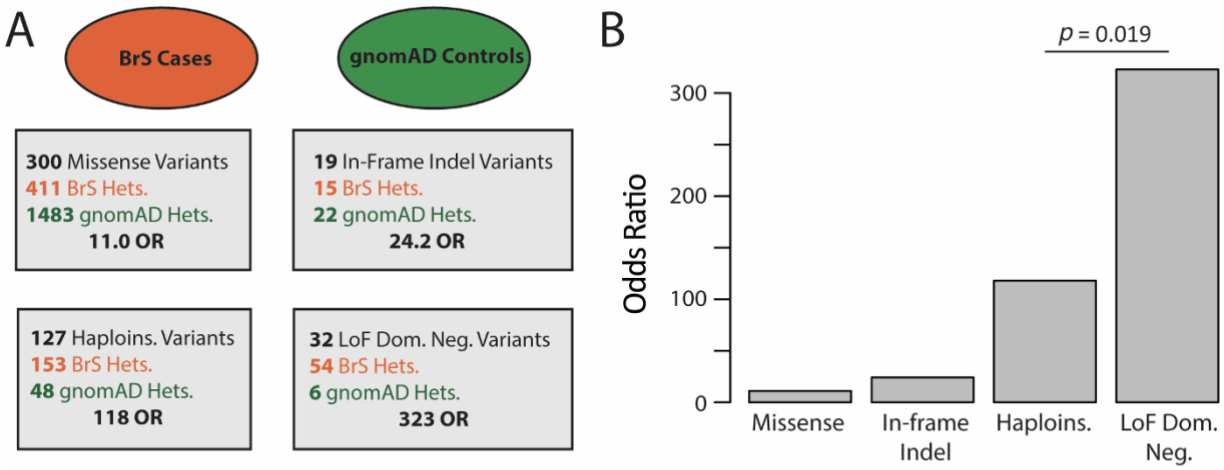


Figure 2.8. Case-control analysis by variant class
A) Case-control breakdown by data source stratified by variant class. BrS cases are shown in red, with putative gnomAD controls shown in green. Heterozygous individuals demarked as Hets. Haploinsufficient (haploins.) indicates nonsense, splice, and frameshift variants. Odds ratios are calculated for each variant class.
B) Barplot of BrS odds ratios by variant class.

Structural Distribution of Dominant Negative Variants

Dominant negative variants were present throughout the structured transmembrane regions of $Na_V1.5$ and did not predominate in any single hotspot region (Figure 2.9A). Structural modeling further showed that dominant negative variants were distributed throughout the three dimensional structure of $Na_V1.5$, with apparent enrichment in the S5-S6 linker domains (Figure 2.9B+C).

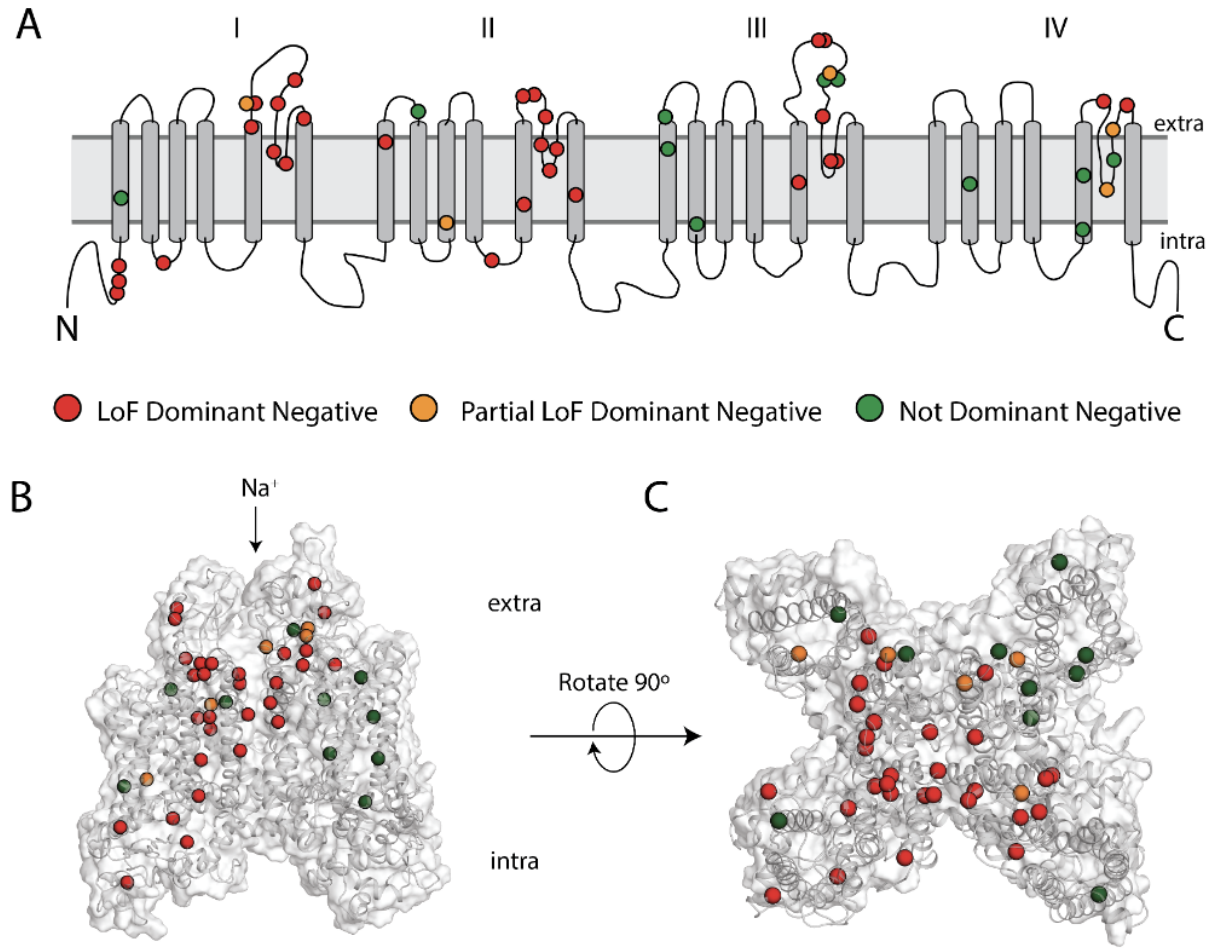


Figure 2.9. Structural distribution of dominant negative variants.
A) Locations of dominant negative variants throughout NaV1.5 in 2D channel rendering. Red indicated LoF dominant negative, orange partial LoF dominant negative, and green non-dominant negative missense variants. Extra: extracellular, intra: intracellular.
B) Side view of NaV1.5 protein with overlaid variant distribution.
C) Top view of NaV1.5 protein with overlaid variant distribution.

DISCUSSION

Dominant Negative Effect Among Most Missense LoF SCN5A Variants. This study assessed the dominant negative properties of 50 LoF and partial LoF variants. A large majority of examined LoF variants (32/35) and some partial LoF variants (6/15) showed dominant negative behavior. Dominant negative effects are pervasive throughout biology, especially for multimeric proteins, and involve several distinct mechanisms to compromise WT function[86]. In the case of *SCN5A,* the dominant negative effect has been posited to arise by both deficient trafficking to the membrane as well as coupled gating at the cell surface. One study showed that the variants p.Arg104Trp and p.Arg121Trp induced a dominant negative effect primarily through endoplasmic reticulum retention of WT protein due to interactions among the channel alpha-subunits[83]. Follow up studies with extensive biochemical analyses showed that the dominant negative variant p.Leu325Arg acted through coupled gating at the cell surface[73].

Previous research suggested that the residues between 493 and 517 are critical for the dimerization and coupled gating of $Na_V1.5$ at the cell surface, and another study found an enrichment of dominant negative variants at the N-terminus of the protein[72,87]. We did not observe an enrichment of dominant negative variants among these previously described residues, but rather a broader distribution of variants spanning the four transmembrane domains of the protein (Figure 2.9) Thus, dominant negative effects appear to be a general property of most LoF missense variants in *SCN5A*, independent of location within the protein. Particularly interesting are examples of disparate effects within close physical proximity, such as the partial LoF variants p.Val1405Met (35.7% peak current in heterozygous expression), p.Val1405Leu (121% peak current), and p.Gly1406Arg (146% peak current). We surprisingly observed a small but non-zero peak current (4.7% of WT) for an early nonsense variant p.Trp822Ter. This variant has an intact Domain I, and therefore may be forming a poorly conducting multimeric channel.

In addition to decreased peak current, we observed that 16/50 variants also influenced voltage of half activation when measured in heterozygous expression with WT. This finding is

consistent with the concept of coupled gating at the cell surface, and reflects the influence of the LoF allele on properties of the WT allele of the protein, possibly through a multi-channel complex[73]. These shifts in $V_{1/2}$ activation in a loss of function direction combine with reduced peak currents to result in additional reduction of channel function in heterozygous expression. $V_{1/2}$ activation was the only additional property that varied substantially from WT $Na_V1.5$ activity, as we did not observe large differences in voltage of half inactivation, recovery from inactivation, or late current.

While dominant negative effects were very common among the variants studied, there were LoF variants and partial LoF variants that did not diminish peak current. These results further support the idea that there are yet to be recognized channel-channel interactions that may or may not induce dominant negative behavior. In experiments assessing protein expression, we found a range of relative expression of $Na_V1.5$ compared to the mCherry reporter by Western blot. We did not observe a strong correlation between degree of dominant negative effects and relative protein expression (Figure 2.4). This may indicate a diversity of underlying mechanisms including protein degradation, altered gating, or aberrant trafficking.

Previous studies[72,73,83] have experimentally demonstrated at least two mechanisms for individual dominant negative variant effects: coupled gating and deficient trafficking. Our observation of 16 variants that modulate voltage of half activation indicate that these variants may act through coupled gating. However, at this point the mechanism of the remaining dominant negative variants is unclear. Through Western blot data, we observe a range of protein expression in selected variants; these remaining variants may act by causing a failure to traffic to the plasma membrane or may traffic properly and reduce channel function at the plasma membrane.

Clinical Implication of Dominant Negative Heterozygote Status. Our results indicate that dominant negative *SCN5A* variants are significantly more likely to cause manifest Brugada Syndrome than other missense variants. Previous work has established that homozygous peak current of *SCN5A*

variants is the strongest *in vitro* electrophysiological predictor of each variant's BrS risk[69,88]. Since dominant negative missense variants cause an especially low peak current, we hypothesized that dominant negative variants would confer an especially high risk for BrS. Importantly, our expanded catalog of 38 dominant negative *SCN5A* variants enabled us to calculate cohort-based estimates of disease risk of this class of variants. Using gnomAD and a recently published cohort of BrS cases[70,77], we demonstrated that dominant negative variants are highly overrepresented in cases vs controls when compared to other variant classes, with a striking odds ratio of 323 for dominant negative LoF missense variants. In contrast, other variant classes have lower odds ratios of 11 (all missense variants) or 118 (putative haploinsufficient frameshift/nonsense/splice site variants). Thus, the relative risk of BrS among dominant negative LoF missense variants compared to putative haploinsufficient variants is 2.7. Previous studies have shown that truncating and functionally inactive missense variants cause a more severe phenotype than partially active missense variants, but the penetrance of dominant negative variants had not yet been extensively studied[89]. Our results indicate that the penetrance of dominant negative missense variants is higher than penetrance of other variant classes. One potential explanation for the different disease penetrance among variant classes is that nonsense mediated decay (NMD) removes aberrant transcripts for splice-altering and nonsense variants, preventing their interaction with WT $Na_V1.5$. Given the data presented here, dominant negative missense variants should arouse high clinical suspicion for BrS risk when detected in patients.

Current evidence suggests that peak $Na_V1.5$ current is highly predictive of clinical BrS penetrance[69,88]. Our finding of a higher BrS penetrance for dominant negative loss of function variants compared to other classes (putative haploinsufficient and non-dominant negative missense variants) further supports a model in which total $Na_V1.5$ peak current (adjusting for dominant negative effects) is correlated with disease severity.

High-throughput Electrophysiological Assays to Study Dominant Negative Effects. High-throughput automated patch clamping has emerged as a tool for rapidly assessing functional consequences of ion channel genetic variation[53]. This technique has been used to assess pathogenicity of variants in *KCNQ1*[90], *SCN5A*[76], and *KCNH2*[91]. Here, we present the most extensive evaluation of heterozygous $Na_V1.5$ expression to date using this platform, studying 51 total variants with 27-164 cells per heterozygous measurement. Heterozygous measurements are already common for the cardiac potassium channels *KCNQ1* and *KCNH2*; this study suggests that heterozygous studies may also be necessary for LoF *SCN5A* variants in future studies. This work shows that high-throughput automated patch-clamp can help establish molecular mechanisms of disease.

Limitations. Results from heterologous expression in HEK293T cells may not fully recapitulate behavior in native cardiomyocytes in human hearts. In particular, contributions such as polygenic modifiers, as has been previously observed in BrS[92], may not be fully captured by this non-native system. Two common alternative splicing events impact SCN5A splicing (p.Gln1077 deletion/insertion and fetal/adult exon 6); only the most common splice isoform in the adult heart was examined in this study. The gnomAD population database does not have available phenotypic information, so a small fraction of individuals included in gnomAD may in fact have BrS.

Conclusions. Most LoF missense variants in SCN5A have a dominant negative effect. These missense dominant negative variants have a 2.7-fold increased risk of BrS when compared to haploinsufficient variants. These results may help refine prediction of BrS risk in dominant negative variant heterozygotes.

Chapter 3

Continuous Bayesian Variant Interpretation Accounts for Incomplete Penetrance among Mendelian Cardiac Channelopathies

Portions of this chapter are published under the same title in *Genetics in Medicine*

Background: The congenital Long QT Syndrome (LQTS) and Brugada Syndrome (BrS) are Mendelian autosomal dominant diseases which frequently precipitate fatal cardiac arrhythmias. Incomplete penetrance is a barrier to clinical management of heterozygotes harboring variants in the major implicated disease genes *KCNQ1, KCNH2*, and *SCN5A*. We apply and evaluate a Bayesian penetrance estimation strategy that accounts for this phenomenon and evaluate penetrance distributions and rationalize their structural underpinnings across four genotype-phenotype pairs.

Methods: We generated Bayesian penetrance estimation models for *KCNQ1*-LQT1 and *SCN5A*-LQT3 using variant-specific features and clinical data from the literature, international arrhythmia genetic centers, and population controls. We analyzed the distribution of posterior penetrance estimates across four-genotype phenotype relationships and compared continuous estimates to ClinVar annotations. Posterior estimates were mapped onto protein structure.

Results: Bayesian models of *KCNQ1*-LQT1 and *SCN5A*-LQT3 are well-calibrated to clinical observations. Variant-informed penetrance estimates of *KCNQ1*-LQT1 and *SCN5A*-LQT3 are empirically equivalent to 10 and 5 heterozygote clinical phenotypes, respectively. Posterior penetrance estimates were bimodal for *KCNQ1*-LQT1 and *KCNH2*-LQT2, with a higher fraction of missense variants with high penetrance among *KCNQ1* variants. *SCN5A*-LQT3 and *SCN5A*-BrS had comparatively more variants with predicted low penetrance. There was a wide distribution of variant penetrance estimates among similar categories of ClinVar annotations. Structural

mapping revealed heterogeneity among ‘hot spot’ regions and featured high penetrance estimates for *KCNQ1* variants in contact with calmodulin and the S6 domain.

Conclusions: Bayesian penetrance estimates provide a continuous framework for variant interpretation, provide higher resolution within ‘hot spot’ domains, and facilitate prospective clinical management of variant heterozygotes.

Introduction

Incomplete penetrance is a hallmark of the inconsistency of genetic determinism, where heterozygotes harboring the same variant may manifest a continuum of the disease phenotype, or none at all[24,93,94]. The nuance of incomplete penetrance is frequently unaccounted for in contemporary genetic methods development; probabilistic approaches may more accurately account for variable genetic contributions to disease presentation. The American College of Medical Genetics and Genomics (ACMG) has established criteria for adjudicating the 'pathogenicity' of variants, drawing to varying extents on population frequency data, segregation within families, *in silico* methods, and functional data, among other sources[43]. However, this categorical approach is unable to account for incomplete penetrance, especially among ClinVar-annotated Pathogenic or Likely Pathogenic variants[45].

Clinical management of patients affected by inherited arrhythmic syndromes is a salient example of this challenge. The Long QT Syndromes Type 1-3 (LQT1-3 [MIM: 192500, 613688, 603830], respectively) and Brugada Syndrome (BrS [MIM 601144]) are frequently responsible for Sudden Cardiac Death[26]. The most rigorously established genotype-phenotype relationships are loss-of-function variants in *KCNQ1* encoding K_V7.1 (LQT1), *KCNH2* encoding K_V11.1 (LQT2), and *SCN5A* encoding Na_V1.5 (BrS) and gain-of-function variants in *SCN5A* (LQT3)[30,36]. These membrane proteins are responsible for orchestrating the cardiac action potential through cardiomyocyte depolarization and repolarization. All four of these genotype-phenotype pairs are characterized by a significant degree of incomplete penetrance[95], which may in part be accounted for by environmental risk factors as well as polygenic liability[96-98]. The ability to accurately predict variant penetrance is especially important, since medical or interventional therapies are well established to lower the risk of fatal ventricular arrhythmias[26]; however, overly aggressive management may conversely cause harm[99,100]. New methods for variant interpretation that capture the granularity of clinical penetrance in a probabilistic manner will help facilitate the clinical management of variant heterozygotes.

Herein, we elaborate a framework where variant-specific features inform a continuous, quantitative probability of disease manifestation (penetrance) for missense variants – in contrast to the probability of 'pathogenic' or 'benign' classification[101,102]. These prior estimates are calibrated to clinically observed penetrance from the spectrum of all available variant-specific feature data. Within a Bayesian framework, we apply clinical evidence from phenotyped individuals from the literature, arrhythmia genetics clinical cohorts, and putative population controls to these calibrated prior estimates of disease probability to yield posterior penetrance estimates (Figure 3.1). This approach recasts nuanced variant features, previously only interpretable by specialists, e.g., structural and channel biophysical properties, to a wider audience as an equivalent number of clinically phenotyped heterozygotes. These estimates can therefore inform clinical management in the absence of prior clinical observations of a variant, or aid in interpretation of variants when few heterozygotes have been clinically phenotyped.

Implementing two instances of this approach, we develop penetrance models for *KCNQ1*-LQT1 and *SCN5A*-LQT3, incorporate data from *KCNH2*-LQT2[103] and *SCN5A*-BrS1[104], and investigate penetrance across these four genotype-phenotype relationships. We demonstrate how penetrance varies among these genotype-phenotype relationships and correlate penetrance with structural features of each gene product.

Materials and Methods

General Information

All analyses in this study use the following transcripts: *KCNQ1* - ENST0000155840; *SCN5A* - ENST00000333535. Penetrance estimates are based on amino acid substitutions. For amino acids substitutions achievable by multiple single nucleotide polymorphisms, we treat the cDNA alterations as equivalent. Statistical analyses were carried out in R Studio Version 4, protein structures were visualized by PyMol Version 2.5.0.

Description of Methodology

Methodology and Definition of Terms

Observed Penetrance we define as the number of affected heterozygotes over total heterozygotes for a particular variant. Observed penetrance is necessarily inaccurate due to limited sample size but approximates the 'true' penetrance at very large sample sizes; α and β represent the number of affected and unaffected individuals heterozygous for the same variant, respectively.

$$Observed\ Penetrance = \frac{\alpha}{\alpha + \beta} \quad \text{(Equation 1)}$$

Empirical Prior. The beta binomial model is used to estimate parameters $\alpha_{prior,\ empirical}$ and $\beta_{prior,\ empirical}$ for all variants across the gene based on a weighted average and mean squared error across all variants. For *KCNQ1-LQT1*, we fit a model that provides an $\alpha_{prior,\ empirical} = 1.13$ and $\beta_{prior,\ empirical} = 1.57$. These parameters then provide an Empirical Prior applied to all variants according to the equation below. In the absence of any additional information about a variant other than it is a missense/inframe indel, this provides an approximation of LQT1 penetrance which is the same for all *KCNQ1* variants (disease probability of 0.42 for *KCNQ1*-LQT1).

$$Empirical\ Prior = \frac{\alpha_{prior,empirical}}{\alpha_{prior,empirical} + \beta_{prior,empirical}}$$ (Equation 2)

Empirical Posterior. The Empirical Prior can be refined using per-variant observed affected (α) and unaffected (β) heterozygotes as a likelihood. This provides a best estimate of penetrance for each variant based on heterozygote observations alone. This per-variant value is also termed the Empirical Posterior penetrance, against which model covariates are correlated.

$$Empirical\ Posterior = \frac{\alpha + \alpha_{prior,empirical}}{\alpha + \beta + \alpha_{prior,empirical} + \beta_{prior,empirical}}$$ (Equation 3)

Expectation-Maximization (EM) Algorithm. The Empirical Posterior can be used to calibrate a Bayesian Prior, with the point estimate and associated uncertainty estimated separately (see section below on *Tuning Parameter*). The variant-specific disease penetrance point estimate is generated using an expectation maximization framework, as previously reported by our lab[103,104]. During the regression step, the Empirical Posterior is used as a dependent variable, and the specified variant-specific features are treated as independent covariates. Through this framework, we iteratively refine variant-specific point estimate of disease probability initially informed by α_{prior} and β_{prior} values until the point estimate converges. The present EM algorithm comprises 3 steps: 1) calculate the expected penetrance from an empirical Bayes penetrance model, 2) regress estimated penetrance to variant-specific features using maximum likelihood, and 3) revise point estimates of disease penetrance using the fit from step 2, iterating steps 2-3 until convergence criteria are satisfied. The EM algorithm was performed using the covariates LQT1-dist, CardioBoost, Provean, REVEL, and homozygous peak current. Missing data was handled using a pattern mixture model, with a model fit for each missing data pattern[105].

Tuning Parameter `v`. To estimate the uncertainty in our disease penetrance point estimates, for each variant, we sampled 100 subjects using a binomial distribution with a probability defined by the Empirical Posterior penetrance (Equation 3) for each variant to generate "LQTS sampled". We then calculated the resulting 95% posterior credible interval from the Beta distribution with shape parameters α) $\text{LQTS sampled} + \alpha_{\text{prior,EM}}$ and β) $100 - \text{LQTS sampled} + \beta_{\text{prior,EM}}$. We repeated this process 1000 times and calculated the rate the posterior credible interval covered the "true" observed LQTS probability (Empirical Posterior). Through this evaluation we can test different values of `v`, interpretable as the number of hypothetically observed and phenotyped heterozygotes generated by the variant feature set. Both large and small `v` values result in coverages which deviate from the expected 95%, many below and above, respectively. We selected the best `v` from the coverage plots which balances the tradeoff of over-coverage in variants with medium-low LQTS probability and under-coverage of variants with high LQTS probability. From this procedure, we chose a tuning parameter of 'v' = 10 for KCNQ1-LQTS and 'v' = 5 for SCN5A-LQTS.

Bayesian Prior. With a per-variant point estimate defined for each variant and estimated uncertainty defined by v, we define new $\alpha_{\text{prior, EM}}$ and $\beta_{\text{prior, EM}}$. The Bayesian Prior for each variant is thus:

$$Bayesian\ Prior = \frac{\alpha_{prior,EM}}{\alpha_{prior,EM} + \beta_{prior,EM}} \text{ (Equation 4)}$$

Bayesian Posterior. When available, per-variant heterozygote observations (α and β as defined above) are again applied as a likelihood to the Bayesian Prior to yield a Bayesian Posterior.

$$Bayesian\ Posterior = \frac{\alpha + \alpha_{prior,EM}}{\alpha + \beta + \alpha_{prior,EM} + \beta_{prior,EM}} \text{ (Equation 5)}$$

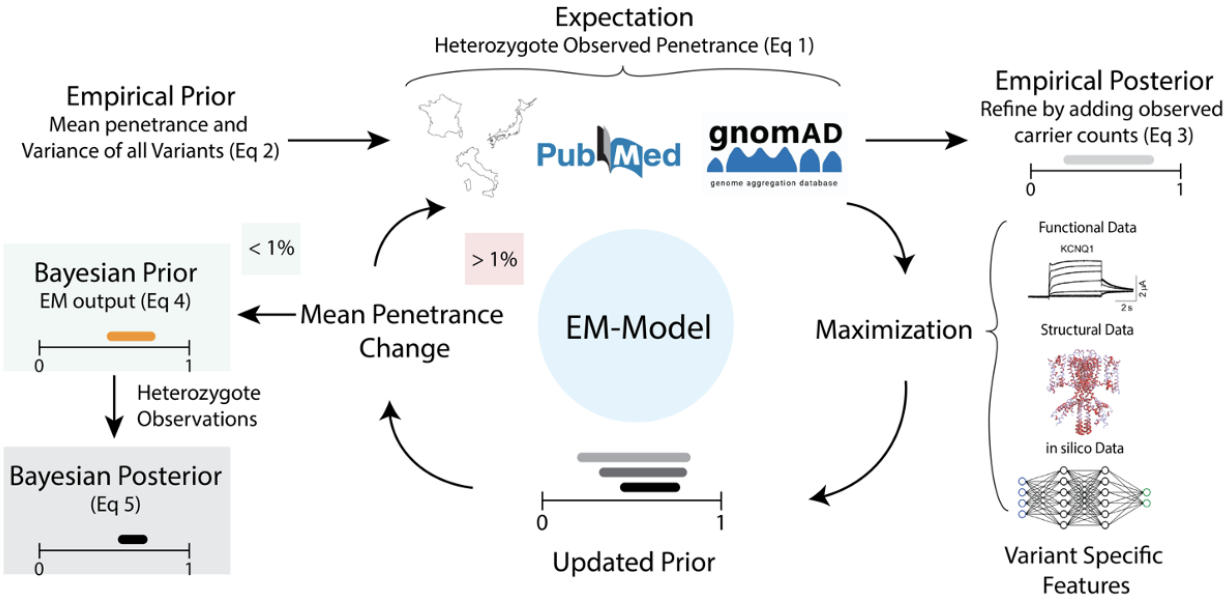


Figure 3.1. Detailed Schematic of Penetrance Methodology. An Empirical Prior is derived from the average penetrance and variance of all variants. We add variant specific counts to derive an Empirical Posterior for each variant. Within the context of an Expectation-Maximization model, we update the Empirical Posterior with variant specific features. This provides an updated prior, which is further iterated with clinical data and variant specific features until the penetrance does not continue to change. This outputs a Bayesian Prior used in future analyses.

Statistical Analyses.

Statistical evaluation. All reported correlation coefficients, Spearman ρ, and Pearson's R^2 were weighted with Equation 6 to ensure variants with higher total heterozygote counts had greater influence in the resulting correlation coefficient estimate.[103,104]

$$F(x) = 1 - \frac{1}{0.01 + total\ heterozygotes} \quad \text{(Equation 6)}$$

10-Fold Cross-Validation. We performed 10-fold cross-validation using the Caret package in R. All variants with at least one heterozygote were included in the model. We calculated a cross-validated Spearman correlation and Pearson correlation between the Empirical Posterior and Bayesian Prior for each variant (see Results). We report summary statistics for the average

correlation coefficients/scores and variances observed for 10 independent implementations of the 10-fold cross-validation.

Penetrance Density Covariate. Information contained within the 3D-environment of each residue was used in our model as a LQT1- or LQT3-dist, a covariate previously described by our group[106]. We implemented this covariate in 2 different ways. To evaluate this covariate, we used a Leave One Out Cross Validation strategy to assess the correlation with Empirical Penetrance (using Empirical Posterior, Equation 3). This ensures the correlation is not inflated by exposure to itself during feature evaluation. For best estimates provided for prospective use on the variant browser website, we leveraged all available variant data.

Curation of Covariates and Clinical Data.

Appraisal of Literature Data. The presence or absence of a clinical diagnosis of LQTS was defined by a Schwartz score ≥3.5 or a single QT interval ≥480 ms[1073]. Affected status of LQT1 was annotated in compliance with the diagnostic guidelines published by the HRS/EHRA/APHRS Expert Consensus Statement on the Management of inherited Arrhythmias[108], withholding point contributions from genotype-positive status. In literature reports where patient level data was not available for manual review, the status was assigned in accordance with that reported by the authors. PubMed was searched for the term “KCNQ1” and identified a total of 2216 manuscripts, of which 326 provided sufficient information to interpret meaningful clinical phenotypes. Each paper was manually reviewed for LQT1 status in affected and unaffected individuals (described below). The full set of heterozygotes for variants in *KCNQ1* from gnomAD was added to this curation. All individuals from gnomAD were assumed not to have LQT1, an assumption previously shown not to alter the model performance[103,104]. From the literature and gnomAD 9969 heterozygotes for 630 unique missense or in-frame insertion/deletion (indel) variants were identified. For *SCN5A*-LQT3, a previously reported dataset was queried for association with

LQTS[104]. This comprised 86118 heterozygotes with 1667 unique missense or in-frame insertion/deletions.

Collation of Cohort Data. Three international centers with arrhythmia genetics expertise contributed clinical phenotypes and genotypes for *KCNQ1* probands and family members when available; Italy (600 heterozygotes harboring 76 unique missense variants), France (438 heterozygotes with 88 unique missense variants), and Japan (350 heterozygotes with 82 unique missense variants). The literature data was reviewed to remove potential overlap of patients between the cohort and literature set. The non-overlapping literature dataset and cohort set were combined and optimism for each presented performance metric was estimated using 10-fold cross validation.

KCNQ1 *Variant Specific Features: Functional Data, Structural Data, and* in silico *Covariates.* Functional data were collected from a previously assembled dataset of *KCNQ1* variant functional data[106], and from recent high-throughput automated patch clamp experiments[90,109]. When multiple functional measurements were available for a single variant, high-throughput SyncroPatch measurements were given priority. A published cryo-EM model of *KCNQ1* enabled high-resolution structural studies[110]. We applied our previously developed penetrance density metric as a covariate in the *KCNQ1*-LQT1 model (see previous manuscript for full description)[106]. We obtained pathogenicity predictions using the *in silico* servers CardioBoost[111], PolyPhen-2[112], and PROVEAN[113]. Conservation scores were obtained by the basic local alignment search tool position-specific scoring matrix (BLAST-PSSM). The meta-predictor REVEL was additionally included given its previous impressive performance in similar applications[114].

Downstream analyses.

ClinVar Analyses. We curated all *KCNQ1* and *KCNH2* variants in the ClinVar database with Benign/Likely Benign and Pathogenic/Likely Pathogenic annotations[45]. We merged the table of categorical classifications with our complete clinical and covariate dataframes. For all variants with at least 1 heterozygote observed in our dataset, we plotted the observed penetrance (affected heterozygotes/total heterozygotes) for each variant, followed by the Bayesian Penetrance estimate for each.

Structural Penetrance. We exported a dataframe containing each amino acid residue and the probability density covariate for each of four genotype-phenotype pairs. We visualized per residue penetrance as a color gradient with probability densities as weights (from 0 to 1) with PyMol.

Description of Code.

We have deposited code used to generate figures and data on the Kroncke lab GitHub page - https://github.com/kroncke-lab/Bayes_Incomplete_Penetrance. There are 3 R-markdown files that together yield the analyses presented in the main manuscript.

Bayes KCNQ1 Data Process – these scripts merge the cohort data with the curated literature data and provide counts of affected LQT1 heterozygotes and total heterozygotes for each variant. Functional, *in silico*, and structural data are combined in a separate dataframe for all variants that could arise from a single nucleotide substitution (not permitting all amino acids at each position). These dataframes are combined and used in the subsequent scripts.

Bayes KCNQ1 Model Evaluation – these scripts implement the prior and posterior probability derivations described above. The significance of each covariate is calculated using Spearman

and Pearson correlations. To assess the robustness of the method, we implemented 10-fold cross validation for Spearman, Pearson, and Brier scores.

Bayes KCNQ1 Model Application – these scripts construct the model and derive prior and posterior probabilities as described above. We then use scripts to generate the prior and posterior penetrance histograms, conduct analyses of variant penetrance with ClinVar annotations, export the penetrance data that is superimposed on structure, and make forest plots of prior and posterior probabilities for all variants, as described in above.

Description of Functional Covariates.

KCNQ1 is a homotetrameric protein that provides a channel for potassium ions to flow out of the cell during electrical repolarization. This is an essential process in homeostasis of the cardiomyocyte following sodium-mediated depolarization. Aberrations in cardiac repolarization are the molecular correlate of the clinical Long QT Syndrome and can precipitate fatal ventricular arrhythmias. Ion channels may exist in a variety of states, such as open, closed, and inactivated. Channel gating describes the process by which ion channels, such as KCNQ1, conformationally change between these states. Genetic variants may affect the kinetics of transitions between these states, and thereby affect conductance of current across the membrane. These general principles also apply to SCN5A ($Na_V1.5$) and KCNH2 (hERG) with small alterations for gating transitions based on relevant structural features. See Chapter 1 for a complete description of electrophysiologic covariates.

Results

Summary of Bayesian Penetrance Methodology. A challenge in genetic medicine is to provide a quantitative probability of disease given the presence of a variant, the variant-induced disease penetrance. This information is critical to facilitate disease diagnosis and is increasingly useful to inform treatment for those heterozygotes. To advance this goal, we constructed a penetrance prediction model that incorporates multiple lines of evidence to estimate this probability and associated uncertainty. The evidence for the model comes from a variety of variant features – protein structural data, functional electrophysiology, and *in silico* predictions of variant effect– in

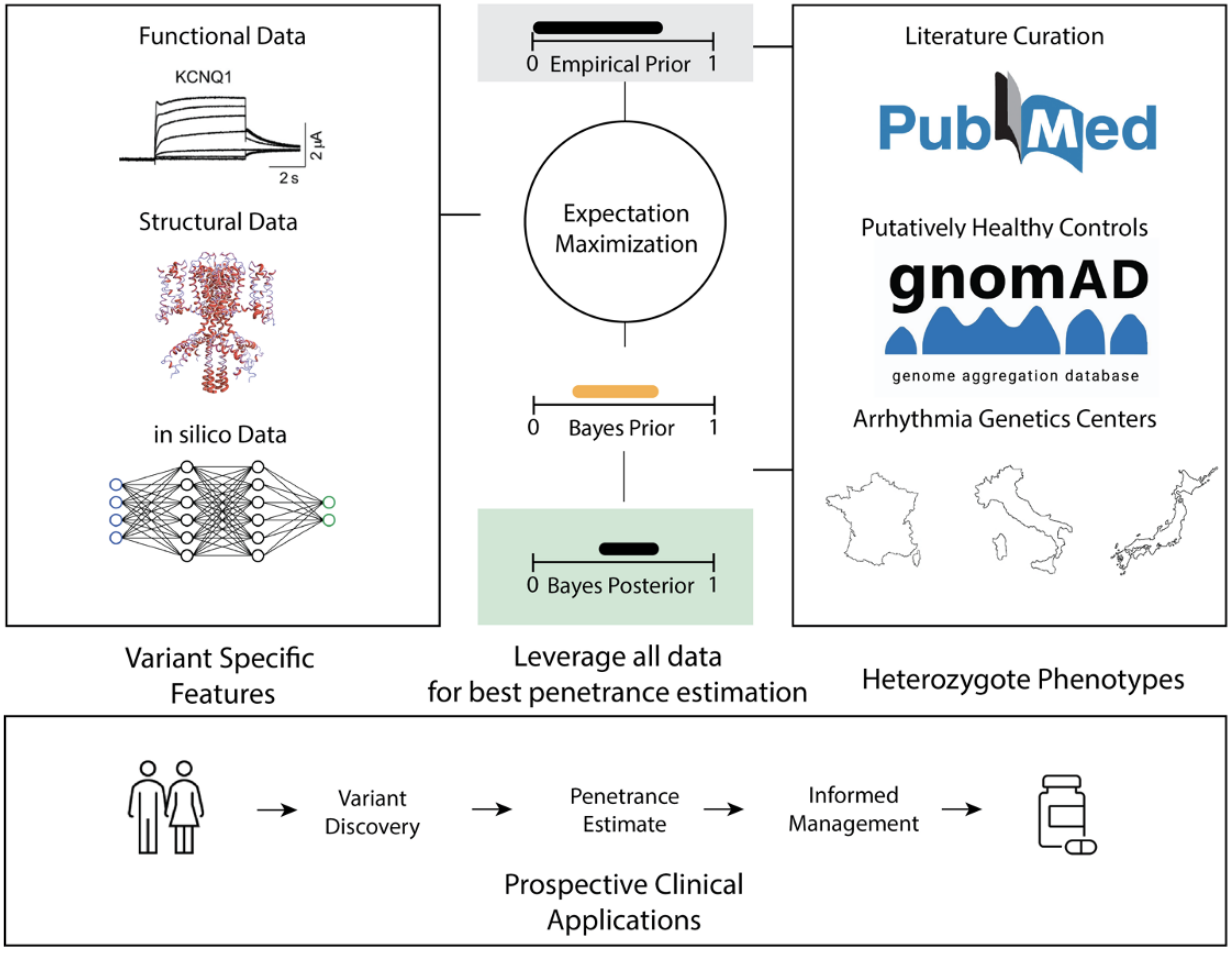


Figure 3.2 – Overview Schematic of Bayesian Penetrance Model.
We combine information contained within variant specific features (channel functional data, protein structural data, *in silico* predictions) with variant heterozygote clinical data (affected vs unaffected status) to estimate disease penetrance for each variant. Data are integrated through an expectation-maximization approach within a Bayesian framework. Quantitative penetrance estimates can then be used for clinical applications to enable personalized medicine.

addition to clinical phenotyping of variant heterozygotes (Figure 3.2; see Methods and Figure 3.1 for detailed derivations).

We begin our model with data contained in phenotyped heterozygotes. We first applied an Empirical Prior to all variants based on the mean penetrance across all variants weighted by their total heterozygote count (see methods for details); per-variant observations were summed across all heterozygotes regardless of proband or kindred status. We updated this Empirical Prior using per-variant phenotype observations within a beta-binomial Bayes model. We term the resulting penetrance estimate the Empirical Posterior. 'True' variant penetrance can only be approximated; this Empirical Posterior estimate therefore provides the 'best guess' estimate of penetrance using heterozygote observations alone.

We posit that these Empirical Posterior estimates, informed only by heterozygote observations, can be preconditioned using complementary information contained in variant-specific features. Accordingly, we update the Empirical Posterior with variant-specific features as a likelihood, implemented through an iterative Expectation-Maximization (EM) model, to yield a Bayesian Prior. Our modified EM algorithm is an iterative technique comprising three steps: 1) calculate an Empirical Posterior for each variant, 2) fit a regression model for Empirical Posterior point estimates on variant-specific characteristics by maximum likelihood and 3) revise our point estimates of disease penetrance using the fit from step 2 then iterate steps 2-3 until convergence criteria are satisfied. This convergence results in disease probability point estimate which, when combined with an estimate of uncertainty, yield a Bayesian Prior. We restrict subsequent analyses to variants with heterozygote observations which can be combined with the Bayesian Prior to yield the Bayesian Posterior, hereafter used interchangeably with Bayesian penetrance.

Evaluation and Performance of KCNQ1 *Variant Penetrance Model.* We implemented a Bayesian penetrance model using all available data for the *KCNQ1*-LQT1 genotype-phenotype pair[103,104].

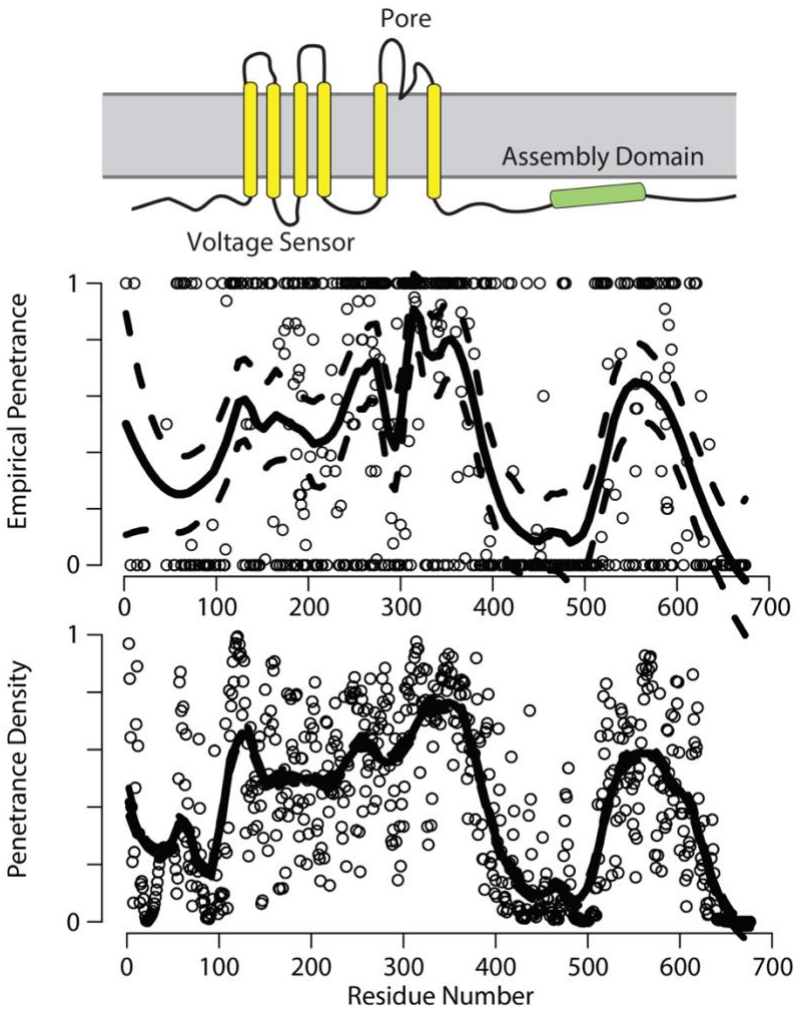


Figure 3.3 – Empirical penetrance and penetrance density by residue.

We curated >2,000 manuscripts from the literature, and in combination with phenotyped heterozygotes from international arrhythmia clinics and putative heterozygote controls from gnomAD[77], found 9,969 *KCNQ1* heterozygotes harboring 630 unique missense and in-frame insertion/deletion variants. Plotting the observed penetrance of each variant by residue recapitulated known associations with 'hot spot' regions of the protein; plotting our structural penetrance density metric[106] (residue weighted by penetrance of adjacent residues in 3D environment) by residue resulted in more continuous estimates (Figure 3.3).

We evaluated the functional, *in silico*, and structural features against the estimated LQT1 penetrance (the observed penetrance modestly biased towards the mean, called the empirical posterior, see Materials and Methods) using Spearman-rank order coefficients, weighted towards variants with many clinical observations (Figure 3.4). We found many channel biophysical properties were statistically significant (nominal $p < 0.05$). For example, reductions in peak current decrease repolarization and, as the primary molecular correlate of LQT1, predict higher LQT1

penetrance (Spearman correlation coefficient -0.55 homozygous, -0.54 heterozygous). CardioBoost[111] had the highest correlation coefficient (+0.64) of the *in silico* predictors, followed by REVEL[114] (+0.64) and Provean[113] (-0.61). The structural LQT1-dist covariate had a Spearman correlation coefficient of +0.54. Integrating the most relevant features using our previously described method[103], we derived a LQT1 Bayesian prior with a coefficient of +0.71 (10-fold cross validated Spearman correlation coefficient +0.65).

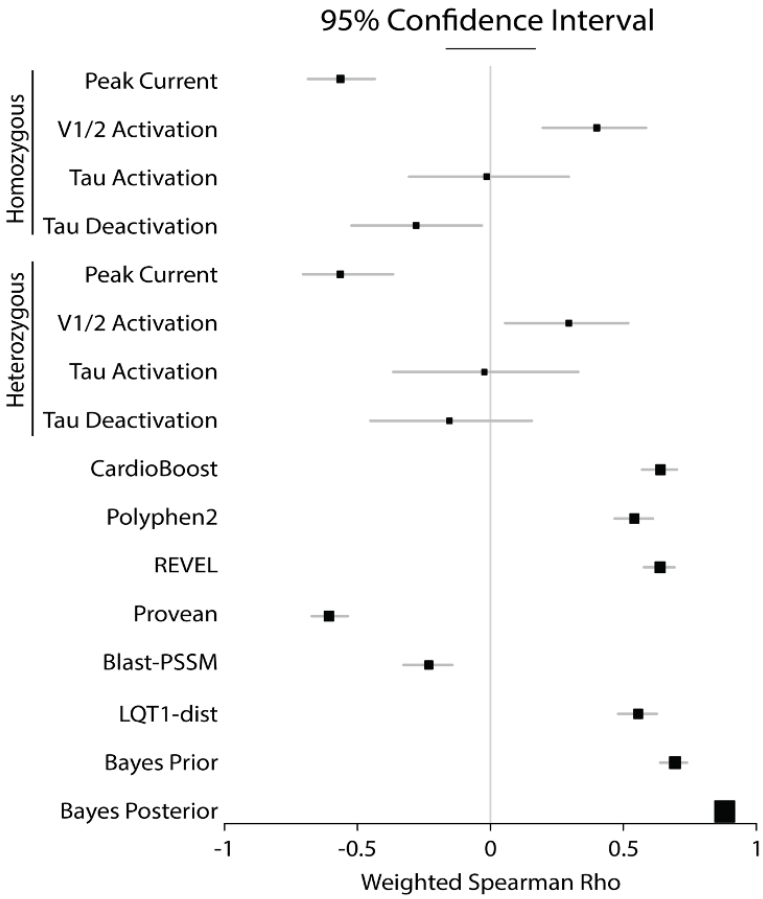


Figure 3.4 Weighted rank order correlations and 95% confidence intervals for each covariate against LQT1 Empirical Penetrance.

A central advantage of our Bayesian penetrance estimate is the quantification of uncertainty[103]. This is expressed as a hypothetical number of heterozygotes, e.g., 7 of 10 clinically phenotyped individuals with a phenotype has a greater uncertainty than 700 of 1,000 clinically phenotyped heterozygotes with a phenotype, though the estimated penetrance is the same. To arrive at this number, we empirically test a tuning parameter, equivalent to the number of hypothetically phenotyped heterozygotes, such that 95% of the true observations (per variant empirical posterior) are contained within the EM posterior's 95% credible interval across the distribution from 0 to 1.0 using boot strap sampling. From this analysis, for *KCNQ1*-LQT1, we suggest the

Bayesian prior is empirically equivalent to the information contained in approximately 10 clinically phenotyped individuals.

In Figure 3.5, we provide a representative example of the prior and posterior penetrance estimates obtained by our approach imposed on the latter part of the S6 transmembrane segment into the C-terminus (residues 350-400). The forest plot shows prior (green, blue, and red) and posterior (black) estimates and their corresponding 95% credible intervals for variants harbored by at least five heterozygotes. The 95% credible interval of the Bayesian prior estimate is narrowed after the application of the likelihood (heterozygote status). For example, for the variant p.Ala370Val, there is a moderate Bayesian prior probability (0.28). After applying the clinical likelihood (1 affected heterozygote of 45 observed heterozygotes), the Bayesian posterior is decreased (.067) along with the respective 95% credible interval. This method provides a best

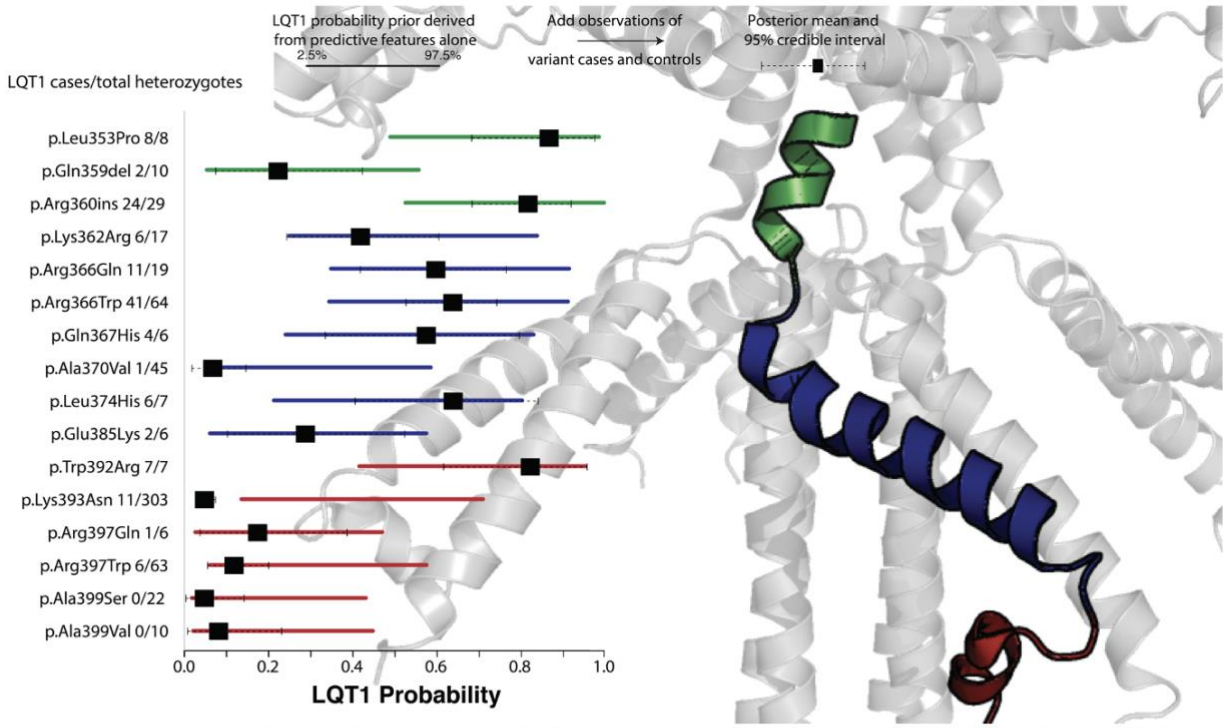


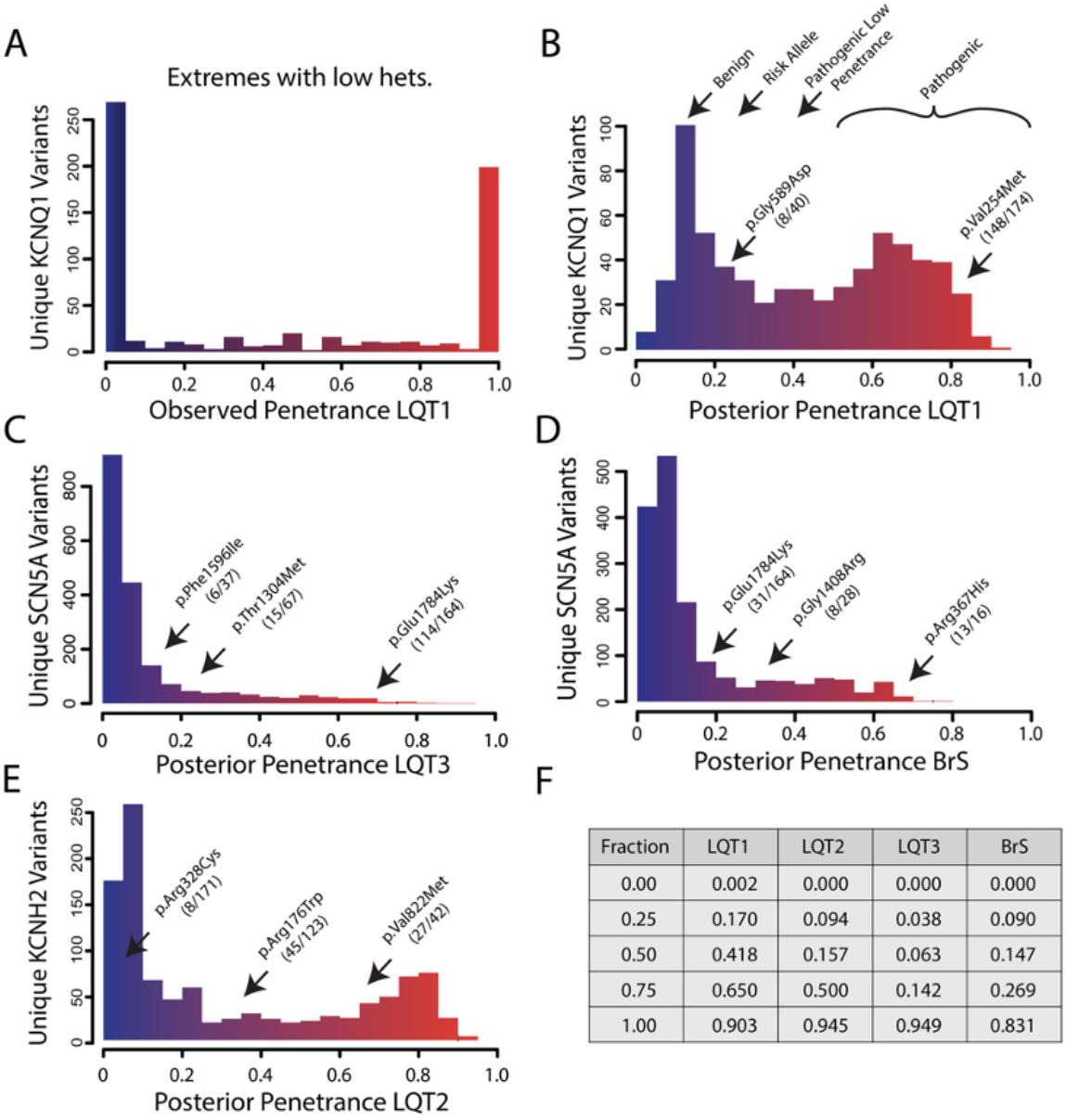
Figure 3.5. Bayesian prior and posterior point estimate probabilities and uncertainty by residue. Depiction of prior and posterior penetrance estimates for residues 350-400 (S6 transmembrane segment into the C-terminus) with >5 heterozygotes. The prior penetrance estimates and 95% credible interval are depicted in red, green, and blue, with the posterior and updated 95% credible intervals in black. Total number of observed heterozygotes are given. 95% credible interval changes proportionally with the number of clinical observations.

probability estimate for these variants, which intrinsically accounts for incomplete penetrance, even among high-penetrance variants.

Evaluation and Performance of SCN5A-LQT3 *Variant Penetrance Model.* We performed a similar analysis for the *SCN5A*-LQT3 genotype-phenotype pair. In contrast to LQT1, the LQT3 phenotype arises from a specific loss of channel inactivation, which may arise through late current, window current, or altered kinetics of inactivation[115]. We previously curated a dataset to evaluate this relationship empirically[104], which includes a total of 86,118 *SCN5A* heterozygotes with 1,243 manifesting the LQT3 phenotype. Analogous to LQT1, we updated prior penetrance estimates based on functional, structural, and *in silico* features with this clinical evidence. The posterior estimates were empirically shown to be equivalent to the information contained in 5 clinical observations of heterozygotes using the same framework described above.

Bayesian Estimates Reflects Gene-specific Penetrance Among Channelopathies. To assess differences in penetrance among the four channelopathies (LQT1-3 and BrS), we compared the Bayesian posterior penetrance estimate distributions for each gene-disease pair (Figure 3.6). Observed penetrance (strictly frequentist) tends towards 0 or 1 owing to most unique variants being observed in a single or few phenotyped heterozygotes (Figure 3.6A). The Bayesian posterior penetrance estimates are distributed over a continuous bimodal distribution (Figure 3.6B). We observe a cluster of variants with low penetrance estimates, followed by a second cluster with higher estimates which may loosely map onto the categorical ACMG framework. We interpret these estimates as indicating ‘benign’ at low estimated penetrance, followed by a continuous gradation of “risk allele’ to ‘pathogenic low penetrance’ to ‘pathogenic’ variants. Formalizing these estimates as continuous disease-presentation probabilities explicitly accounts for incomplete penetrance along this continuum. *SCN5A* variants associated with BrS and LQT3 do not display the same extent of bimodal distribution as those in *KCNQ1* and feature

comparatively more variants predicted to have very low penetrance (Figure 3.6C and 3.6D). This finding partially reflects the composition of *SCN5A,* at approximately four times the size of *KCNQ1*



Fraction	LQT1	LQT2	LQT3	BrS
0.00	0.002	0.000	0.000	0.000
0.25	0.170	0.094	0.038	0.090
0.50	0.418	0.157	0.063	0.147
0.75	0.650	0.500	0.142	0.269
1.00	0.903	0.945	0.949	0.831

Figure 3.6. Histograms of variant penetrance for Channelopathy genotype-phenotype pairs.
A) Observed penetrance frequencies for *KCNQ1*-LQT1 variants. Note bimodal distribution of penetrance frequencies in the observed penetrance.
B) Frequencies of Bayesian Posterior penetrance estimates for *KCNQ1*-LQT1. Labels indicate clinical 'gestalts' for different ranges of penetrance fractions to map ACMG classifications.
C) Bayesian Posterior frequencies for *SCN5A*-LQT3. This distribution has the lowest penetrance estimates with few very highly penetrant variants.
D) Bayesian Posterior frequencies for *SCN5A*-BrS. Covariates predict a more left-skewed distribution with a mild bimodal distribution. The highest decile of observations has less penetrance than those of LQT1 and LQT2.
E) Bayesian Posterior frequencies for *KCNH2*-LQT2. Bimodal distribution of penetrance frequencies.
F) Quantile table of distributions of posterior penetrance fractions for each genotype-phenotype pair.

with large linkers (400 residues) between membrane spanning segments largely uninvolved in channel function and correspondingly low estimated penetrance. Finally, *KCNH2* variants also have a bimodal distribution like those for *KCNQ1* variants, but with a smaller fraction predicted at high penetrance (Figure 3.6E). These observations also reflect the relative fraction of functionally relevant residues within the gene. These differences are quantified in Figure 3.6F. These distributions indicate half of identified *KCNQ1* variants have a penetrance estimate of at least 0.41, suggesting a randomly chosen *KCNQ1* variant has a 50% chance of producing a LQT1 penetrance greater than 41%. The probability distributions skew comparably lower for *KCNH2*-LQT2 and *SCN5A*-LQT3 where half of observed variants have penetrance estimates lower than 0.157 and 0.063, respectively.

Distribution of Posterior Estimates within Individual ClinVar Annotations. We next investigated how our Bayesian estimates and observed penetrance compared with ClinVar categorical designations. In Figure 3.7A, we present posterior estimates for Benign/Likely Benign (B/LB) and Pathogenic/Likely Pathogenic (P/LP) variants for *KCNQ1*. Plotting the observed penetrance vs category reveals that many variants are incompletely penetrant. We observe that the Bayesian estimates for B/LB cluster close to 0, although we show remarkable heterogeneity for estimates of the P/LP variants (ranging from 0.17-0.90). We repeated this analysis for variants within *KCNH2* (Figures 3.7C/D). We again observe distributions of Bayesian penetrance estimates (0.093-0.95) as well as a high degree of incomplete penetrance among P/LP variants. Some P/LP variants have been observed in relatively few individuals without a phenotype, and therefore have an observed penetrance of 0 (Figure 3.7A/C). Further investigation revealed that several of these variants have a P/LP ClinVar annotation which includes private communications about heterozygote clinical phenotypes (such as *KCNH2* p.Gly657Ser), which were therefore unavailable to incorporate into our model.

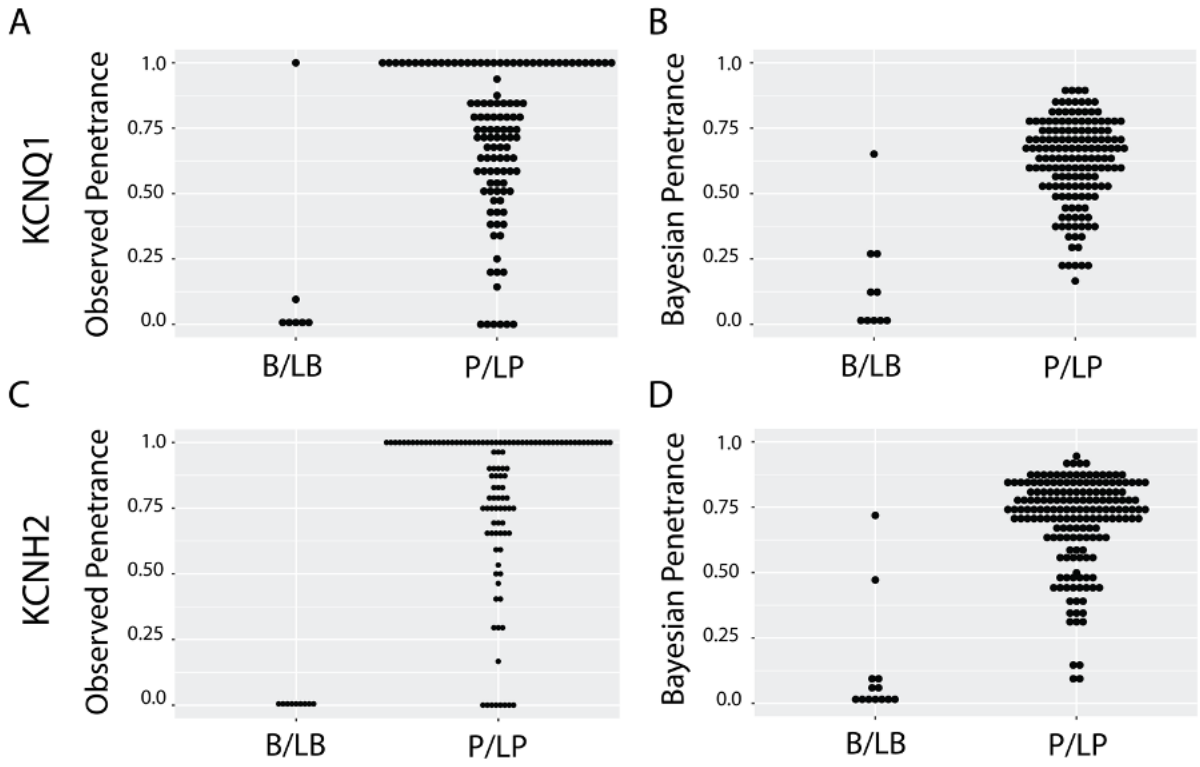


Figure 3.7. Penetrance among ClinVar-annotated variants.
A) Plotting the observed penetrance by ClinVar category shows that many *KCNQ1* P/LP variants are incompletely penetrant.
B) Most ClinVar Benign and Likely Benign *KCNQ1* variants have low Bayesian penetrance estimates. Notably, Pathogenic and Likely Pathogenic variants have a wide distribution of Bayesian penetrance.
C) P/LP *KCNH2* variants have a wide distribution of Bayesian penetrance scores in our model.
D) Several ClinVar P/LP variants within *KCNH2* are incompletely penetrant.

Structural Implications of Variable Penetrance. We mapped posterior penetrance estimates onto Cryo-EM structures (PDBIDs 5VMS, 5VA1, and 5X0M for KCNQ1, KCNH2, and SCN5A [homology modeled to include the c-terminal EF hand domain], respectively) of each protein and interrogated global and 'hot spot' penetrance heterogeneity. We observed substantial heterogeneity in penetrance across traditionally labelled 'hot spot' regions in $K_V7.1$, including the pore domain, voltage sensing domains, and subunit assembly domains (Figure 3.8A and 3.8B)[116]. Whereas several variants in these 'hot spot' domains are observed to have high posterior estimates, as indicated by the color and thickness of the putty rendering, there remain several residues with comparatively low penetrance within the same domain. Interestingly, we observe

enriched penetrance across the known binding site of calmodulin (Figure 3.8C)[110]. For example, surface residues on two helices in contact with calmodulin both have high penetrance estimates, including p.Trp379 (3 observed variants, 4 of 5 heterozygotes affected) and p.Met520Arg (4 of 4 heterozygotes affected). Structural comparisons revealed that the K_V7.1 S5-S6 domain have higher penetrance estimates than those of K_V11.1 (Figures 3.8D/E). We observe that the high

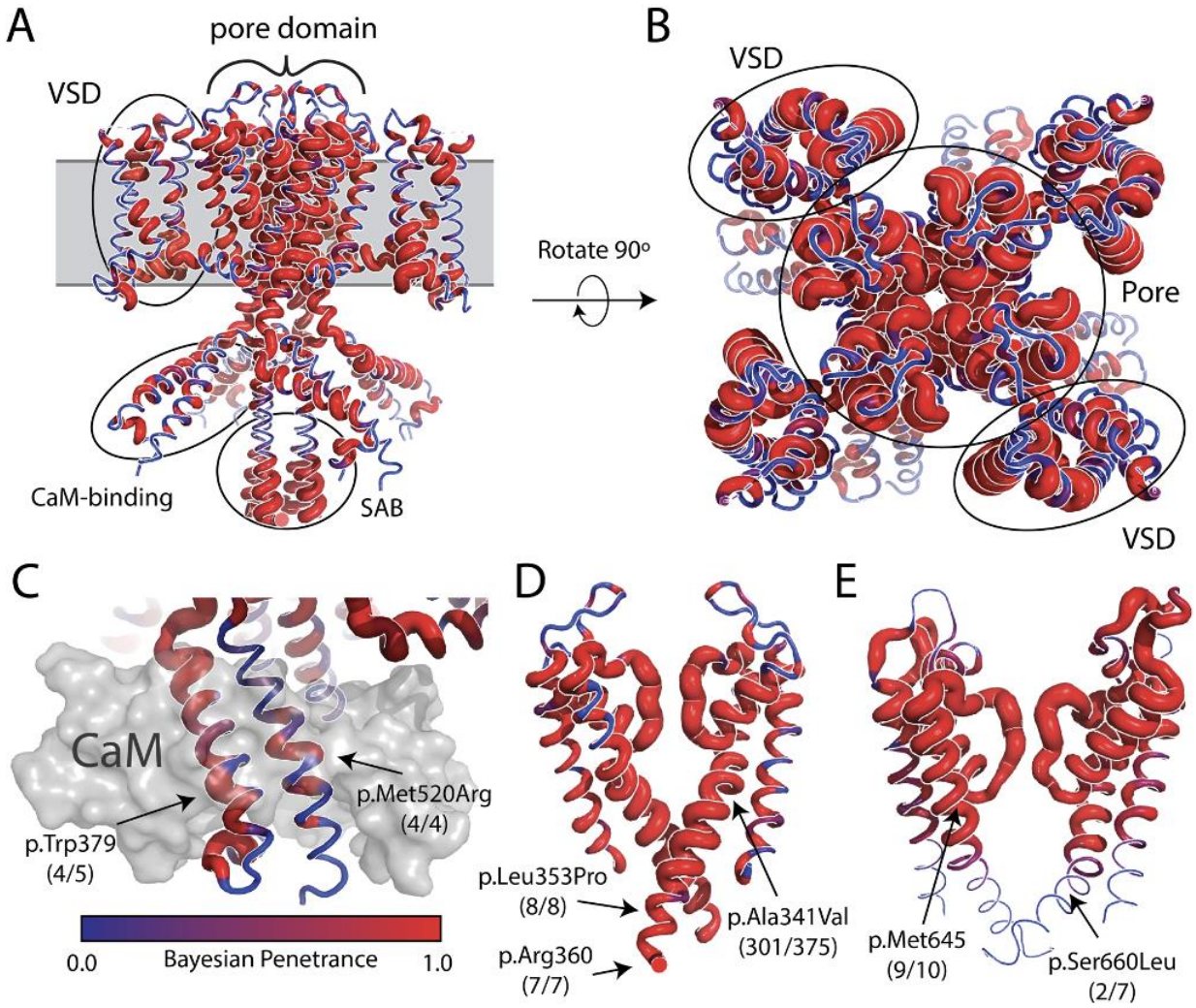


Figure 3.8. Structural Basis of Penetrance and 'Hot-Spot' Domain Heterogeneity in KCNQ1 and KCNH2.
A) Posterior penetrance estimates imposed upon the tetrameric experimental K_V7.1 structure. The subassembly domain (SAB) and pore domain have a high degree of penetrance. Calmodulin (CaM) binding domains and Voltage Sensing Domains (VSD) also exhibit high, although heterogenous, penetrance.
B) View of channel embedded in membrane from extracellular perspective. We observe very high penetrance in the channel pore and along VSD.
C) Close up of calmodulin binding domain with experimentally observed Calmodulin binding. Residues with high contact show enlarged penetrance estimates.
D) K_V7.1 S5-S6 helices with nearly uniform high penetrance estimates. Example variants provided.
E) K_V11.1 S5-S6 helices. Penetrance tapers along the S6 transmembrane domain compared to that observed for K_V7.1.

pathogenicity of variants in the S5 helix continues into the S6 helix of K_V7.1 (p.Ala344Val, 17 of 20 heterozygotes affected; p.Gly345Arg, 2 of 2 heterozygotes affected; p.Leu353Pro, 8/8 heterozygotes affected), whereas those for K_v11.1 rapidly taper (residue p.Met645 – 4 unique variants – 9/10 heterozygotes affected; p.Ser660Leu, 2/7 heterozygotes affected). This finding is consistent with a recent report implicating variants in the Kv7.1 S6 domain as particularly pathogenic, causing a significant decrease in event-free survival[117]. This may be structurally rationalized by the greater sensitivity of flexible motifs within the Kv7.1 S6 helix including the 'GSG' motif[118] (p.Ser349Trp, 2 of 2 heterozygotes affected, and p.Gly350, 3 variants, 9 of 9 heterozygotes affected).

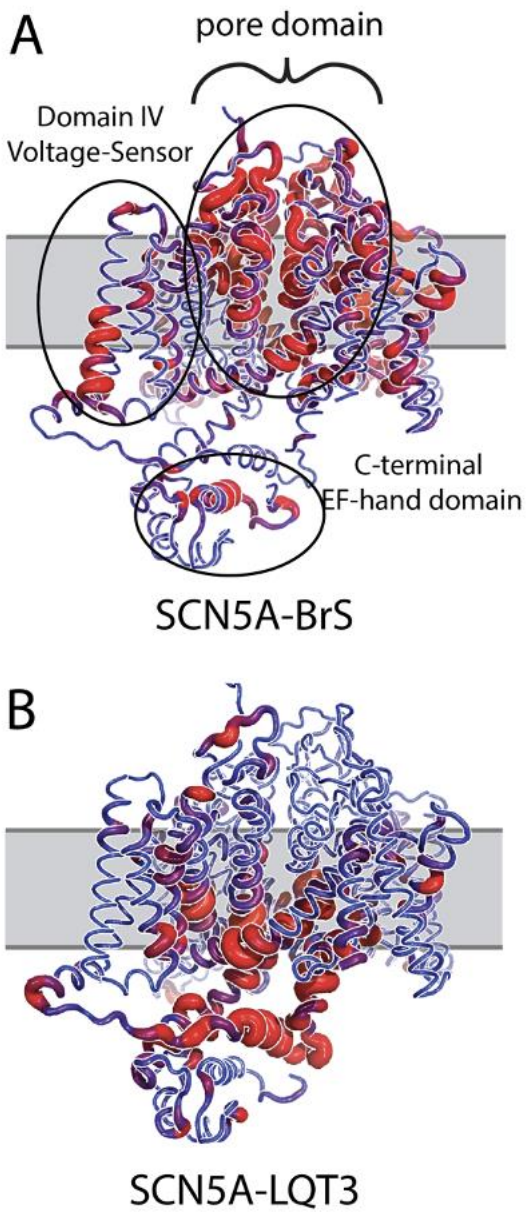


Figure 3.9 BrS and LQT3 penetrance reflected on structure.
A) Bayesian Posterior estimates for *SCN5A*-BrS. High degree of penetrance along the channel pore and voltage sensing domain (VSD).
B) Bayesian Posterior estimates for *SCN5A*-LQT3. Fewer residues show high penetrance for LQT3 compared to BrS, consistent with the molecular defects precipitating each disease. Some overlap is observed in the C-terminal EF Hand domain, whereas variants in the pore are highly tolerated compared to BrS.

We next mapped Bayesian posteriors for BrS and LQT3 onto the structure of *SCN5A* (Figure 3.9A and 3.9B). The BrS posterior estimates were highest in the pore region, especially near the selectivity filter, with some additionally high penetrance variants in the voltage sensing domains (VSDs) and the C-terminal EF-hand domain; however, even in the highly conserved pore helices, there were several BrS low-penetrance residues. These results contrast with penetrance

estimates for LQT3, where high penetrance residues were concentrated in and near the inactivation gate on the intracellular side of the channel. While some domains are associated with both phenotypes (such as the C-terminal EF-hand domain), other regions of $Na_V1.5$ are exclusively linked to BrS or LQT3 (VSD and pore for BrS and inactivation gate for LQT3). For example, of the 164 heterozygotes for the variant p.Glu1784Lys, we observed 31 with BrS and 114 with LQT3 (posterior estimates 0.19 and 0.69, respectively). This variant is in the C-terminus of the S6 helix, and recent structural studies have implicated its role in fast inactivation[119]. In contrast, the in-frame deletion p.Gln1507-1509del in the domain III-domain IV linker (inactivation-gate) was LQT3 specific (85 of 87 heterozygotes LQT3 affected, 0 BrS). Our highest BrS posterior estimate was for the variant p.Arg878Cys (32 of 36 heterozygotes affected) in the domain II pore.

Discussion

Bayesian Penetrance Estimates for Mendelian Arrhythmias. We applied a probabilistic model to generate prior and posterior penetrance estimates for the *KCNQ1*-LQT1 and *SCN5A*-LQT3 genotype-phenotype pairs that explicitly account for incomplete penetrance in the model framework. Functional, *in silico*, and structural features of these channels were significantly associated with LQT1 and LQT3 penetrance. We found that these models carried information equivalent to 10 and 5 heterozygote clinical observations, respectively. The *KCNQ1*-LQT1 model had a Bayesian prior 10-fold cross-validated Brier scores of 0.032, indicating excellent probabilistic forecasting. We then analyzed the distributions of penetrance estimates for four genotype-phenotype pairs and found *KCNQ1*-LQT1 had the largest fraction of variants with high estimated penetrance (3.6F). These continuous estimates were then mapped onto each protein structure, which revealed a structural basis of these findings as well as significant heterogeneity

in traditionally annotated 'hot spot' domains. We curated these data and host them in a convenient format for research and prospective clinical management.

Variable Penetrance Among Genotype-Phenotype Pairs. Incomplete penetrance is observed even among highly deleterious variants. For example, the *KCNQ1* variant p.Ala341Val has sharply decreased event-free survival compared to other *KCNQ1* missense variants[117], but still provokes the phenotype incompletely (301 of 375 heterozygotes manifest LQT1). Extended to all variants observed in at least one heterozygote, Figure 3.6 shows penetrance estimates for relatively few variants exhibit >90% penetrance (quantified in Figure 3.6F) and a bimodal distribution for LQT1 and LQT2 penetrance, a loosely bimodal distribution for BrS, and a taper for LQT3. The fraction of significantly high penetrance variants is consistent with the fraction of LQTS cases attributed to variants in these genes: roughly 50% of LQTS cases are attributable to variants in *KCNQ1*, 40% in *KCNH2* and 10% in *SCN5A*[120]. This trend inversely correlates with gene size: (K_V7.1 676 amino acids, K_V11.1 1159 amino acids, and Na_V1.5 2016 amino acids). For BrS and LQT3, more variants can confer loss-of-function for BrS, compared to the more limited set of possible variants that could induce a 'gain-of-function' phenotype associated with LQT3. This difference may explain the relative decrease in precision of the *SCN5A*-LQT3 estimates reflected in the lower number of phenotype observations compared to *KCNQ1*-LQT1 (5 vs 10). The predictive features are comparatively weaker due to the bias to estimate loss-of-function rather than gain-of-function as a single score. Compared to LQT1 and LQT2, BrS (also associated with loss-of-function) penetrance distribution has a lower fraction of variants with penetrance >0.5, likely due to contributions to the phenotype from other genes or polygenic modifiers. This is consistent with recent suggestions of the oligogenic nature of BrS as compared to more strictly Mendelian LQT1-3 with polygenic modifiers[121]. The complication of both arrhythmia-associated gain-of-function and loss-of-function phenotypes for missense variants in SCN5A likely

contributes to the worse correlation between SCN5A predictive features and LQT3 penetrance leading to lower hypothetical heterozygote count for SCN5A-LQT3.

Continuous Bayesian Penetrance Provides Higher Resolution Variant Interpretation. The ACMG variant classification scheme is a clinically useful categorical framework to interpret variants[43]. While this approach enables convenient approximations for clinical decision making, variants within each of these categories may nevertheless exert divergent risk due to incomplete penetrance. The *KCNQ1* variant p.Val254Met has not been observed in gnomAD and is associated with a high frequency of sudden cardiac death and syncope in Japanese heterozygotes[122]. From variant-specific features alone, we obtained a penetrance estimate of 0.65 and a posterior LQT1 penetrance estimate of 0.84. In contrast, the disease-associated variant p.Gly589Asp has a higher allele-frequency (4.96e-5 in gnomAD) than expected for a highly pathogenic variant (2.5e-5 rare allele cutoff[82]), despite being annotated as Pathogenic in ClinVar[45]. Accordingly, our model predicted a prior penetrance estimate of 0.37, and a posterior estimate of 0.23 after incorporating clinical data (8 of 40 affected heterozygotes). Such a variant clearly carries risk above the baseline population rate, but not as high as that conferred by p.Val254Met. A strength of our approach is therefore the ability to capture this granular level of detail, whereas classification methods lose this nuance. Our continuous estimates suggest that certain variants may be considered 'modifiers' or exert subtle impact, consistent with the terminology of "LQT-lite" proposed by Ackerman and colleagues[123-125]. This idea is additionally consistent with recent work showing variable environment interactions in the presence of common *SCN5A* variants, where the variant may act as a 'modifier' of penetrance[126].

The divergence of our approach, leveraging information contained within the variant-specific features, can also be appreciated when contrasted with ClinVar categorical descriptions (Figure

3.7). ClinVar annotated P/LP variants display a range of predicted penetrance estimates (0.167-0.903 *KCNQ1* and 0.0930-0.945 *KCNH2*); the effect is also evident from non-Bayesian penetrance observations (Figure 3.7A and 3.7C). Furthermore, our findings in Figure 3.7A show that observed penetrance is not sufficient for forecasting penetrance, that most unique variants are observed in very few people leading to too many LQTS penetrance estimates of 0 or 1. Variant-specific features modify these crude estimates from low numbers of phenotyped variants to shift estimates away from the extremes to more intuitive and biologically reasonable estimates (Figure 3.7B). Collectively, these observations support the notion that our continuous Bayesian estimates offer a higher resolution alternative to categorical classification.

Penetrance Heterogeneity within Structured Functional Domains. Previous works have implicated 'hot spot' regions within the three protein products based on collapsing of variants in cases and controls; however, many of these studies preceded recent large-scale population genome sequencing efforts[35,116]. Our combination of genetic and structural data shows variable penetrance throughout many traditional 'hot spot' domains (Figure 3.8 and Figure 3.9). For example, not all residues in the *SCN5A* domain IV VSD or pore are associated with high BrS penetrance (Figure 3.9A). These differences are further rationalized between the *SCN5A*-BrS and *SCN5A*-LQT3 pairs, where different electrophysiological parameters correspond to discrete structural features. Even for higher penetrance variants within 'hot spots', many provoke the phenotype incompletely. In *KCNH2,* the ClinVar Pathogenic variant p.Val822Met is located in the pore domain 'hot spot' but affects only 27 of 42 heterozygotes with a Bayesian penetrance estimate of 0.65. We also observe heterogeneity between homologous S6 segments in K_V7.1 and K_V11.1, with a much higher LQTS penetrance throughout the helix in the former. These results highlight important structural differences in these proteins that contribute to clinical penetrance, despite similar channel characteristics (both being voltage-gated potassium channels).

Altogether, these data suggest using variant location in 'hot spot' domains in ACMG variant classifications may be premature[70]. For instance, recent work from our lab performed a Deep Mutational Scan of a 'hot spot' exon in KCNH2, and found experimentally that only 42% of all possible amino acid substitutions substantially decreased trafficking[54]. We view the current approach as providing higher resolution interpretation in these domains – this work provides more information than categorical classification, or yes/no 'hot spot' annotations.

Mapping penetrance to structure also provides unique insights into the underlying channel biology. For example, K_V7.1 residues 364-383 and 510-533 form the A and B helices, respectively, with several high-penetrance sites, including non-buried residues which interact with calmodulin (Figure 3.8C). Calmodulin is global calcium sensor within the cardiomyocyte and is important for regulating excitation-contraction coupling[127]. The functional implications of Calmodulin binding K_V7.1 are speculative, however, a recent study implicated calmodulin as a state-dependent switch to modulate channel gating[128]. Our study suggests a high LQT1 penetrance when these interactions are abrogated by genetic variation, likely being mediated through this loss of gating modulation. These results provide a complementary set of data to implicate loss of the Calmodulin-K_V7.1 signaling axis as an etiology for LQT1.

Comparison with Other Approaches. Recent studies in the cardiac space have leveraged machine learning to facilitate variant reclassification[101,102]. Approaches such as GENESIS[102] and MLb-LDLr[101] make progress towards this goal but rely on ClinVar data as a critical input. The approach we take relies on classic regression strategies and expectation-maximization to integrate data, and instead draws on all available variant-specific features to output a *continuous* penetrance estimate calibrated to curated, granular clinical data. Rather than focusing on classification, other interesting approaches have instead focused primarily on functional impact, such as prediction of

functional properties in the presence of missense variants[129,130]. The use of granular clinical data, each heterozygote evaluated as above or below a clinically meaningful threshold represents an advance by which we can assert not only a penetrance estimate, but also the uncertainty of the estimate – an approach well-tailored to incompletely penetrant BrS and LQTS variants.

Precision Medicine Applications. An increasingly important goal of precision medicine is to enable 'genetics first approaches', wherein sequencing of unascertained clinical cohorts may prioritize individuals for tailored follow-up[50,131]; however, prioritizing variant heterozygotes for follow up remains difficult. Additionally, sequencing of large population cohorts has revealed an unexpectedly large frequency of missense variants in genes associated with Mendelian disease[77,132,133]. It is unlikely there will ever be sufficient discovered heterozygotes for clinical phenotyping to adjudicate the clinical consequence of all possible single nucleotide variants in Mendelian disease genes[134]. It is therefore imperative to maximally leverage all variant-specific features to help facilitate an understanding of clinical consequence[52]. As many phenotypes demonstrate variable penetrance, our method provides an initial gestalt from variant-specific features which help tailor prospective precision medicine in concert with other clinical observations. To calibrate the strength of these features, we provide an empirical equivalent of heterozygote observations, an interpretable form of these data. Based on this progress, we envision a future in which genetic testing forms a pre-test probability of disease which is updated by clinical phenotyping. This schematic would represent an inversion of the current use of genetics in medical practice, in which a genetic test most commonly modifies a clinical pre-test probability.

Limitations. One limitation of our study is the ascertainment bias of highly penetrant variants in clinical practice. Many asymptomatic heterozygotes will not be clinically phenotyped or adequately sampled in current population sequencing studies. This bias is additionally reflected in the availability of functional data, as most studies have interrogated variants with a high suspicion of pathogenicity; however, this bias is increasingly tempered by functional evaluation of less unascertained variants such as those in a recent study of arrhythmia genes in the eMERGE network[51]. We are currently unable to satisfactorily account for variant combinations such as compound heterozygotes[135]. Data availability remains a limitation for model construction, as several ClinVar annotations were based off 'personal communications' which resulted in low or zero observed penetrance among P/LP variants in select cases.

Conclusion

Incomplete penetrance demands nuance to variant interpretation for the realization of genome-first precision medicine approaches[50]. We make progress towards this goal by elaborating a probabilistic approach that considers each variant as having a continuous penetrance probability and provides higher resolution in 'hot spot' domains to facilitate prospective clinical management. This approach provides more information than classification and explicitly treats incomplete penetrance in model construction. We render variant-specific features as equivalents to clinically-phenotyped heterozygotes to democratize these data to a broader audience without electrophysiology and structural biology expertise.

Chapter 4

Functional Assays Reclassify Suspected Splice-Altering Variants of Uncertain Significance in Mendelian Channelopathies

Portions of this chapter are published under the same title in *Genetics in Medicine*

Background: Rare protein-altering variants in *SCN5A*, *KCNQ1*, and *KCNH2* are major causes of Brugada Syndrome (BrS) and the congenital Long QT Syndrome (LQTS). While splice-altering variants lying outside 2-bp canonical splice sites can cause these diseases, their role remains poorly described. We implemented two functional assays to assess 12 recently reported putative splice-altering variants of uncertain significance (VUS) and 1 likely pathogenic (LP) variant without functional data observed in BrS and LQTS probands.

Methods: We deployed minigene assays to assess the splicing consequences of 10 variants. Three variants incompatible with the minigene approach were introduced into control induced pluripotent stem cells (iPSCs) by CRISPR genome editing. We differentiated cells into iPSC-derived cardiomyocytes (iPSC-CMs) and studied splicing outcomes by reverse transcription-polymerase chain reaction (RT-PCR). We used the American College of Medical Genetics and Genomics functional assay criteria (PS3/BS3) to reclassify variants.

Results: We identified aberrant splicing, with presumed disruption of protein sequence, in 8/10 variants studied using the minigene assay and 1/3 studied in iPSC-CMs. We reclassified 8 VUS to LP, 1 VUS to Likely Benign, and 1 LP variant to pathogenic.

Conclusions: Functional assays reclassified splice-altering variants outside canonical splice sites in BrS- and LQTS-associated genes.

Introduction

The arrhythmia syndromes Brugada Syndrome (BrS) and congenital Long QT Syndrome (LQTS) are rare autosomal dominant Mendelian diseases, mainly involving variants in cardiac ion channels. BrS is associated with loss-of-function (LoF) variants in the *SCN5A* sodium channel gene in 20% of patients, while LoF variants in the potassium channel genes *KCNQ1* and *KCNH2* and gain-of-function (GoF) variants in *SCN5A* are found in 80% of patients with congenital LQTS [30,36]. These diseases contribute to the >250,000 cases of sudden cardiac death (SCD) each year in the US through fatal ventricular arrhythmias. When affected heterozygotes are recognized early, BrS and LQTS can be clinically managed through medications and/or interventional therapies. However, when not recognized through medical or genetic screening, SCD may present as the sentinel disease manifestation[26]. Genetic sequencing therefore offers the opportunity to uncover risk in 1) clinically unrecognized heterozygotes identified in unascertained population sequencing studies, or 2) family members of a recognized proband. Multiple approaches are in development to support genome-first approaches and cascade screening, but these efforts are still hampered by issues of clinical interpretability[50,109]. The American College of Medical Genetics and Genomics (ACMG) has provided a framework for variant interpretation, spanning benign (B) to pathogenic (P) based on criteria including variant functional evidence, population minor allele frequencies, segregation within families, and computational predictions, among others[43]. While panel sequencing and whole exome/genome sequencing are improving our broader understanding of the genetic basis of disease, a large proportion of detected variants are classified as Variants of Uncertain Significance (VUS) and are therefore not clinically actionable[77].

Protein-altering variants (including nonsense, frameshift, and missense variants) are a major focus of clinical attention. Variants affecting RNA splicing occupy a comparatively less

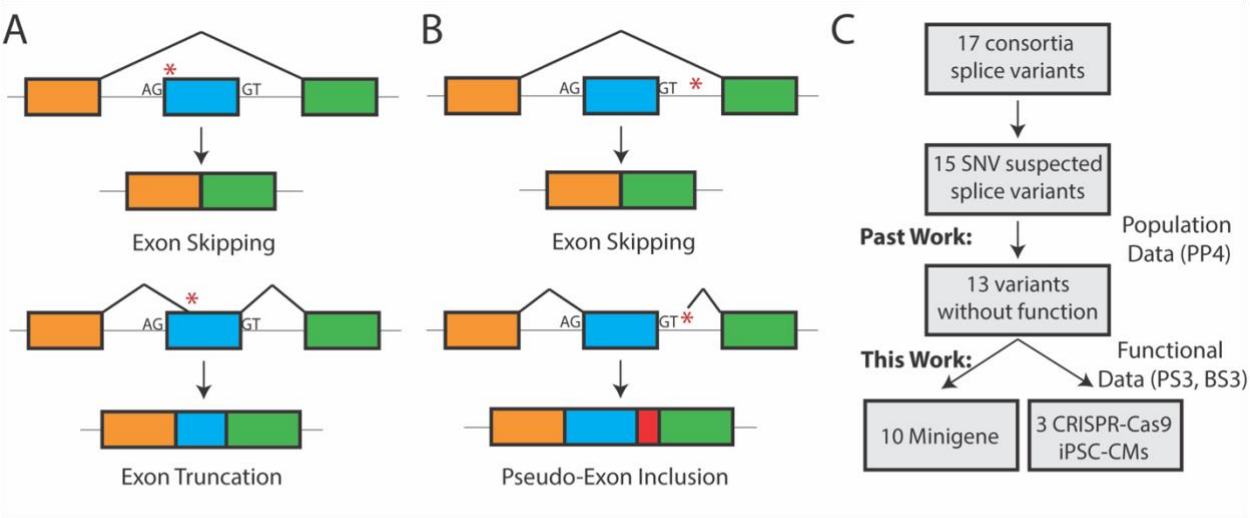


Figure 4.1. Aberrant splicing from *cis*-genetic variation and previous variant prioritization.
A) Variants in an exon near the 2-bp splice acceptor (AG) or donor (GT) can disrupt the recognition of the canonical site and lead to exon skipping. Alternatively, variants within the exon may introduce a new splice acceptor or donor site and lead to truncation of the exon, which may introduce frameshifts or damage protein function. This schematic features canonical AG-GT splice sites; rare alternate splice sites are possible.
B) Intronic variants adjacent to the 2-bp splice acceptor or donor may disrupt spliceosome recognition of the canonical splice site and lead to exon skipping. Deeper intronic variants may introduce cryptic splice acceptor or donor sites that may lead to transcription of intronic regions (pseudo-exons) that may disrupt the reading frame or compromise protein function. This schematic features canonical AG-GT splice sites; rare alternate splice sites are possible.
C) A recent variant interpretation effort curated LQTS and BrS heterozygotes and reported variants. 17 putative splice-altering variants were reported, of which 15 were Single Nucleotide Variants. Of these, 13 were previously not functionally characterized. We analyzed 10 of 13 using a standard minigene approach and 3 using RT-PCR assays from iPSC-CMs.

explored fraction of the genome, despite estimates that they contribute to up to 10% of pathogenic variants in Mendelian diseases[64]. Aberrant splicing from rare variants that alter the 2-bp canonical splice sites flanking each exon is typically characterized as meeting the PVS1 criterion, and therefore often these variants are often classified as P/LP. However, rare splice-altering variants falling outside of these 2-bp sites are more difficult to interpret. Aberrant splicing may arise from more distant intronic or exonic variation by introducing or ablating splice acceptors or donor sequences or by affecting regulatory splicing enhancers or silencers (Figure 4.1)[136]. *In silico* predictors of splicing consequences have historically had only modest predictive ability.[137,138] However, recent tools leveraging advances in machine learning and large RNA-seq datasets raise the possibility that *in silico* splicing predictors may be increasingly used to facilitate interpretation of suspected splice-altering variation[64].

In this study, we used minigene assays in Human Embryonic Kidney 293 (HEK293) cells to study variant consequences on splicing for 10 putative splice-altering variants in arrhythmia

genes identified in a recent cohort of BrS and LQTS patients (Figure 4.1C). For three variants incompatible with minigene assays, we examined the impact on splicing in induced pluripotent stem cell-derived cardiomyocytes (iPSC-CMs) that were edited with CRISPR-Cas9[139]. We applied these functional splicing assays to reclassify a total of 10 putative splice-altering variants within the ACMG framework, including 8 VUS to Likely Pathogenic.

Methods

Selection of Variants. Putative splice-altering variants were identified from a recent curation of variants from Walsh et al.[17] Variants were considered putative splice-altering variants if they were identified in at least 1 case of BrS (*SCN5A*) or LQTS (*KCNQ1* and *KCNH2*), and were not predicted to affect the coding sequence of the protein. Additionally, all candidate variants had an allele frequency lower than 2.5e-5 in the Genome Aggregation Database (gnomAD)[7], a cutoff derived from theoretical predictions and empirical measurements of maximum allele frequency for Mendelian arrhythmia variants[17, 29]. Variant case[17] and gnomAD[7] counts are shown in Table 4.1. We used the following transcripts throughout our study: ENST00000155840 (*KCNQ1*), ENST00000262186 (*KCNH2*), and ENST00000333535 (*SCN5A*).

Interpretation of Functional Assays. For minigene assays, we performed triplicate amplification of RT-PCR products and quantified the fraction of Percent Spliced In (PSI) of the wild-type (WT) exon cassette from gel intensities using ImageJ[41]. All bands were confirmed by Sanger sequencing. We deemed minigene results inconclusive if the PSI of the Sanger sequence-confirmed canonical exon included product in the control (WT) plasmid was less than 30% of the total band intensity for that lane. We considered a >50% disruption of WT splicing to fulfill the PS3 criteria, while the absence of splice perturbation (<10% change in PSI compared to WT) fulfilled the BS3 criteria. Statistical analyses were performed with a non-parametric independent 2-group Mann-Whitney U Test implemented in Stata version 17.0.

Minigene constructs. A schematic of the minigene construct is shown in Figure 4.2A. PCR products bearing the exon of interest surrounded by 100 nucleotides of intronic DNA were amplified from healthy control gDNA. PCR products were gel extracted (Qiagen) then restriction digested with SalI and NotI (New England Biolabs). The plasmid pET01 (MoBiTec GmbH, Göttingen, Germany) similarly underwent double digestion with SalI and NotI, followed by

treatment with Calf Intestinal Phosphatase (Promega). The digested products were ligated together with T4 ligase (New England Biolabs). An aliquot of the ligation mix was transformed into DH5α competent cells by heat shock (Thermo Scientific) and cells were grown overnight at 37°C incubated with LB broth and 1 mg/mL ampicillin. Plasmid DNA was extracted using the Qiagen Miniprep kit. Each variant of interest was introduced using the QuikChange Lightning Multi Kit (Agilent) with 1 primer per variant using primers designed using the online QuikChange Primer Design tool. The sequences of wild-type and variant plasmids were confirmed using Sanger sequencing and the entire insert sequences were compared to the GrCh38 reference genome sequence to ensure there were no unanticipated variants present.

Cell Culture and Transfection. HEK cells were cultured at 37°C in humidified 95% air/5% CO_2 incubator in "HEK media": Dulbecco's Eagle's medium supplemented with 10% fetal bovine serum, 1% non-essential amino acids, and 1% penicillin/streptomycin. At 30% confluency, HEK cells were transfected with 500 ng of plasmid using FuGENE 6 (Promega) following manufacturer's instructions.

RNA Isolation and Reverse Transcription-Polymerase Chain Reaction (RT-PCR). 48 hours after transfection, HEK cells were washed with PBS, treated with Trypsin for 2 minutes, and harvested with HEK media. Following centrifugation at 300g for 5 minutes, the supernatant was aspirated, and the pellet was immediately placed on ice. RNA was extracted from the pellet using the Qiagen RNeasy Minikit. Reverse transcription was performed using SuperScript® III System (Invitrogen) using a primer specific to pET01. PCR was performed using GoTaq PCR Master Mix (Promega). For exon skipping events, a different primer was used to confirm the splice junction by Sanger sequencing. DNA was amplified using a touchdown protocol: 98°C for 30", (98°C for 10s, 65°C

for 30s [decreasing 1°C/cycle], 72°C for 60s) x 10 cycles, (98°C for 10s, 55°C for 15s, 72°C for 60s) x 30 cycles.

Gel Electrophoresis and Extraction of RT-PCR Products. RT-PCR products were separated by gel electrophoresis using 2% agarose gels in TAE buffer. Bands were visualized with UV light and excised followed by DNA isolation (Qiagen Gel Extraction kit). Extracted DNA was analyzed by Sanger sequencing.

CRISPR-Cas9. Primers for CRISPR gRNAs were designed using the online CRISPOR tool[42]. Guides were cloned into SpCas9-2A-GFP (pX458)[20] plasmid (Addgene #48138, a gift of Feng Zhang). The guide plasmid and a 150 bp single-stranded homology directed repair oligonucleotide template with the variant of interest and a PAM site variant to disrupt re-cutting were co-electroporated into control iPSCs from a healthy donor using a Neon Transfection System (ThermoFisher MPK5000)[43]. PAM sites were designed to introduce a synonymous exonic variant when possible or an intronic variant within 40 bp of the introduced primary variant. In all cases, the SpliceAI score for the PAM disrupting variant was <0.15 for each predicted splice change. After 48 h, GFP+ single cells were sorted on a BD Fortessa 5-laser instrument and sub-cloned. Each sub-clone was genotyped by variant-specific primers. Sanger sequencing of amplified DNA identified clones with the correct heterozygous variant.

Cardiac Differentiation of iPSC. iPSCs were maintained on plates coated with Matrigel (BD Biosciences) in mTeSR plus media (STEMCELL) at 37°C in a humidified 95% air/5% CO_2 incubator. At 60-80% confluency, iPSCs were differentiated into iPSC-CMs using a monolayer chemical method as previously described[44].

RNA isolation and RT-PCR. iPSC-CMs were harvested at day 30 of differentiation following the RNeasy Minikit protocol (Qiagen). RT-PCR was performed as above with variant-specific primers. The amplified cDNA was separated using 2% agarose gel electrophoresis, extracted as above, and analyzed with Sanger sequencing. For TA cloning, the gel extracted product was ligated into the pGEM®-T Vector (Promega) using the supplied kit per manufacturer's instructions. After ligation and transformation, single colonies were grown up, DNA was extracted with the Qiagen Miniprep kit as above, and transcript composition determined by Sanger sequencing.

SpliceAI. The *in silico* predictor SpliceAI was used to predict the splicing consequences of each assayed variant. Variants were manually interrogated using the web-based interface hosted at https://spliceailookup.broadinstitute.org/ using GRCh38 coordinates and a maximum distance of 1000 bp. For each variant, four predicted probabilities were obtained: probabilities of introducing a splice Acceptor Gain (AG), Acceptor Loss (AL), Donor Gain (DG), or Donor Loss (DL) accompanied by a predicted position of such change relative to each variant. We computed an aggregate SpliceAI score to incorporate contributions from each of the 4 predicted categories using a previously described formula[45]:

$$P(\text{aberrant splicing}) = 1 - ((1-(AG)) * (1-(AL)) * (1-(DG)) * (1-(DL))$$

We considered a high likelihood of disrupting splicing to be greater than 0.80, and a low likelihood less than 0.25.

ACMG Classification of Variants. There has been extensive use and validation of this minigene assay in our study and previous studies[13-15,46,47] and there is a clear link between heterozygous LoF variants of the 3 studied genes with Long QT Syndrome (*KCNQ1* and *KCNH2*) or Brugada Syndrome (*SCN5A*)[1,2]. Therefore, we considered a conclusive assay result (as defined above) that resulted in >50% disruption of splicing to fulfill the PS3 criterion, while the absence of splice perturbation fulfilled the BS3 criterion. We applied the PP3 criterion (computational prediction

shows deleterious effect) to variants with an aggregate SpliceAI score >0.8 and the BP4 criterion (computational prediction shows no deleterious effect) to variants with an aggregate SpliceAI score <0.25. We supplemented the previously published ACMG criteria from Walsh et al. (Table 4.1) with these functional assays and computational findings for reclassification, using the University of Maryland ACMG online tool to implement the criteria[17,48].

Results

Minigene Assays Reveals Aberrant Splicing. We used a minigene construct previously deployed by our group[140] and others[141,142] to study the effects of *SCN5A* potential splice-altering variants that have been observed in one or more patients with BrS (Figure 4.2A). This assay allows the direct comparison of WT versus variant splicing outcomes by studying a specific exon and flanking intronic regions of interest. We first studied the known likely pathogenic (LP) exon-truncating *SCN5A* synonymous variant c.4719C>T in our system. This variant was previously shown to disrupt splicing in both a minigene assay and RNA isolated from patient peripheral blood lymphocytes[143]. Compared to the respective WT minigene construct, the size of the RT-PCR product from the c.4719C>T minigene was consistent with cDNA truncation (Figure 4.2B). This was confirmed by Sanger sequencing (Figure 4.2C). All major bands observed by gel electrophoresis were subjected to characterization by Sanger sequencing. For the bands that could be unambiguously identified, we inferred the splice product. For each major band, we inferred the splicing consequence based on reconstruction of the minigene vector and native genomic sequence. We next studied the *SCN5A* VUS c.1890G>A located at the terminal coding nucleotide position of exon 12 and its WT counterpart. In the minigene assay, c.1890G>A resulted in retention of intron 10 by a splice donor loss as evidenced by RT-PCR product gel band size and confirmatory Sanger sequencing (Figures 4.2D and 4.2E). Quantification of the gel band intensity showed significant changes in the WT exon cassette percent spliced in (PSI) in both cases (Figure 4.2F). The effects on transcript composition of these two variants are shown schematically in Figure 4.2G. This approach was attempted for two additional *SCN5A* VUS, c.4299G>C and c.4299+6T>C, but the results of the minigene assay were inconclusive due to poor splicing in of the WT exon.

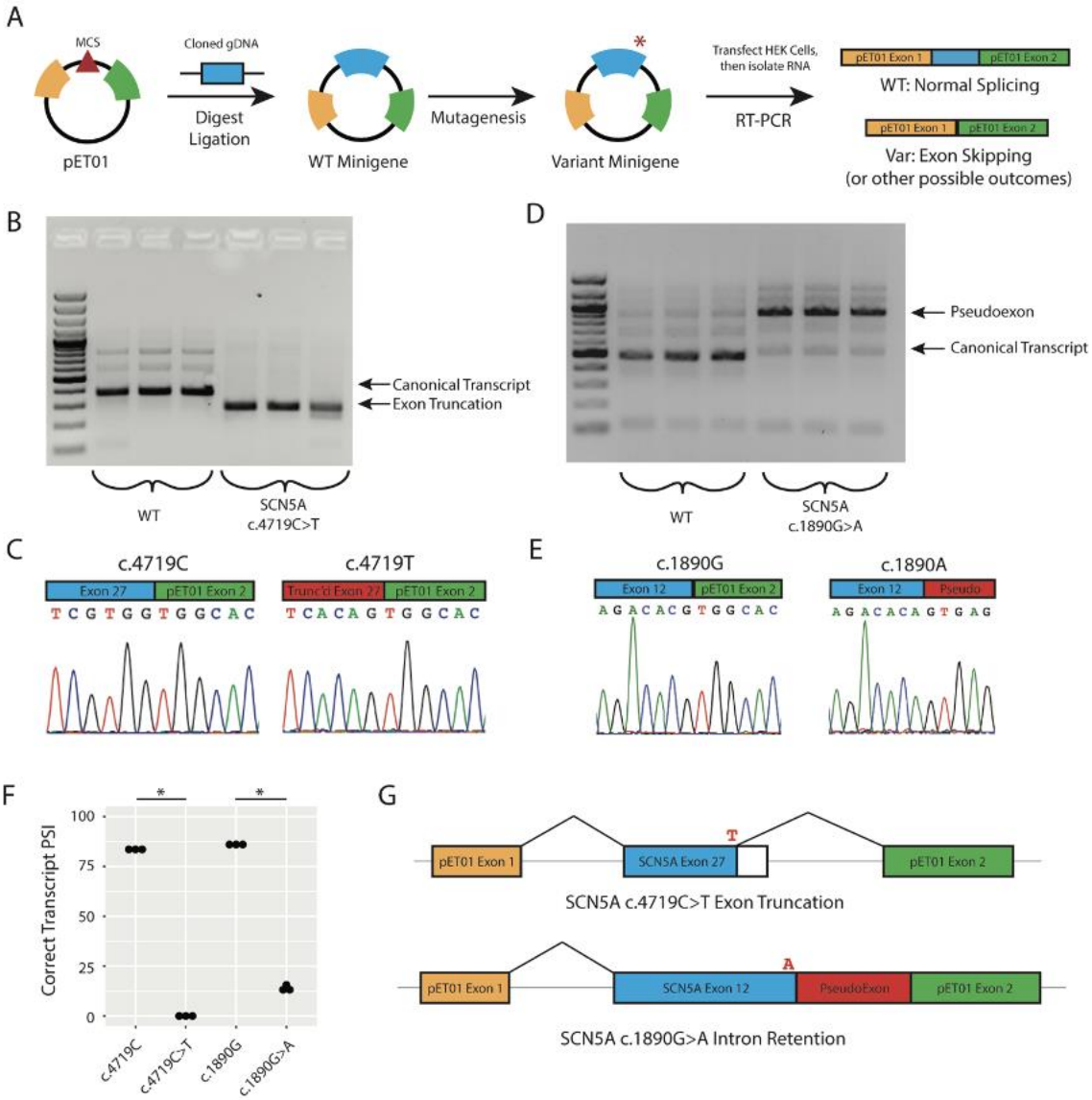


Figure 4.2. Minigene assay and studies on *SCN5A*.
A) Schematic of minigene assay. The PCR product of an exon and segments of flanking introns of interest is cloned into a multiple cloning site (MCS) between two known exons. Mutagenesis inserts the variant. After transfection of the WT and variant plasmids into HEK cells, RNA is isolated, and RT-PCR provides cDNA for analysis of transcript composition.
B) Gel of PCR products for the known exon truncating variant c.4719C>T and WT plasmid. The WT band is consistent with higher molecular weight compared to the truncated form induced by the aberrant splice donor gain. All minigene assays are presented in triplicate.
C) Sanger sequencing of the exon-exon junctions of WT and truncated RT-PCR products.
D) Gel of the WT exon 12 and c.1890G>A RT-PCR products implicating a pseudo-exon gain induced by the variant.
E) Confirmatory Sanger sequencing of RT-PCR products for the WT and c.1890G>A VUS.
F) Quantification of WT and variant Percent Spliced In (PSI) for 3 independent replicates, respectively. * indicates $p < 0.05$, independent 2-group Mann-Whitney U Test.
G) Schematic of assay results. *SCN5A* c.4719C>T truncates exon in our assay, whereas the *SCN5A* variant c.1890G>A induces intron retention.

We studied three *KCNH2* VUS c.1128G>A, c.2145G>A, and c.2398+5G>T discovered in the LQTS cohort[70]. The variant c.1128G>A induced exon skipping, which is predicted to disrupt

the downstream reading frame based on exon size. The variant c.2145G>A led to an exon skipping event at exon 8, similarly resulting in a predicted frameshift. Both of these nucleotides are located at the last nucleotide position of the coding exon. The intronic variant c.2398+5G>T behaved similarly, leading to exon skipping. This exon 9 skipping event is also predicted to lead to a frameshift. Quantification of the respective WT exon cassette PSI showed significant changes in all cases ($p < 0.05$, Mann-Whitney U Test) (Figure 4.3A). Schematic depictions of the major splicing outcomes on the transcript are shown in Figure 4.3B.

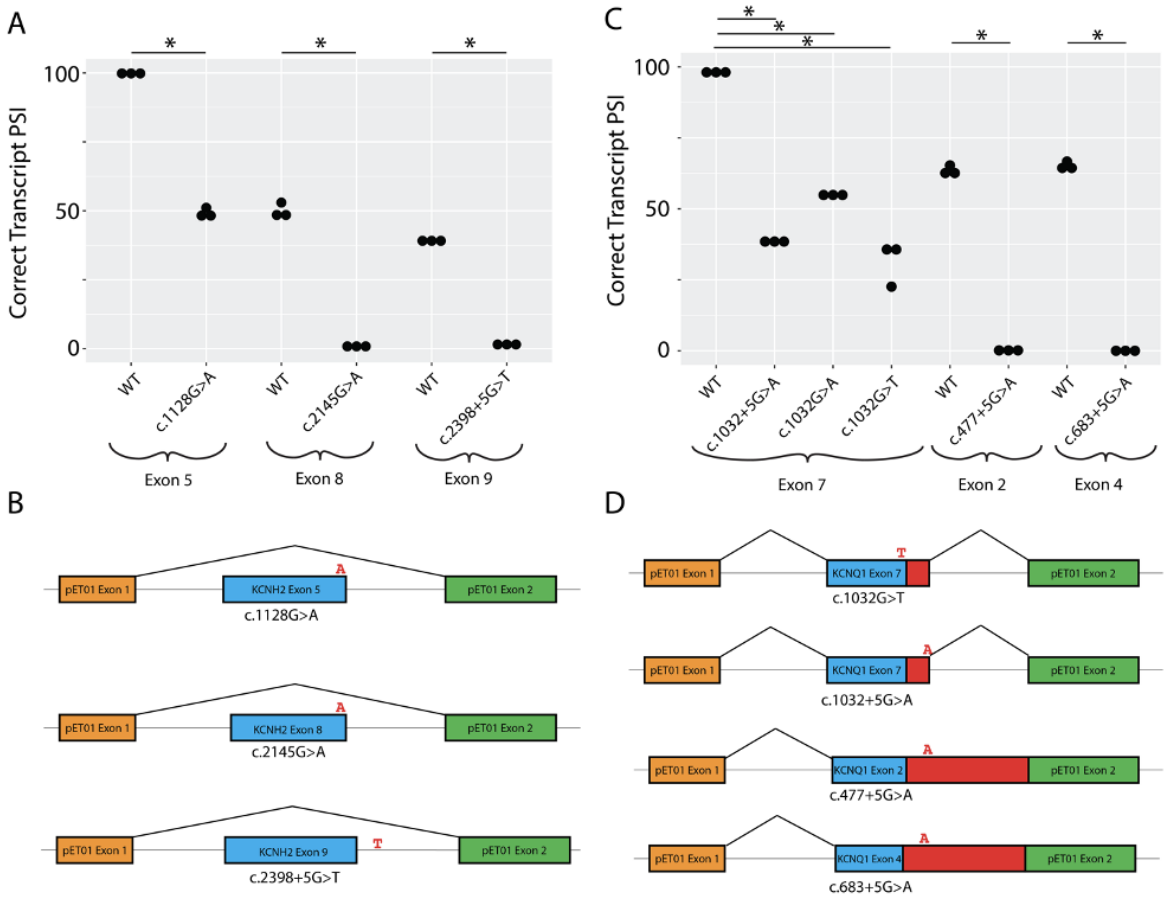


Figure 4.3. Minigene analysis of *KCNH2* and *KCNQ1* splice variants.
A) Quantification of *KCNH2* band intensity corresponding to the PSI of the WT and variant bands. 3 independent replicates per group (* indicates $p < 0.05$, independent 2-group Mann-Whitney U Test).
B) Schematic depiction of the splicing outcome associated with the highest intensity band for each *KCNH2* variant.
C) Quantification of normal exon PSI across all samples. 3 independent replicates per group (* indicates $p < 0.05$, independent 2-group Mann-Whitney U Test).
D) Schematic depictions of the major splicing outcome for each variant.

We studied 5 candidate LQTS splicing variants in *KCNQ1*. 3 *KCNQ1* VUS (c.683+5G>A, c.1032+5G>A, c.1032G>T) and 2 *KCNQ1* LP variants (c.477+5G>A and c.1032G>A). The c.1032

position is the terminal coding position of this exon and is a hotspot for variation. As a control, we first studied c.1032G>A, one of the most common causes of LQTS among Japanese probands

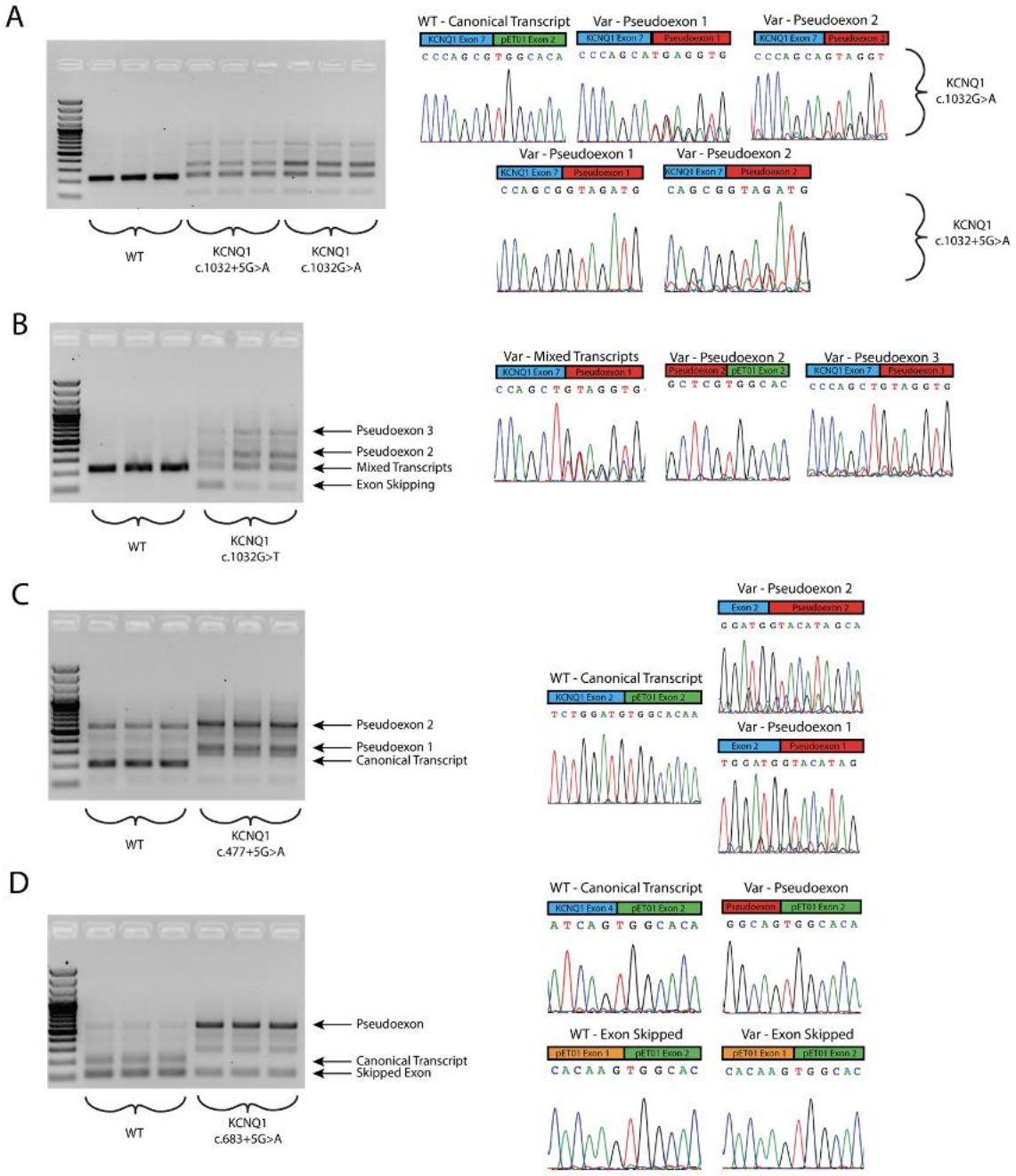


Figure 4.4. *KCNQ1* Variant Minigene Assay Gels and Sanger Traces.
A) Gel analysis of *KCNQ1* c.1032G>A and c.1032+5G>A variants. Each variant introduces multiple pseudoexons profiled by Sanger sequencing.
B) The c.1032G>T VUS introduces similar pseudoexons, while inducing a higher degree of exon skipping than that observed in c.1032G>A.
C) Gel electrophoresis of RT-PCR products show that the LP variant c.477+5G>A was observed to introduce at least 2 discrete pseudoexons and ablate WT splicing.
D) The VUS c.683+5G>A variant introduces a major pseudoexon while also introducing a degree of exon skipping.

(29 cases, 0 gnomAD controls).[63] This variant has been previously shown to affect splicing *in vitro* and did so in our assay as well, leading to 2 alternative pseudoexon inclusions (Figure 4.4A). The adjacent intronic VUS, c.1032+5G>A, had a similar gel electrophoresis profile and led to multiple pseudoexon inclusions with variable respective cryptic donor sites. A second VUS near this splice junction, c.1032G>T, was also functionally assayed. In this case, 3 pseudoexons and an exon skipping event were observed, with a small amount of WT transcript overlapping with a pseudoexon (Figure 4.4B). These consequences are predicted to have similarly deleterious effects on the as those previously studied for c.1032G>A[63]. These results show that the *KCNQ1* exon 9 splice donor site is highly intolerant to variation. The previously unstudied LP variant, c.477+5G>A, was assayed and showed nearly complete loss of WT splicing, with inclusion of 2 unambiguously defined pseudoexons (Figure 4.4C). The retained intron sequence contains a stop codon that would disrupt protein function. In our functional studies, introduction of the VUS c.683+5G>A variant completely abrogated WT splicing, while also introducing a pseudoexon and retaining an exon skipping event moderately observed in the WT (Figure 4.4D). The retained intronic sequence would also lead to a stop codon inclusion, which is predicted to disrupt protein function. Quantifications of splicing outcomes for the WT PSI are shown in Figure 4.3C ($p < 0.05$, Mann-Whitney U Test). Major splicing outcomes for each variant are depicted schematically in Figure 4.3D.

CRISPR/Cas9 and iPSC-CMs Reveal Aberrant Splicing. Many non-standard splice sites exist at a small percentage of splice sites throughout the human genome[144]. Although the minigene assay can reliably assess many variant effects on splicing, it is an *in vitro* assay that best performs with splicing to an exon on both sides with the canonical AG-GT splice sites. Therefore, outcomes may be confounded when assaying the first or last exon of a gene, or assaying rare splice sites that do not use the standard AG-GT sites. Three candidate splice-altering VUSs were incompatible with minigene assays: *KCNQ1* c.386+6T>G (first exon), *SCN5A* c.393-5C>T (AC-GT splice site),

and *SCN5A* c.4437+5G>A (AG-AT splice site). These three variants were therefore studied by introducing each variant with CRISPR-Cas9 into healthy control iPSCs , differentiating into iPSC-CMs, and then assessing splicing consequences by RT-PCR of isolated RNA[62].

The RT-PCR products of the *SCN5A* VUS c.4437+5G>A and WT RNA were approximately the same size by gel electrophoresis (Figure 4.5A); however, Sanger sequencing of relevant exon junctions showed aberrant splicing of the edited allele (Figure 4.5B). We used negative and positive controls in all RT-PCR experiments corresponding to no PCR input and a

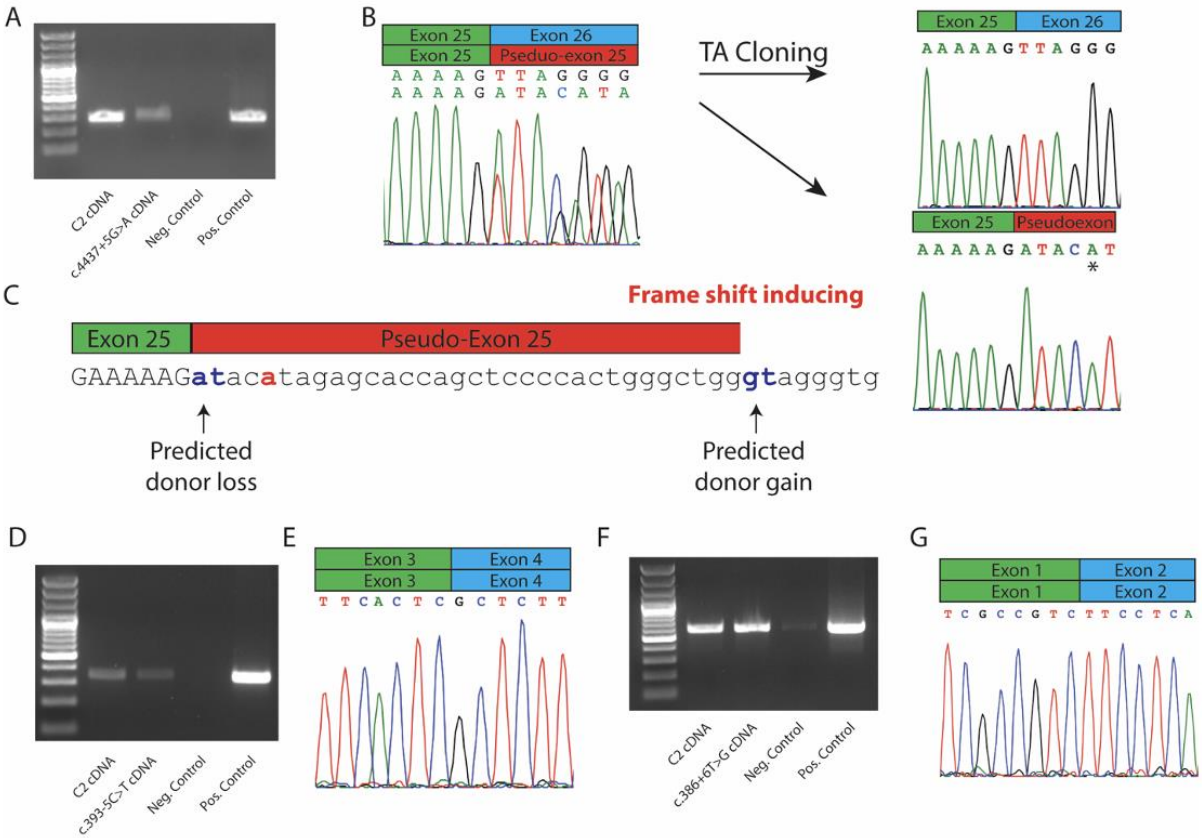


Figure 4.5. Variant functional analysis through CRISPR-edited iPSC-CMs.
A) RT-PCR analysis of population control iPSC-CM (designated C2) vs *SCN5A* c.4437+5G>A iPSC-CM. Negative control indicates no PCR input, and positive control indicates *SCN5A* cDNA construct.
B) Sanger trace of exon 25 junction shows normal splicing of WT allele and intron retention in the CRISPR-edited allele. * indicates variant DNA base from CRISPR edit.
C) Schematic depiction of predicted splicing outcome of c.4437+5G>A, consistent with experimental findings.
D) RT-PCR analysis of population control iPSC-CM (designated C2) vs *SCN5A* c.393-5C>T iPSC-CM.
E) Sanger trace of exon 3/exon 4 junction shows WT splicing of both *SCN5A* alleles c.393-5C and c.393-5T.
F) RT-PCR analysis of population control iPSC-CM (designated C2) vs *KCNQ1* c.386+6T>G iPSC-CM.
G) Sanger trace of exon 1/exon 2 junction shows canonical splicing of both *KCNQ1* alleles c.386+6T>G.

WT *SCN5A* or *KCNQ1* cDNA construct. TA cloning and sequencing revealed 2 major transcripts corresponding to the WT and predominant edited allele product, inducing a frameshifting variant

(Figure 4.5B). Splicing predictions suggested a cryptic donor site that would lead to a frameshift of the transcript, matching the experimentally observed transcript (Figure 4.5C). The variant *SCN5A* c.393-5C>T was previously studied by minigene assays and shown to induce exon skipping[145]; however, due to the non-standard splice site usage, this specific implementation of the assay may lead to inappropriate conclusions. In the CRISPR-Cas9 iPSC-CM model, we observed no change in splicing by RT-PCR (Figures 4.5D and 4.5E). RT-PCR of the *KCNQ1* VUS c.386+6T>G also showed no disruption of splicing by gel electrophoresis or by Sanger sequencing (Figures 4.5F and 4.5G). Further investigations into this variant using primers to detect large pseudoexons and NMD inhibition by cycloheximide failed to show aberrant splicing. Altogether, these studies implicated *SCN5A* c.4437+5G>A as splice-altering, and *SCN5A* c.393-5C>T and *KCNQ1* C.386+6T>G as not splice-altering.

SpliceAI Scores of Affected Variants. Aggregate SpliceAI scores are shown in Table 4.1. SpliceAI provides probabilities of losing or gaining a splice acceptor or donor site due to *cis*-genetic variation. Most variants were predicted to introduce and/or ablate a donor site as their major consequence, consistent with their position at the 3' end of each exon (12/13). SpliceAI results were generally concordant with our experimental findings. For example, the *SCN5A* c.393-5C>T VUS was not predicted to alter splicing, matching the experimental findings (Figure 4.5C). 7/11 variants that disrupted splicing *in vitro* had SpliceAI scores >0.5. One exception is *KCNH2* c.2145G>A, which was not highly predicted to disrupt a splice site by SpliceAI (0.25 aggregate score), but was observed in 2 LQTS patients[70] and caused exon skipping *in vitro*. SpliceAI also successfully identified the exonic cryptic donor splice site introduced by the LP variant *SCN5A* c.4719C>T (0.96 aggregate score).

ACMG Reclassification of Putative Splice-altering Variants. Following functional assays, we applied the PS3 criteria to 8 variants and the BS3 to 2 variants. Computational predictions from SpliceAI were also included: 2 variants meeting BP4 and 3 meeting PP3 criteria. After integrating these functional and computational findings with the previous ACMG criteria calculated by Walsh et al.[70], we reclassified 8/12 VUS to LP, 1 VUS to LB, and 1 LP variant to P (Table 4.1).

Table 4.1. Summary of variants with aggregate SpliceAI scores, assay results, case and control frequencies, and ACMG reclassifications.

Gene	Variant	Aggregate SpliceAI	Assay Type	Assay Result	Case / gnomAD	ACMG Post-Assay	Reclassification
SCN5A	c.393-5C>T	0.21	iPSC-CM	Benign	1 / 0	PM2, BS3, BP4	VUS → LB
SCN5A	c.1890G>A	0.72	Minigene	Aberrant	1 / 0	PM2, PS3,	VUS → LP
SCN5A	c.4299+6T>C	0.75	Minigene	N/A	1 / 0	PM2	VUS → VUS
SCN5A	c.4299G>C	0.80	Minigene	N/A	1 / 0	PM2	VUS → VUS
SCN5A	c.4437+5G>A	0.69	iPSC-CM	Aberrant	2 / 0	PM2, PS3	VUS → LP
SCN5A	c.4719C>T	0.96	Minigene	Aberrant	2 / 2	PM2, PS3, PS4_strong, PP3	LP → P
KCNH2	c.1128G>A	0.34	Minigene	Aberrant	2 / 0	PM2, PS3	VUS → LP
KCNH2	c.2145G>A	0.25	Minigene	Aberrant	2 / 0	PM2, PS3, BP4	VUS → LP
KCNH2	c.2398+5G>T	0.45	Minigene	Aberrant	2 / 0	PM2, PS3	VUS → LP
KCNQ1	c.386+6T>G	0.96	iPSC-CM	Benign	1 / 0	PM2, BS3, PP3	VUS → VUS
KCNQ1	c.477+5G>A	0.75	Minigene	Aberrant	7 / 3	PM2, PS3, PS4_strong	LP → P
KCNQ1	c.683+5G>A	0.29	Minigene	Aberrant	3 / 4	PM2, PS3, PS4_moderate	VUS → LP
KCNQ1	c.1032+5G>A	0.95	Minigene	Aberrant	2 / 0	PM2, PS3, PP3	VUS → LP
KCNQ1	c.1032G>T	0.74	Minigene	Aberrant	1 / 0	PM2, PS3	VUS → LP
KCNQ1	c.1032G>A	0.68	Minigene	Aberrant	9 / 0	PM2, PS3, PS4_strong	LP → P

Discussion

Splicing as Disease Mechanism. It is estimated that up to 10% of pathogenic variants may arise from aberrant splicing[64]. In addition to the widespread annotation of variants disrupting the canonical 2-bp splice sites as meeting the PVS1 criterion, the importance of splice-altering variants outside the canonical 2-bp splice sites has also been shown for Mendelian diseases[146,147]. While not all variation close to the exon/intron junction should be assumed to affect splicing, functional assays provide expedient evidence for prospective variant reclassification when observed. Although we did not use SpliceAI to help decide which variants to investigate, SpliceAI scores correlated well with functional assay results. This recently developed convolutional deep neural network appears to predict aberrant splicing with much higher accuracy than previous computational predictors[64]. Future *in vitro* splicing studies might therefore prioritize variants that are predicted to be splice-altering by this algorithm. Notably, the LQTS/BrS consortium dataset did not include any examples of deep intronic variation, commonly defined as >100 bp away from the exon/intron junction[148].

Functional Assays Reveal Aberrant Splicing. We applied functional studies to reclassify a set of VUS that have been observed in clinical cohorts of BrS and LQTS patients. The minigene remains a standard assay to address the fidelity of splicing and has been deployed in many applications[149,150]. Although a few reports of high-throughput adaptations of the minigene system have been published, these often remain restricted to a small genomic area[151] or are limited by in-frame exon triplet cassettes[152]. A complementary tool is RT-PCR of primary patient tissue or a secondary tissue developed from patient-derived iPSCs when primary tissue is unavailable. Studying candidate arrhythmia variants in a cardiomyocyte context naturally provides more information than the *in vitro* minigene assays alone. For example, investigators found many more splicing aberrations in an iPSC-CM model than in minigene assays alone for the LQTS-associated splice variant c.1032G>A[63]. We observed aberrant splicing in 1/3 of our CRISPR edited iPSC-CM

assays. We did not detect aberrant splicing for the *SCN5A* c.393-5C>T variant in iPSC-CMs, which was discordant with a previous study that showed exon skipping in a minigene assay[145]. As mentioned above, however, the use of non-canonical splice sites in minigene assays may have caused the previously observed exon skipping events. Further supporting our results is the high frequency of the related c.393-5C>A variant: an allele count of 45 in gnomAD, ~8x higher than the 2.5e-5 cutoff that we and others have used[82].

Aberrant Splicing-associated Cardiac Morbidity. Aberrant splicing is increasingly recognized as a driver of cardiovascular disease. Most of such variation has been implicated in cardiomyopathy genes[146,147,153-155]. A recent functional genomics variant reclassification effort examined aberrant splicing from a set of 56 hypertrophic cardiomyopathy (HCM) probands with *MYBPC3* variants. The investigators performed RNA analysis of 9 such variants outside the canonical splice sites from venous blood or myocardial tissue, and were able to reclassify 6 variants (4 VUS -> LP, and 2 LP -> P)[155]. In a follow up study, the investigators used patient-derived iPSC-CMs followed by RNA-seq to identify 2 known and 1 novel deep intronic splice variants in HCM probands[153]. The investigators also found that rationally designed antisense oligonucleotides (ASOs) could correct the aberrant splicing of one of the three variants. In a similar study, investigators combined SpliceAI with cohort WGS to identify known splice-altering variants and a novel splice-altering variant in *MYBPC3* from studies of peripheral blood and RT-PCR[154]. A complementary study jointly investigated splice-altering variants in *LMNA*, a dilated cardiomyopathy (DCM) associated gene, and *MYBPC3* using computational tools to prioritize variants, and a high-throughput sequencing-based minigene platform[147]. The authors also extended this system to study non-canonical splice-altering variants in *TTN*[146]. Aberrant splicing in the cardiac arrhythmias is comparatively less studied. A seminal study identified a branch point-altering variant in *KCNH2* that elicited the LQTS phenotype in a multigenerational family using patient tissue and minigene assays[156]. More recently, a deep intronic splice-altering variant in *KCNH2* was identified in a

genotype-negative LQTS family and functionally characterized using a similar CRISPR iPSC-CM framework as used in this study[157].

In silico *Prediction Concordance.* We found that SpliceAI agreed with our functional assays in most cases (Table 4.1). For example, our RT-PCR studies of the *SCN5A* c.393-5C>T variant agreed with the benign predictions of SpliceAI to have little effect. The *SCN5A* c.4719C>T variant was also strongly predicted by SpliceAI to introduce a new cryptic exonic donor site, consistent with the truncation event observed in our study. In contrast, we observed divergent outcomes with the *KCNQ1* c.386+6T>G variant, where we observed no splicing abnormality despite high predictions of disruption (0.96 aggregate score). Further studies excluding large pseudoexon retention and NMD inhibition experiments did not yield evidence of aberrant splicing. SpliceAI also was highly predictive of disrupted splicing at the +5 position (4/5 variant studied had >0.5 aggregate SpliceAI score). The +5 position has frequently been invoked in disease and is particularly intolerant to variation in evolutionary studies[158]. Experimentally, a recent high-throughput study of non-canonical splice-altering *TTN* variants also found a large enrichment of splice-altering variants at this position[146].

Therapeutic Opportunities. Splice-altering variants comprise an eminently targetable class of genetic variation, with therapeutic modalities spanning ASOs, traditional small molecules, and emerging gene editing platforms. ASO technologies that leverage coding variation benefit by retaining pharmacokinetic and pharmacodynamic properties over a broad range of possible sequence targets[159]. Indeed, 'N of 1' drug discovery has already been implemented clinically[160], and has even precipitated new regulatory processes in anticipation of increased translational opportunities[161]. Complementary mechanism-agnostic gene therapy approaches such as the 'suppression-replacement' strategy may be capable of treating a variety of variant classes in a

specified gene, including those that alter splicing[162]. Altogether, the characterization of splice-altering variants may prioritize therapeutic opportunities in the genetic arrhythmias.

Future Directions. Future studies can apply advances multiplexed assays to enable rapid high-throughput prospective evaluation of splice-perturbing variants. Further characterizing the fraction of splice-altering variants responsible for the Mendelian channelopathies may inspire future investigations in this area.

Limitations. HEK293 cells do not capture the full environment of a native cardiomyocyte. Trans-acting proteins may also affect splicing outcomes in native cardiomyocytes. Although minigene assays offer clear functional readouts, the exact consequences of the reading frame in a human gene may remain cryptic due to the limited intronic window of the assay. Predicted termination codons beyond the 100 bp of cloned intronic DNA are therefore necessarily inferred from the observed reading frames. The variants in this study represent only a small fraction of the total quantity of possible splice site-adjacent variants in these three arrythmia-associated genes. Splicing is a complex and stochastic process with a continuum of effect. The current ACMG guidelines do not capture the nuance of this behavior, and only provide a dichotomous (yes/no) framework for variant effect interpretation in this context. An additional challenge is the use of *in silico* predictors for variants adjacent to exons that use non-canonical splice sites, given the relative paucity of training data. A limitation of the CRISPR iPSC-CM approach is the necessity of an additional PAM site-disrupting variant to prevent Cas9 re-cutting. Although the SpliceAI predictions were low for each variant, we cannot exclude the possibility that a PAM variant could influence the splicing outcome. In the future, technologies such as base editors or prime editors that do not require a PAM-disrupting variant may be preferred[163].

Conclusions

We deployed functional assays to reclassify 10 variants near canonical splice sites observed in BrS and LQTS probands. This work suggests that splice-altering variation may play a prominent role in the genetic etiology of Mendelian arrhythmias, consistent with recent work defining splice-altering substrates in the cardiomyopathies[146,147,153-155]. Methods to rapidly assess suspected splice-altering variation will play an important role in the continued realization of personalized medicine for diagnostic, and potentially therapeutic purposes.

Chapter 5

Multicenter Clinical and Functional Evidence Reclassifies a Recurrent Non-canonical *FLNC* Splice-altering Variant

Portions of this chapter are under review with the same title at *Heart Rhythm*

Background: Truncating variants in Filamin C (*FLNC*) can cause arrhythmogenic cardiomyopathy (ACM) through haploinsufficiency. Non-canonical splice-altering variants may contribute to this phenotype.

Objective: To investigate the clinical and functional consequences of a recurrent *FLNC* intronic variant of uncertain significance (VUS), c.970-4A>G.

Methods: Clinical data in 9 variant heterozygotes from 4 kindreds were obtained from 5 tertiary healthcare centers. We used *in silico* predictors and functional studies with peripheral blood and patient-specific induced pluripotent stem cell-derived cardiomyocytes (iPSC-CMs). Isolated RNA was studied by reverse transcription polymerase chain reaction (RT-PCR). iPSC-CMs were further characterized at baseline and following nonsense-mediated decay (NMD) inhibition, using quantitative PCR (qPCR), RNA-seq, and cellular electrophysiology. American College of Medical Genetics and Genomics (ACMG) criteria were used to adjudicate variant pathogenicity.

Results: Variant heterozygotes displayed a spectrum of disease phenotypes, spanning mild ventricular dysfunction with palpitations, to severe ventricular arrhythmias requiring device shocks or progressive cardiomyopathy requiring heart transplantation. Consistent with *in silico* predictors, the c.970-4A>G *FLNC* variant activated a cryptic splice acceptor site, introducing a 3-bp insertion containing a premature termination codon. NMD inhibition upregulated aberrant spliced transcript by qPCR and RNA-seq. Patch clamp studies revealed irregular spontaneous action potentials, increased action potential duration, and increased sodium late current. These findings fulfilled multiple ACMG criteria for pathogenicity.

Conclusion: Clinical, *in silico*, and functional evidence support the prediction that the intronic c.970-4A>G VUS disrupts splicing and drives ACM, enabling reclassification from VUS to pathogenic.

Keywords: Arrhythmogenic cardiomyopathy, arrhythmia, splicing, nonsense-mediated decay, functional genetics

Introduction

Filamin C is an actin-binding protein encoded by *FLNC* that is expressed in cardiac and skeletal muscle. Variants in *FLNC* have been associated with autosomal dominant hypertrophic cardiomyopathy [MIM: 617047] and distal [MIM: 614065] and myofibrillar [MIM: 609524] skeletal myopathy. More recently, truncating *FLNC* variants have been shown to cause a particularly severe and highly penetrant form of arrhythmogenic cardiomyopathy (ACM) in the absence of skeletal muscle phenotypes[39]. ACM encompasses a spectrum of cardiomyopathies, including arrhythmogenic right ventricular cardiomyopathy, and biventricular and arrhythmogenic dilated cardiomyopathy (DCM)[39,164,165]. Most of the truncating *FLNC* variants reported to date have been nonsense variants or function-altering variants in canonical splice sites. Splice-disrupting intronic variants outside the 2-bp canonical splice sites are an increasingly recognized cause of inherited cardiomyopathies and arrhythmia syndromes[146,155,166]. These variants may cause frameshifts or inclusion of premature termination codons, which may converge on the molecular mechanism of haploinsufficiency identified in previous studies of *FLNC* truncating variants[165,167].

A challenge in precision medicine is interpreting the pathogenicity of variants discovered in disease-associated genes. The American College of Medical Genetics and Genomics (ACMG) has provided a framework to interpret variants that integrates multiple variant features, including clinical phenotyping, variant frequency, *in silico* tools, and functional assays[43]. These criteria result in categorical classifications of variants ranging from benign to pathogenic; however, the majority of variants do not have sufficient evidence for benign or pathogenic classification, and therefore fall into the nebulous category of Variant of Uncertain Significance (VUS)[44]. This presents an increasing clinical problem as VUS's cannot be used to guide clinical management, but they can raise alarm for providers and patients, especially in genes such as *FLNC* that are associated with malignant ventricular arrhythmias[168]. Accordingly, we sought to investigate a *FLNC* VUS (c.970-4A>G) that is located nearby a canonical splice acceptor site and has been reported in 4 separate families worldwide. By combining clinical phenotyping with computational

and functional characterization of the variant using patient peripheral blood and patient-derived induced-pluripotent stem cell-derived cardiomyocytes (iPSC-CMs), we provide multiple lines of evidence to reclassify this recurring variant from VUS to pathogenic within the ACMG framework.

Methods

Clinical Evaluation. The *FLNC* VUS c.970-4A>G was identified in 4 unrelated families across 5 tertiary care centers as part of a *FLNC* ACM registry[39]: Vanderbilt University Medical Center (VUMC), University of Colorado Cardiovascular Institute, Victor Chang Cardiac Research Institute, Amsterdam University Medical Center, and Erasmus Medical Center Rotterdam. Informed consent was obtained, and studies were approved by local Institutional Review Boards and Research Ethics Committees. Clinical data for variant heterozygotes was obtained by retrospective review of medical records. All primary data were collected and shared using a comprehensive phenotypic evaluation of available clinical data. Arrhythmogenic cardiomyopathy was defined as ventricular dysfunction not explained by ischemic, hypertensive, or valvular heart disease, together with conduction disease, atrial arrhythmias, or ventricular arrhythmias[164].

Illegitimate Transcription. RNA was isolated from peripheral blood using PAXgene tubes and the PAXgene blood RNA kit (Qiagen). cDNA was produced by RT-PCR with the use of Superscript III (Invitrogen) and random hexamers. PCR was performed using *FLNC* specific primers. cDNA was sequenced with the Brilliant Dye Terminator Cycle Sequencing system (Thermofisher Scientific) and Big Dye Terminator kit (Applied Biosystems) and analyzed using CodonCode Aligner version 8.0.2.

Generation of iPSCs. Peripheral blood mononuclear cells (PBMCs) were isolated from peripheral blood cells from the consenting participants as previously described[169]. PBMCs were reprogrammed into iPSCs using electroporation of non-integrating episomal vectors (Epi5™ Episomal iPSC Reprogramming Kit, ThermoFisher; Neon Transfection System, ThermoFisher)[170]. Following electroporation, single cells were plated on a 10 cm plate. Colonies were visually inspected for ‘iPSC-likeness’ and moved towards a Matrigel (Corning) 24 well plate.

Colonies were expanded, validated, and maintained for future studies. iPSCs were cultured in mTeSR media (STEMCELL), and passaged at a 1:20 ratio every 3-5 days.

iPSC-CM Differentiation. iPSCs were grown to approximately 80% confluency before initiating cardiac differentiation. At day 0, cells were treated with 'M1' media composed of RPMI 1640 (Life Technologies) and 2% B-27 minus insulin supplement (Life Technologies). Cells were also treated with 6 μM CHIR99021 (LC Laboratories) from day 0 to day 2. On day 3, cells were treated with 'M1' media that also contained 5 μM IWR-1 (Sigma). Afterwards, media changes occurred every second day. On days 13-17, cells were treated with 'M2' media for a metabolic selection – this was comprised of RPMI 1640 without glucose (Life Technologies) and 2% B-27 minus insulin supplement (Life Technologies). Beginning on day 19, cells were treated with 'M3' media, containing RPMI 1640 (Life Technologies) and 2% B-27 supplement (Life Technologies). Cells were cultured for a total of 30-40 days before functional studies. To measure the differentiation efficiency of our method, we stained cells for the cardiac marker α-actinin at day 35 of differentiation. Cardiomyocytes in a monolayer were dissociated using TrypLE-select (ThermoFisher), were harvested as single cells, and resuspended in DPBS (Gibco). Cells were centrifuged at 300g for 5 minutes, resuspended in DPBS a second time, and filtered into a polystyrene round-bottom culture tube with cell strainer cap (Falcon). Cells in suspension were then stained with mouse α-actinin antibody (Sigma), followed by staining with a GFP anti-mouse secondary antibody (Invitrogen). Cells then underwent analytical flow cytometry using a 3-laser Fortessa flow cytometer.

Nonsense-mediated decay (NMD)-inhibition experiments. iPSC-CMs between day 30 and 32 of differentiation were treated with either vehicle dimethyl sulfoxide (DMSO) or cycloheximide (100 μM). After 6 hours, RNA was isolated using the Qiagen RNeasy Minikit. RT-PCR was carried out

using the SuperScript III One-Step RT-PCR Kit (Invitrogen). For TA cloning, the gel extracted product was ligated into the pGEM®-T Vector (Promega). After ligation, transformation, and plating, single colonies were grown up, DNA was extracted with the Qiagen Miniprep kit as above, and transcript composition was determined by Sanger sequencing.

qPCR and RNA-seq. For qPCR experiments, RNA was reverse transcribed into cDNA using random hexamers with the SuperScript III kit (Invitrogen). Quantitative PCR was carried out using TaqMan Gene Expression Assays (Applied Biosystems), TaqMan Fast Universal PCR Master Mix (Applied Biosystems), and a CFX-96 Real-Time PCR system (BioRad). We used predesigned probe Hs00155124_m1 (ThermoFisher Scientific). All qPCR reactions were performed in triplicate and expressed as a normalized value to the 60S ribosomal protein L19 (RPL19) using the comparative ΔCt method. Statistical analyses were performed with a two-tailed unpaired t-test implemented in R Studio. For RNA-seq, RNA was isolated from vehicle and CHX treated cells as described above. RNA integrity was assessed on a 2100 Bioanalyzer (Agilent). Sequencing was completed by VANTAGE (Vanderbilt Technologies for Advanced Genomics), which prepared a polyA-selection library, sequenced the library with a requested 50 million PE100 reads on an Illumina Nova-seq sequencer, and returned FASTQ files. FASTQ quality was assessed using fastqc. We performed Sashimi plot analysis by analyzing fastq files with fastp[171], followed by alignment with hisat2[172], indexing of bam files[173], and visual analysis by Integrated Genomics Viewer[174].

General Electrophysiology Methods. All data acquisitions were carried out using a MultiClamp 700B patch-clamp amplifier and software Clampex 10.7 (Molecular Diagnostics). Currents were filtered at 5 kHz (-3 dB, four-pole Bessel filter) and digitized using an analog-to-digital interface (Digidata 1550B, Molecular Diagnostics). To minimize capacitive transients, capacitance was compensated ~80%. Series resistance was 1-4 MΩ. Voltage-clamp protocol to record late sodium

current is presented in the figure inset. Linear leakage currents were digitally subtracted online by the P/N protocol. All action potential (AP) and ion current parameters were analyzed using Clampfit 10.7 software. All experiments were conducted at room temperature (~23 ˚C).

Patch-clamp Electrophysiology. Recordings of action potentials: In current-clamp mode, non-paced spontaneous action potentials (APs) in patient-derived and healthy-control iPSC-CMs were recorded after breaking-in cell membrane. Paced APs were elicited by injection of a brief stimulus current (1-2 nA, 2-6 ms) at a stimulation rate of 0.5 Hz. For AP experiments, the bath (extracellular) solution was normal Tyrode's, which contained (in mmol/L) NaCl 135, KCl 4.0, $CaCl_2$ 1.8, $MgCl_2$ 1.0, HEPES 5.0 and glucose 10, with a pH of 7.4. The pipette-filling (intracellular) solution contained (in mmol/L) KCl 130, ATP-K_2 5.0, $MgCl_2$ 1.0, $CaCl_2$ 1.0, BAPTA 0.1 and HEPES 5.0, with a pH of 7.3. Ten successive AP traces were recorded and averaged for analysis of AP durations at 90% repolarization (APD_{90}). All AP experiments were repeated among three independent differentiations of the patient-derived and control iPSCs.

Recordings of late sodium current. To record late sodium currents in patient-derived and healthy-control iPSC-CMs, a K^+- and Ca^{2+}-free external solution with 135 mM NaCl was used to optimize late sodium current, which was recorded by 200-ms depolarization pulses to -30 mV from the holding potential of -120 mV. The level of late sodium current was measured at a time window of 195-198 ms before the ending of testing pulses. Ten continuous current traces were recorded and averaged for analysis of the late sodium current level as the percentage of peak current at a testing potential of -30 mV. The pipette (intracellular) solution contained (in mmol/L) NaF 5, CsCl 120, BAPTA 5 and HEPES 10, with a pH of 7.3 adjusted by CsOH. Glass recording pipettes were pulled and heat polished with a tip resistance of 0.5 ~ 1.5 MΩ.

Results

Clinical Phenotypes. We studied 9 c.970-4A>G variant heterozygotes from 4 unique families collected as part of the *FLNC* ACM Registry[39]. The clinical phenotypes and pedigrees of each family are individually described below (Figure 5.1).

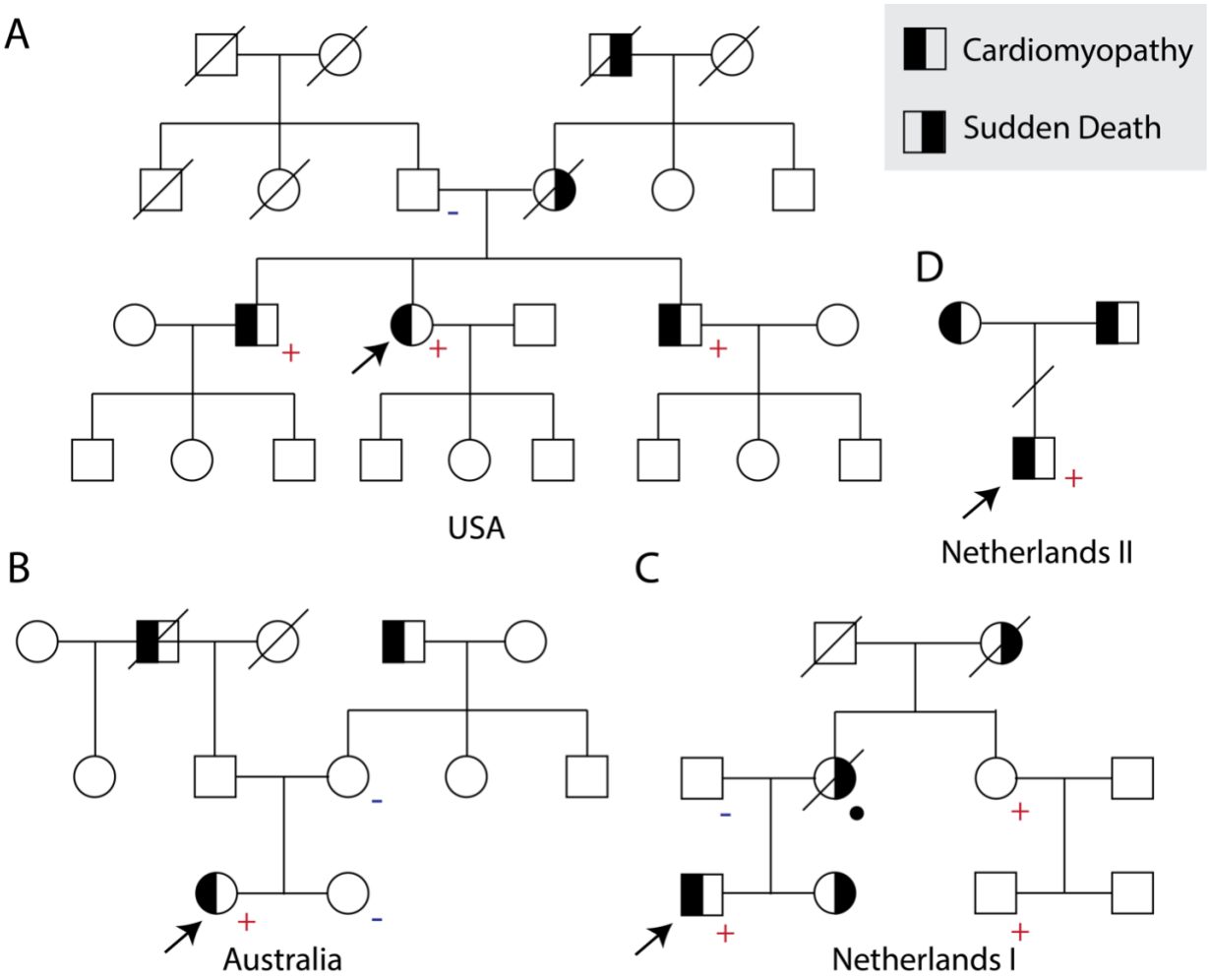


Figure 5.1. Pedigrees and Phenotypes of *FLNC* c.970-4A>G families.
A) 4-generation pedigree of variant heterozygotes cared for by USA centers. All siblings of proband generation are heterozygous for the variant and have a history of ACM. The mother died at 42 of sudden death following the diagnosis of an unknown arrhythmia, and the maternal grandfather suddenly died at the age of 60 one year after a negative ischemic evaluation.
B) 3-generation pedigree of a young proband followed at an Australian center. The index patient had early onset DCM requiring heart transplant. Her mother and sister were both negative for the variant. Family history was otherwise remarkable for both paternal and maternal grandfathers with diagnosed DCM.
C) 3-generation pedigree from Netherlands. The proband presented later in life with severe DCM. The mother and maternal grandmother both died suddenly. A maternal aunt and child were variant heterozygotes in apparently normal health, who were unable to be contacted for additional phenotyping.
D) 2-generation pedigree of additional family from the Netherlands. The proband presented in mid-adulthood with DCM. Both parents had a history of cardiomyopathy, but further were precluded by estrangement from his family.
+ indicates genetic testing and heterozygous status for c.970-4A>G. – indicates wildtype status. • indicates obligate heterozygote. + or – absent when genetic testing was not completed. Arrow indicates proband within each family. Dash line across relationships indicates estrangement.

Family A (U.S.A.). The proband was a female who first presented with fatigue and palpitations associated with premature ventricular contractions (PVCs) that were first documented during pregnancy at the age of 33. At the time she had a cardiac workup that showed multifocal PVCs with no evidence of ACM by cardiac MRI or echo. Family history was notable for sudden unexplained death of both her mother at the age of 42 and maternal grandfather at the age of 60. Both deaths occurred during mild to moderate physical activity. Following a gap in care and recurrence of symptoms, the patient was referred for additional management 7 years later. Repeat imaging showed normal left ventricular (LV) ejection fraction (LVEF; 54%) and right ventricular (RV) ejection fraction (RVEF; 57%) with mild LV (5.3 cm) and RV (3.9 cm) dilation; there was no late gadolinium enhancement (LGE) (Table 5.1). An exercise stress test showed ventricular ectopy at rest that increased with exercise and episodes of non-sustained ventricular tachycardia (NSVT) of multiple ventricular morphologies (Figure 5.2). The tentative clinical

Covariate	USA	Australia	Netherlands I	Netherlands II
Sex	Female	Female	Male	Male
Age at Dx	41	11	59	41
Hypertension	Yes	No	No	No
NYHA Class	I	III	II	II
Symptoms	Nausea	Dyspnea	Dyspnea	Dyspnea
ECG (base eval)	Sinus	Sinus	Sinus	Sinus
Echo				
LVEDD, cm	-	7	-	6.8
LVEF, %	-	27	-	15
MRI				
LVEF, %	54	8	30	35
RVEF, %	57	-	44	43
LGE	-	-	Non-specific fibrosis	3% LV
Arrhythmias				
Atrial	SVT	-	-	-
Ventricular	PVCs, NSVT	PVCs, NSVT	VF	PVCs

Table 5.1. Clinical Phenotypes of c.970-4A>G Probands.
SVT – supraventricular tachycardia. PVC – premature ventricular contraction. NSVT – non-sustained ventricular fibrillation. VF – ventricular fibrillation. LGE – late gadolinium enhancement. LVEDD – left ventricular end diastolic dimension. LVEF – left ventricular ejection fraction. RVEF – right ventricular ejection fraction.

diagnosis was catecholaminergic polymorphic ventricular tachycardia (CPVT), and genetic testing with an arrhythmia panel was negative (genes tested at each site in Table 5.2).

Clinical evaluation of the proband's 2 brothers showed that they both had PVCs with multiple morphologies. The brother seen at Colorado had NYHA Class I HF with a preserved LVEF (55%) with borderline dilation of the LV (5.6 cm) at baseline. Holter monitor recorded PVCs (4597), ventricular couplets (121), and a 4-beat run of NSVT during a 24-hour period. Cardiac MRI showed a normal LVEF (57%) and RVEF (52%) along with segmental LGE of the mid-inferoseptum at the RV insertion. Cardiac MRI of the second brother revealed a normal LVEF (61%), reduced RVEF (45%), and borderline dilation of the LV (5.5 cm) and RV; no LGE was noted by MRI. Combining these findings lead to the reassignment of the clinical diagnosis to ACM in all siblings. Repeat genetic testing with a cardiomyopathy

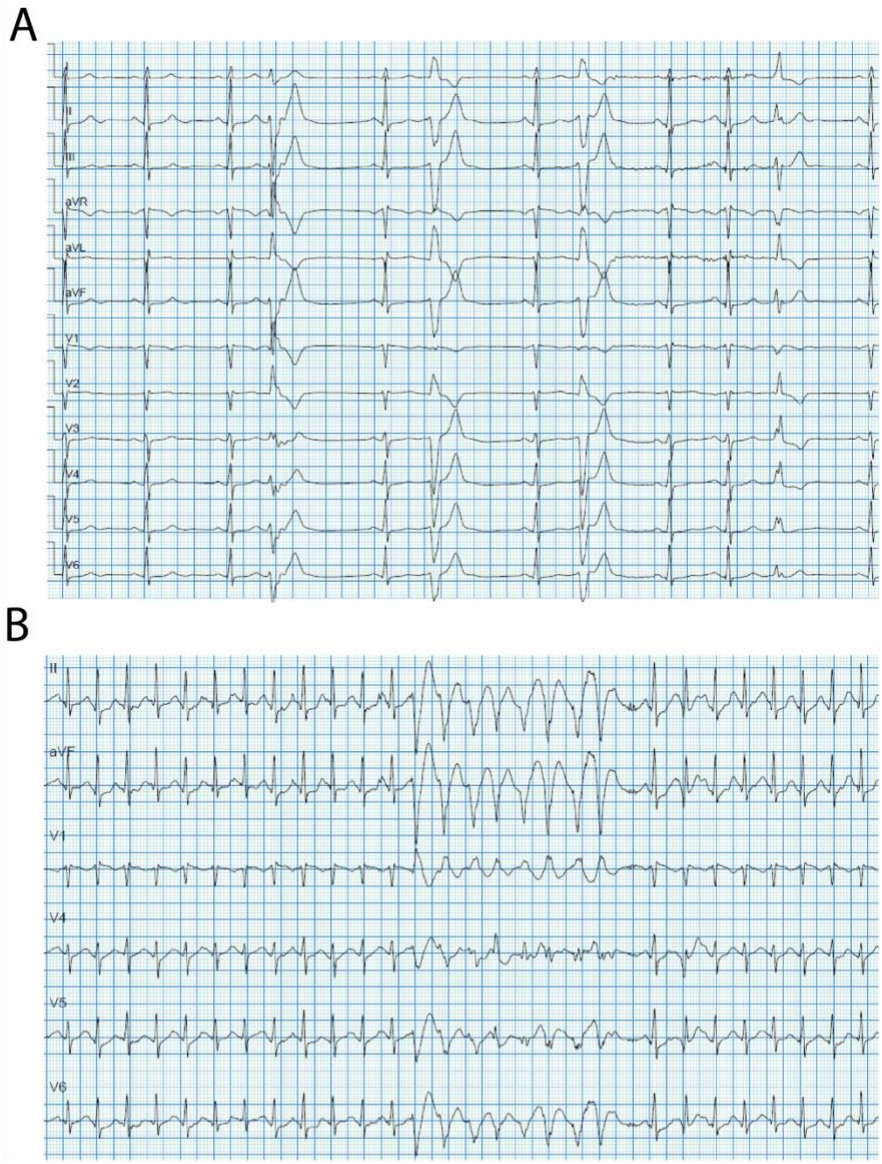


Figure 5.2. Exercise stress test of USA proband.
A) Proband ECG (Family 1) at rest prior to exercise. Late-coupled PVCs (3 morphologies).
B) Proband ECG at Stage 3 of stress test with an 8-beat episode of relatively monomorphic NSVT.

gene panel identified the intronic *FLNC* variant c.970-4A>G, which was classified as a VUS[45]. This variant co-segregated in all 3 phenotype-positive siblings. The variant was absent in the father, suggesting that the mother was a variant heterozygote. Together, the 3 siblings have 9 children who are currently phenotypically normal and have not completed genetic testing.

Family B (Australia). The proband was a female diagnosed with ACM at 11 years of age (Table 5.1). She first presented with dyspnea and NYHA Class III heart failure. A baseline echo showed severe LV dilation (7.0 cm), mild mitral regurgitation, and depressed LVEF (27%). A 24-hour ambulatory monitor showed PVCs and a single instance of a ventricular couplet. Panel sequencing of 105 genes revealed the *FLNC* VUS and an additional VUS in *MYLK2* (c.1360G>T; p.Val454Leu) (all genes tested listed in Table 5.2). She had a family history of DCM in both her maternal and paternal grandfathers. The paternal grandfather was diagnosed with DCM in his 7th decade of life and the maternal grandfather died of DCM complications in his 5th decade of life – further details are unknown. Neither her mother or sister carried the *FLNC* variant and were both healthy; the father is unaffected and declined completion of genetic testing. The proband experienced progressive contractile dysfunction that necessitated heart transplantation at age 12. The histology report post-transplant identified nuclear hyperchromasia and enlargement without inflammation or fibrosis, ultimately read as DCM of uncertain etiology. She has remained healthy after successful transplantation.

Family C (Netherlands I). The familial proband was a male diagnosed with DCM at age 59 (Table 5.1). Cardiac MRI revealed a severely reduced LVEF (30%), mildly reduced RVEF (45%), and non-specific myocardial fibrosis. He had a negative ischemic evaluation by coronary angiography. Genetic testing revealed the *FLNC* c.970-4A>G VUS and a missense VUS in *TTN*. The proband's sister suffered a cardiac arrest at the age of 50 but declined genetic testing. His father was phenotype-negative and did not carry the *FLNC* variant. His mother died of unspecified cardiac

causes and is an obligate carrier. A maternal aunt and cousin were genotype-positive for the *FLNC* variant but were cared for by different providers. Due to local ethical and privacy concerns, they were unable to be contacted to obtain prior testing results or to perform additional phenotyping. The proband has since received an implanted cardioverter defibrillator and has received multiple appropriate shocks for ventricular arrhythmias.

Family D (Netherlands II). The proband was a male diagnosed with ACM at 41 years of age. He presented with heart failure and severally depressed LVEF (15 %) with diffuse hypokinesia of the left ventricle by echo and PVCs (Table 5.1). Subsequent MRI evaluation revealed depressed LVEF (35%), RVEF (43%) and late gadolinium enhancement (3% of LV). With treatment LVEF improved to 45% at the age of 45 years. Genetic testing with a cardiomyopathy panel revealed the *FLNC* VUS c.970-4A>G variant along with another VUS in *FLNC*, c.7399C>T/p.Arg2467Cys, a VUS in *TTN*, c.71310C>G/p.Asp23770Glu, and a VUS in *MYBPC*, c.1471G>A/p.Val491Met. Family history noted maternal heart failure and an enlarged heart in the father; however, the proband is estranged from his family, eluding the possibility of variant segregation within the family.

Table 5.2. Genes tested by each site.

Center	Genes
Vanderbilt and Colorado (arrhythmia panel)	*ABCC9, AKAP9, ANK2, CACNA1C, CACNA2D1, CACNB2, CALM1, CALM2, CALM3, CASQ2, CAV3, DES, DSC2, DSG2, DSP, GPD1L, HCN4, JUP, KCND3, KCNE1, KCNE2, KCNE3, KCNE1L, KCNH2, KCNJ2, KCNJ5, KCNJ8, KCNQ1, LMNA, NKX-2-5, PKP2, PLN, RANGFR, RYR2, SCN10A, SCN1B, SCN2B, SCN3B, SCN4B, SCN5A, SNTA1, TGFB3, TMEM43, TRDN, TRPM4, TTN*
Vanderbilt and Colorado (cardiomyopathy panel)	*ABCC9, ACTC1, ACTN2, AKAP9, ALMS1, ALPK3, ANK2, ANKRD1, BAG3, BRAF, CACNA1C, CACNA2D1, CACNB2, CALM1, CALM2, CLM3, CASQ2, CAV3, CHRM2, CYRAB, CSRP3, CTNNA3, DES, DMD, DOLK, DSC2, DSG2, DSP, DTNA, EMD, EYA4, FHL1, FKRP, FKTN, FLNC, GAA, GATA4, GATA5, GATA6, GATAD1, GJA5, GLA, GNB5, GPD1L, HCN4, HFE, HRAS, ILK, JPH2, JUP, KCNA5, KCND3, KCNE1, KCNE2, KCN3, KCNE1L, KCNH2, KCNJ2, KCNJ5, KCNJ8, KCNQ1, KRAS, LAMA4, LAMP2, LDB3, LMNA,*

	LRRC10, MAP2K1, MAP2K2, MIB1, MT-ND1, MT-ND5, MT-ND6, MT-TD, MT-TG, MT-TH, MT-TI, MT-TK, MT-TL1, MT-TL2, MT-TM, MT-TQ, MT-TS1, MT-TS2, MURC, MYBPC3, MYH6, MYH6, MYL2, MYL3, MYL4, MYLK2, MYOZ2, MYPN, NEBL, NEXN, NKX2-5, NRAS, PDLIM3, PKP2, PLN, PPA2, PRDM16, PRKAG2, PTPN11, RAF1, RANGRF, RBM20, RIT1, RYR2, SCN10A, SCN1B, SCN2B, SCN3B, SCN4B, SCN5A, SGCD, SHOC2, SNTA1, SOS1, TAZ, TSBX20, TCAP, TECRL, TGFB3, TMEM43, TMPO, TNNC1, TNNI3, TNNT2, TOR1AIP1, TPM1, TRDN, TRPM4, TTN, TTR, TXNRD2, VCL
Australia	*A2ML1, ABCC9, ACADVL, ACTC1, ACTN2, AGL, ALMS1, ANKRD1, BAG3, BRAF, CACNA1C, CALR3, CAV3, CBL, CHRM2, CPT2, CRYAB, CSRP3, CTF1, CTNNA3, DES, DMD, DNAJC19, DOLK, DSC2, DSG2, DSP, DTNA, ELAC2, EMD, EYA4, FHL1, FHL2, FKRP, FKTN, FLNC, GAA, GATA4, GATA6, GATAD1, GLA, HCN4, HRAS, ILK, JPH2, JUP, KRAS, LAMA4, LAMP2, LDB3, LMNA, LRRC10, MAP2K1, MAP2K2, MTO1, MYBPC3, MYH6, MYH7, MYL2, MYL3, MYLK2, MYOM1, MYOZ2, MYPN, NEBL, NEXN, NF1, NKX2-5, NPPA, NRAS, PDLIM3, PKP2, PLEKHM2, PLN, PRDM16, PRKAG2, PTPN11, RAF1, RASA1, RBM20, RIT1, RRS, RYR2, SCN5A, SDHA, SGCD, SHOC2, SLC22A5, SOS1, SOS2, SPRED1, TAZ, TCAP, TGFB3, TMEM43, TMEM70, TMPO, TCCN1, TNNI3, TNNT2, TPM1, TTN, TTR, TXNRD2, VCL*
Netherlands I	*ACTC1, ACTN2, ALPK3, ANKRD1, BAG3, CALR3, CAV3, CDH2, CRYAB, CSRP3, CTNNA3, DES, DSC2, DSG2, DSP, EMD, FHL1, FHL2, FLNC, GLA, HCN4, JPH2, JUP, LAMA4, LAMP2, LDB3, LMNA, MIB1, MYBPC3, MYH6, MYH7, MYL2, MYL3, MYOZ2, MYPN, NEXN, PKP2, PLN, PPA2, PRDM16, PRKAG2, RBM20, SCN5A, TAZ, TCAP, TMEM43, TNNC1, TNNI3, TNNT2, TPM1, TTN, TTR, VCL*
Netherlands II	*ABCC9, ACAD9, ACTC1, ACTN2, ALPK3, ANKRD1, ANO5, BAG3, CALR3, CASQ2, CAV3, CRYAB, CSRP3, CTNNA3, DES, DSC2, DSG2, DSP, DTNA, EMD, EYA4, FHL1, FKTN, FLNC, GATAD1, GLA, HCN4, ILK, JPH2, JUP, KIF20A, LAMA4, LAMP2, LDB3, LMNA1, MIB1, MYBPC3, MYH6, MYH7, MYL2, MYL3, MYLK2, MYOZ1, MYOZ2, MYPN, NEBL, NEXN, NKX2-5, PDLIM3, PKP2, PLN, PPA2, PPCS, PRDM16, PRKAG2, RAF1, RBM20, RYR2, SCN5A, SGCD, TAZ, TBX20, TCAP, TGFB3, TMEM43, TNNC1, TNNI3, TNNI3K, TNNT2, TPM1, TRIM63, TTN, TTR, TXNRD2, VCL*

FLNC c.970-4A>G Variant Characteristics. The *FLNC* variant c.970-4A>G is located adjacent to the canonical splice acceptor site of exon 6, within the rod 1 functional domain. The frequency of this variant is 5.35 * 10^{-5} (15 alleles) in the gnomAD database (v2.1.1)[77]. Given the proximity to the canonical splice acceptor site, the impact on splicing was assessed with *in silico* tools. SpliceAI is an in silico tool that calculates a probability, from 0 to 1, of a genetic variant to introduce or weaken a splice acceptor or donor site. The algorithm also outputs a predicted location of those splice sites with respect to the variant. In this case, the algorithm strongly predicts (0.88) that a competing splice acceptor will be gained 1 nucleotide downstream, and the canonical splice acceptor lost 4 nucleotides downstream. The molecular consequence would be a 3-bp insertion of intronic sequence coincidentally containing a PTC. Both SpliceAI (Figure 5.3A) and AlaMut predicted the activation of a cryptic splice acceptor site 3 nucleotides upstream of the canonical acceptor site (Figure 5.3B). Based on this prediction and strong evidence of *FLNC* haploinsufficiency as a disease

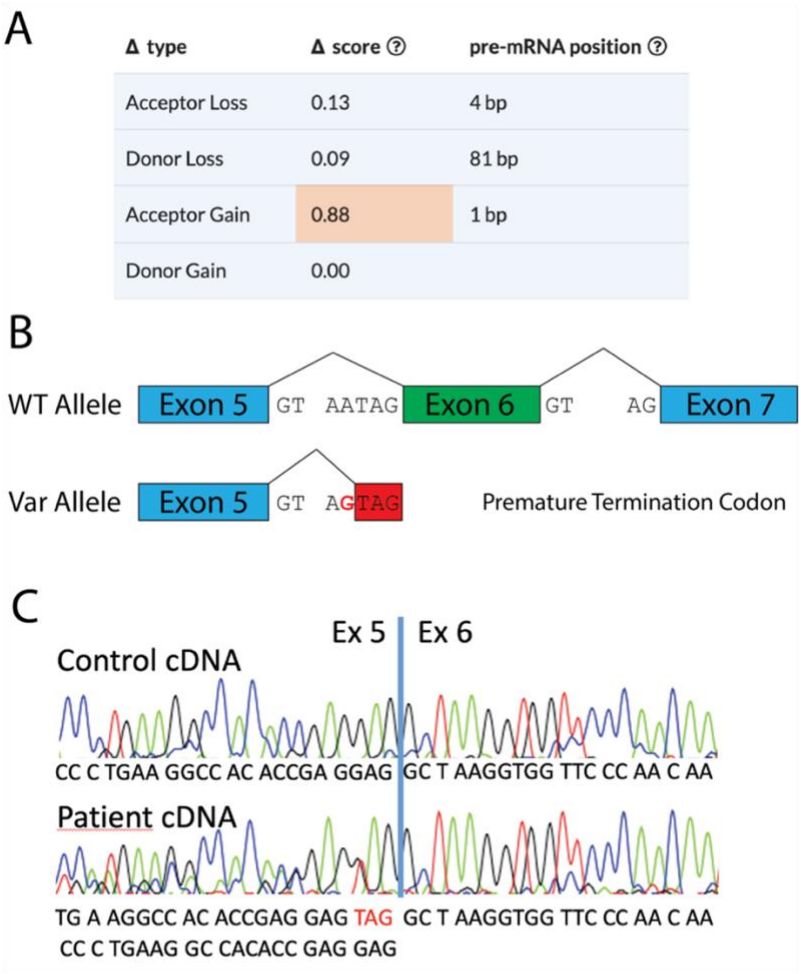


Figure 5.3. *FLNC* c.970-4A>G variant characteristics.
A) SpliceAI scores for the FLNC c.970-4A>G. Strongly predictive of the introduction of a competing splice donor site immediately adjacent to the variant.
B) Schematic of the functional consequences for variant heterozygotes. Competing splice site introduces a 3-bp premature termination codon into the aberrantly spliced exon.
C) Sanger trace of cDNA analysis of peripheral blood obtained from the proband of Family C (The Netherlands I). Trace shows introduction of the predicted premature termination codon in one transcript.

mechanism, we sought to investigate the effects of the variant on splicing in human peripheral blood through illegitimate transcription[58]. Sanger sequencing of the RT-PCR product from isolated peripheral blood RNA from the Family C proband revealed the presence of a 3 base pair insertion (Figure 5.3C). This in-frame insertion is predicted to introduce a premature truncation (UAG stop codon).

Transcriptional Studies of Patient-derived iPSC-CMs. Peripheral blood mononuclear cells were obtained from the VUMC proband, reprogrammed into iPSCs, and differentiated into iPSC-CMs (see Materials and Methods). The cells were confirmed to be heterozygous for the intronic variant. Analysis of the splice junction was completed using RT-PCR with RNA isolated from the patient and control iPSC-CMs (Figure 5.4A). We observed no difference in mobility of the RT-PCR amplicons between the patient and healthy control; Sanger sequencing of the patient line showed a normal splice junction between exons 5 and 6 (Figure 5.4B).

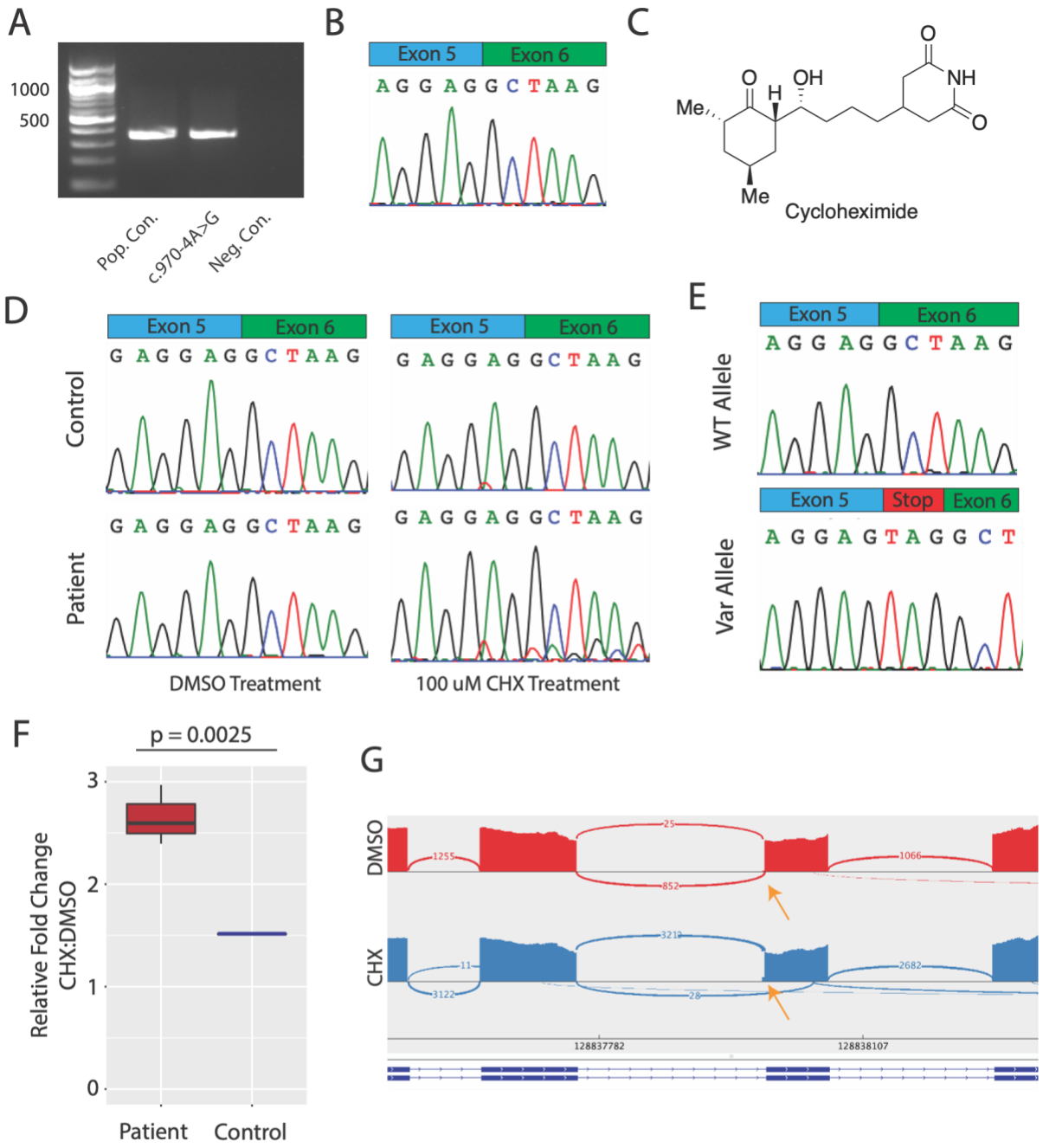


Figure 5.4. *FLNC* variant c.970-4A>G iPSC-CM transcriptional investigations.
A) iPSC-CM RNA was reverse transcribed into cDNA and amplified by PCR. Gel shows amplified DNA from a healthy control (Pop. Con.), USA proband, and a negative control (Neg. Con.) without RNA input to RT-PCR.
B) Sanger sequencing result of the gel extracted band from USA proband iPSC-CM RT-PCR product. Sequence shows only normal splicing of exon 5 to exon 6.
C) Chemical structure of cycloheximide, a natural product inhibitor of nonsense-mediated decay.
D) Sanger sequencing of RT-PCR products from patient and control lines treated with DMSO or 100 μM cycloheximide (CHX).
E) TA cloning of gel extracted bands from 100 μM CHX treatment in panel E. Identification of aberrantly spliced transcript degraded in absence of NMD inhibition.
F) Relative fold change of transcript abundance as determined by qPCR. Two-tailed t-test.
G) Sashimi plot visualization of RNA-seq data in patient cell line. Low level indel incorporation is observed after vehicle treatment, which is markedly upregulated after inhibition of NMD.

This result was at variance with the illegitimate transcription result above, so we further tested the idea that the aberrantly spliced variant was degraded by NMD. Non-sense mediated decay (NMD) is a conserved cellular process by which premature termination codons (PTCs) in nascent transcripts are recognized and degraded in the transcriptional milieu[175]. This homeostatic process can nevertheless precipitate pathology when degraded PTC-containing transcripts are associated with a haploinsufficient disease. Chemical tools, such as cycloheximide (CHX), may be used to inhibit this machinery and implicate it in disease through functional studies. After exposure to CHX (Figure 5.4C), the patient iPSC-CM RT-PCR amplicons showed the emergence of a 3-bp in-frame insertion on sequence analysis that was confirmed by TA cloning to represent otherwise degraded variant transcript (Figures 5.4D and 5.4E). Figure 5.4F shows experiments with qPCR displaying increased abundance of *FLNC* transcript in patient vs control lines after treatment with cycloheximide ($p < 0.01$, N=3), presumably by decreasing degradation of variant transcript. The significantly higher increase in the patient *FLNC* transcripts suggests a recovery of aberrantly spliced transcript compared to the mild upregulation observed in the control, consistent with global dysregulation of transcript expression after CHX treatment. To further investigate the transcriptional alterations in the patient-derived iPSC-CMs, we performed RNA-seq of cells treated with vehicle DMSO or CHX and visualized alignments using Sashimi plots (Figure 5.4G). We observed minimal aberrant splicing in the vehicle control that could not be appreciated by Sanger sequencing alone (25 aberrant reads vs 852 canonical reads) and increased aberrant splicing after treatment with CHX (321 aberrant reads vs 1740 canonical reads).

Electrophysiological Studies of iPSC-CMs. To determine whether aberrant splicing led to arrhythmogenic behavior, we next sought to explore the electrophysiologic properties of the patient iPSC-CMs compared to a healthy control iPSC-CMs. Spontaneous action potentials of each line were recorded by manual patch-clamp at day 35 of differentiation. Multiple differentiations of the healthy control line (N=3) featured regular, spontaneous depolarizations

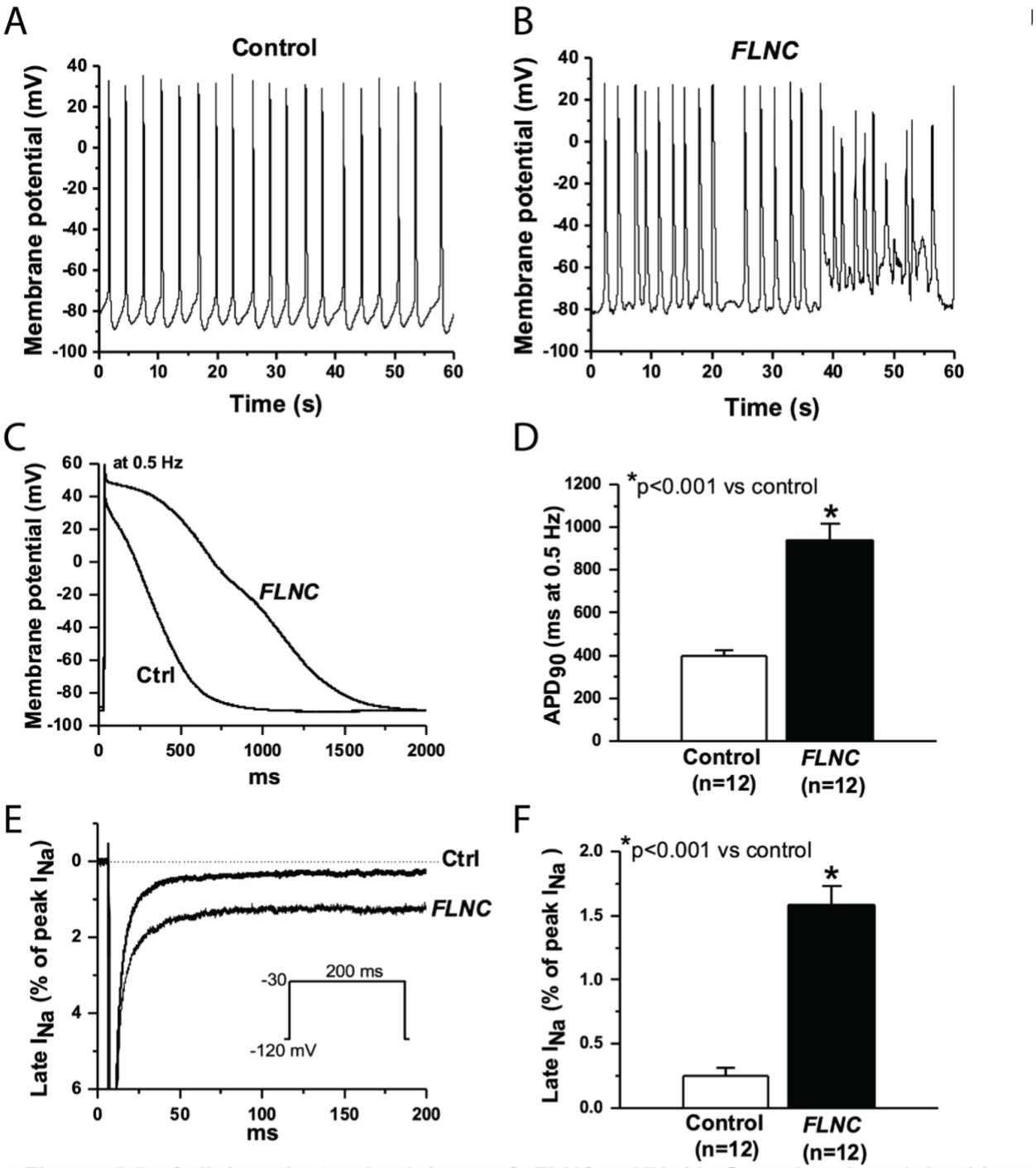


Figure 5.5. Cellular electrophysiology of *FLNC* c.970-4A>G proband and healthy control iPSC-CMs.

A) Example spontaneous action potentials measured in control iPSC-CMs.

B) Example spontaneous action potentials in *FLNC* proband iPSC-CMs.

C) Individual traces of paced action potential duration among *FLNC* variant and control lines.

D) Quantification of APD_{90} in each cell line. Longer duration observed among *FLNC* cells from 3 independent replications (2 sided t-test) .

E) Representative late sodium current traces. Voltage protocol shown in the inset.

F) Quantification of late current as a % of peak sodium current (2 sided t-test).

spontaneous depolarizations and irregularity of spontaneous depolarizations, including early after depolarizations (Figure 5.5B). Action potential duration was significantly longer in patient vs healthy control lines (Figure 5.5C and 5.5D). Sodium late current has been implicated as a terminal manifestation of aberrations in the PDGFRA signaling axis[169], a signaling pathway previously invoked in the study of other FLNC truncating variants[176]. To investigate the impact of sodium late current for this splice-altering variant, we recorded late current for patient and control lines. We observed significantly increased late current among proband vs. control iPSC-CMs (Figure 5.5E and 5.5F).

ACMG Variant Reclassification. The variant c.970-4A>G was previously classified as a VUS in ClinVar (ClinVar: 432231[45]) following application of PP3 criteria alone. We implemented evidence based on the investigations above to reclassify this variant to Pathogenic as follows[43]. Cosegregation of the phenotype with variant heterozygote status supported the implementation of PP1 at the moderate level (cosegregation of disease in multiple affected family members in a gene definitively known to cause the disease). Functional data in 2 separate assays from 2 independent heterozygotes supported the application of PS3 (well-established functional studies supportive of a damaging effect) following electrophysiological assays and PVS1 (null variant in a gene where loss-of-function is a known mechanism of disease) following transcriptional studies. The multiple concordant deleterious splicing predictions supported application of PP3 (multiple supporting lines of computational evidence). Given the frequency of 5.35e-5 (15 alleles) in gnomAD, we did not apply variant frequency criteria based on recommendations for highly penetrant variants in ARVC and DCM[82]. Altogether, these data support variant reclassification to pathogenic.

Discussion

Here we report the clinical phenotypes and probable molecular mechanisms associated with a heterozygous non-canonical splice-altering *FLNC* variant, c.970-4A>G, identified in 4 unrelated kindreds. We show that the variant presents with a range of severity, including early-onset heart failure necessitating heart transplantation in a pediatric patient and life-threatening arrhythmic events, to mild ventricular dysfunction and symptomatic ventricular ectopy during exercise. Using patient lymphocytes and a patient-specific iPSC-CM model, we demonstrate aberrant splicing consistent with computational predictions. Furthermore, we discover that the resulting in-frame premature termination codon is a substrate for NMD, and that the iPSC-CMs recapitulate the arrhythmogenicity observed in patients. Using the ACMG criteria, we were able to fulfill multiple criteria to reclassify this variant to pathogenic. This work highlights the utility of international clinical and functional collaboration to address recurrent VUS.

Spectrum of FLNC Phenotypes. Variants in *FLNC* have been associated with many cardiac phenotypes, including DCM[177], ARVC[178], and restrictive cardiomyopathy[179,180]. One extensive study of *FLNC* variants in patients with familial DCM found variants in 2.2% of affected individuals[181]. Further studies from the *FLNC* ACM Registry defined the natural history of variant carriers, with 49% meeting criteria for DCM, 25% meeting criteria for ARVC, and 3% arrhythmogenic left-dominant cardiomyopathy[39]. Beyond selected cohorts, studies of *FLNC* loss-of-function variants in a large biobank have also supported an increased risk of ACM among unselected individuals carrying such variants[182]. Our clinical findings of predominant LV dysfunction and ventricular arrythmias are consistent with previous reports of truncating *FLNC* variants that were highly arrhythmogenic and favored the LV in contrast to ACM presentations with different genetic substrates[165]. An interesting observation of our clinical case series is that heterozygotes of the c.970-4A>G variant appear to have less LGE than those harboring other truncating variants. Genetic testing will be critically important to help guide these interventions

based on genotype, especially for those with minimal structural disease by MRI upon initial clinical evaluation. Collectively, the findings in these studies support the idea that arrhythmias and sudden death may often be the first manifestation of a genetic cardiomyopathy[38]. This is especially relevant for related individuals of genotype-positive patients in our 4 families, who may benefit from completing genetic testing and detailed clinical phenotyping.

Utility of Functional Assays to Facilitate Variant Interpretation. Our work highlights the utility of functional assays in disease-relevant systems to illuminate molecular mechanisms. We show consistent results with both secondary (peripheral blood) and pseudo-primary tissue (iPSC-CMs) from 2 different patients harboring the same variant. The degree of NMD is different between the 2 tissues, with the variant allele completely degraded in the iPSC-CMs. This result may closely resemble the biology in the native cardiomyocyte and contribute to haploinsufficiency provoking the phenotype. NMD serves as regulatory machinery to degrade transcripts containing premature termination codons, and may be a confounder when not explored *in vitro* splicing assays[183]. In our study, we could only identify the aberrantly spliced transcript in venous blood, but not in iPSC-CMs without CHX. Cell-type specific influence of NMD may contribute to divergent outcomes of splicing among different assays (e.g. venous blood, minigene, primary tissue)[155]. Due to this confounding, the yield of splice-altering variants may be underestimated. Future studies should consider inhibition of NMD before ruling out pathogenic variation in haploinsufficient disease. These assays may be applied to other intronic variants in *FLNC* that may be unrecognized as causes of ACM through non-canonical aberrant splicing. Moving beyond transcriptional defects, the arrhythmogenicity of the patient-derived iPSC-CMs provides preliminary evidence to link aberrant splicing to an electrophysiologic property associated with arrhythmic events – sodium late current. Furthermore, the presence of early after depolarizations observed in the spontaneous action potentials may provide an additional point of initiation for severe ventricular arrhythmias. Further elucidating this exact mechanism will benefit from both pharmacological and genetic

approaches, including chemical tools and isogenic controls in multiple lines, which are beyond the scope of the current study.

Recurrence of the Intronic Variant. An interesting aspect of this variant is the relatively high frequency of this variant in gnomAD (allele count 15; allele frequency $5.35 * 10^{-5}$)[77]. This is a rare, but not ultra-rare frequency which does not clearly lend itself to classification as benign or pathogenic[82]. A characteristic related to the frequency is the recurrence of this variant in geographically distinct families. This may reflect a mutational hotspot adjacent to this exon (the variant c.970-5A>G is also observed in gnomAD 8 times but annotated as likely benign in ClinVar). The natural history of this variant among the 4 families is characterized by phenotypic heterogeneity, with some heterozygotes experiencing mild ectopy, while others have experienced aborted cardiac death. Although cosegregation is weakened by 2 putatively healthy variant heterozygotes in Family C, we cannot exclude the possibility of limited phenotype given the inability to access additional clinical data; we therefore cannot distinguish between clinically unaffected (incomplete penetrance) and unknown phenotype[184].

Limitations. Both functional assays are limited as they do not capture human primary tissue. As above, we are currently unable to account for the mechanisms underpinning variable penetrance. For certain individuals that are purportedly unaffected, we cannot exclude the possibility they have subclinical disease that could be detected if comprehensive clinical evaluation was performed or will develop clinical signs and symptoms in the future. Another limitation is that variable expressivity among individuals carrying the FLNC variant remains unexplained. One possibility is that the degree of NMD may have varied between individuals, partially accounting for the differences in phenotypic severity. Other possibilities include clinical and polygenic modifiers that were not ascertained in this study. A complete elucidation of the underlying arrhythmogenic

mechanism will require isogenic controls from multiple independent lines, which we were unable to provide in this clinical reclassification effort.

Conclusions. Current ACM guidelines now take into account the results of genetic testing, including the 2019 HRS Consensus Statement[164] and the Padua Criteria[185]. Reclassification of this variant will allow for genetic screening of at-risk family members thereby eliminating the need for continued follow-up in genotype-negative individuals and justifying the need for closer follow and risk reduction strategies in genotype-positive individuals. Inspection of the 4 kindreds in this report demonstrates that several individuals who are at-risk but currently unaffected may directly benefit from our reclassification efforts. Together, this work demonstrates the utility of international collaboration and data sharing in VUS reclassification, and sheds light on an interesting molecular mechanism of a splice-altering variant.

Chapter 6

ParSE-seq: A Calibrated High-throughput Assay to Facilitate Clinical Interpretation of Putative Splice-altering Variation.

Abstract

Background: Up to ten percent of disease-associated variants are predicted to affect RNA-splicing. Interpreting the clinical significance of putative splice-altering variants outside of the 2-base pair canonical splice sites remains difficult, often owing to lack of functional data.

Methods: We developed Parallel Splice Effect Sequencing (ParSE-seq) to test variant effects on RNA splicing. We developed a multiplexed minigene assay and quantified variant effects on splicing through high-throughput sequencing. We deployed the assay on a library of 271 exonic and intronic variants in *SCN5A*, which encodes the major voltage-gated cardiac sodium channel. The assay was performed in Human Embryonic Kidney (HEK293) cells and induced pluripotent stem cell cardiomyocytes (iPSC-CMs). The strength of the assay in the American College of Medical Genetics and Genomics classification scheme was determined by calibration with ClinVar-curated control variants. Variants were reclassified using the calibrated functional data. Experimental outcomes were compared to *in silico* predictions with the tool SpliceAI. We further studied missense variants by automated patch clamp and validated 2 variants by CRISPR-editing of iPSC-CMs.

Results: Our high-throughput method revealed the splicing effect of 199 variants in HEK293 cells and 180 variants in iPSC-CMs after strict filtering for multiple testing and biological effect size. In disease relevant iPSC-CMs, 41/114 exonic variants disrupted splicing and 30/66 intronic variants disrupted splicing. The assay had concordant scores for 43/45 ClinVar-annotated benign and pathogenic control variants, achieving a PS3- and BS3-strong level of evidence after calibration. 9 of 42 Variants of Uncertain Significance were reclassified, and 30 of 34 variants with conflicting interpretations were reinterpreted using the functional evidence. SpliceAI had a high sensitivity

and specificity for splice-altering variants (AUC = 0.96). We identified 5 missense variants that had minimal disruption of protein function in a cDNA-based assay but disrupted proper *SCN5A* splicing.

Conclusions: ParSE-seq is an efficient, high-throughput assay that is clinically calibrated to facilitate splice-altering variant interpretation. Some splice-altering variants may be incorrectly assigned benign functional criteria in cDNA-based assays.

Keywords: Splicing, Variant interpretation, RNA, Non-coding, Clinical genetics, Arrhythmias, iPSC-CM, SCN5A

Introduction

Understanding the clinical consequences of genetic variation is one of the greatest challenges in contemporary human genetics research. While most research has focused on understanding exonic variation at the protein level, there is an increasing need to explore variant effects on splicing[186]. Sequencing studies have shown that up to 10% of pathogenic variation may act through disruptions of splicing outside of the canonical AG-GT splice sites[64]. Despite these estimates, both intronic and exonic splicing variation has been poorly studied compared to protein-coding missense variation, largely due to the difficulty of identifying and studying this class of variants. Several assays exist for assaying splicing outcomes, including direct measurements from patient tissue, minigene assays, and CRIPSR-editing at the endogenous locus. However, these assays are typically not calibrated to putative benign and pathogenic variation and are too low-throughput to decipher the thousands of potential splice-altering variants in Mendelian disease genes. An alternative approach is high-throughput multiplexed assays that use next generation sequencing (NGS) to read out pools of splicing outcomes. Several such assays have been adopted but have focused on saturation mutagenesis of small regions of specific genes[187,188], capture incomplete variant effect through an exon couplet configuration[146,147], or are limited to small inserts by size requirements of pooled oligo synthesis[152,189]. A versatile, accurate multiplexed assay of putative splice-altering variants would greatly benefit the clinical interpretation of splicing outcomes[186].

In the 2015 American College of Medical Genetics and Genomics (ACMG) classification guidelines, “well-validated” functional data could be implemented at the strong level for abnormal assay results (PS3) or normal assay results (BS3).[43] The ClinGen Sequence Variant Interpretation Working Group developed a mathematical formulation to “calibrate” functional assays. The working group recommended that functional assays be deployed on a large number of control benign/likely benign (B/LB) and pathogenic/likely pathogenic (P/LP) variants[45,47]. The assay results for control variants could be converted into a variable range of evidence strengths

in the ACMG classification scheme, from supporting to very strong evidence, depending on the number of tested control variants and the concordance of the assay results. Several high-throughput assays of protein-coding variants have been calibrated using this scheme, but it not been applied to functional assays of splice-altering variants.

A clinically calibrated, high-throughput splicing assay would facilitate the interpretation of Variants of Uncertain Significance (VUS) and Conflicting Interpretation (CI) variants in disease-relevant genes. Herein, we studied variants in *SCN5A*, which encodes the principal voltage-gated sodium channel in the heart. Loss-of-function variants in *SCN5A* are the leading genetic cause of the arrhythmia disorder Brugada Syndrome (BrS, MIM #601144),[30] and are also associated with other cardiac diseases[190]. The majority of BrS variants reported to date are missense, frameshift, indels, or "canonical" splice site variants (*e.g.* variants in 2-bp exon-adjacent splice sites)[29]. We and other groups previously deployed low-throughput functional assays to demonstrate splicing defects in several *SCN5A* exonic and intronic variants outside the 2-bp splice sequences[143,191,192].

In this work, we present a multiplexed splicing assay, Parallel Splice Effect Sequencing (ParSE-seq), to determine the splice-altering consequences of intronic and exonic variants. We implement ParSE-seq for 271 *SCN5A* variants in Human Embryonic Kidney (HEK293) cells and induced pluripotent-derived cardiomyocyte cells (iPSC-CM). We calibrate the assay with ClinVar-annotated variants and assess our experimental outcomes against the *in silico* tool SpliceAI. Using our calibrated strength of evidence, we posit reclassifications for 9 VUS, and put forth orthogonal information to help adjudicate 30 conflicting interpretation variants in ClinVar. Furthermore, we demonstrate that some missense variants may be incorrectly described by cDNA-based assays that cannot assess splicing outcomes.

Methods

Selection of Variants. Variants were broadly selected from 3 categories – 1) Benign/likely benign and pathogenic/likely pathogenic ClinVar variants to calibrate the assay; 2) VUS and CI variants in *SCN5A* that could be reclassified with splicing functional data; 3) exonic missense and synonymous variants with aggregate SpliceAI scores greater than 0.8. Control B/LB and P/LP variants were chosen from ClinVar, accessed on October 2, 2022. We excluded B/LB variants with gnomAD allele count of 4 or less (MAF < $2.5x10^{-5}$), consistent with previous calculations of highly penetrant Brugada variants in large population databases[77,82]. All selected variants were within 200 nucleotides of the exon/intron junction. Pathogenic variants were those that were annotated in ClinVar to act through aberrant splicing rather than protein-altering variants. We used gnomAD v3.1.2 accessed on October 6th 2022, with Ensembl ID ENST00000333535.9. The assay design requires an acceptor and donor splice site on each end of the construct exon, and is therefore incompatible with the first or last coding exons (2 and 28). In addition, *SCN5A* uses 2 instances of non-canonical AC-AT splice sites between exon 3 and 4, and 25 and 26. Therefore, we did not study variants in or near these exons. Furthermore, we were unable to include constructs with exon 15 due to synthesis incompatibility (high GC content) and exon 17 due to overlap of restriction enzymes used for barcoding.

Twist Plasmid Constructs. The exon trapping vector pET01 (MoBiTec) was mutagenized to introduce two restriction sites (AscI and MfeI) in the 3' rat insulin exon to produce pAG424 (used above)[193]. Briefly, we used an inverse PCR mutagenesis method with primers ag491 and ag492 to introduce the restriction sites which were then verified by Sanger sequencing[193]. 291 "clonal genes" containing an exon surrounded by 200-300 bp of intronic sequence on each side were directly synthesized and cloned into pAG424 by Twist Biosciences (South San Francisco, CA). We resuspended each plasmid in water to a concentration of 25 ng/μl. We pooled 50 ng of each

plasmid to form a pre-barcoded library. We also manually cloned two plasmids that were incompatible with Twist synthesis using previously described methods[191].

Barcoding the Construct Library. We digested the plasmid pool with AscI (NEB) and MfeI (NEB) followed by incubation with Calf Intestinal Phosphatase (NEB). A cloning insert including an 18-mer barcode was produced using overlap-extension PCR. Briefly, ag1371 and ag1372 were annealed, followed by extension to make fully double stranded DNA using Klenow polymerase (NEB)[194]. All primers are presented in Table 1. The double stranded DNA was then phenol/chloroform extracted and digested using AscI and MfeI. The ligation product was PCR purified (QIAGEN) and electroporated into electrocompetent cells. The resulting bacterial culture was then grown overnight, and DNA was isolated by a maxiprep (QIAGEN) to yield the barcoded plasmid library. Barcode diversity was estimated by plating dilutions of the library on ampicillin plates and counting colonies.

Assembly Steps. We deciphered the relationship between barcodes and inserts (WT or variant exon and introns) using PacBio long-read SMRT sequencing[195]. We used the pooled barcoded library as template for a touchdown PCR with primers mo37 and ag489 using Q5 polymerase (NEB) per manufacture's protocol. Multiple reactions were pooled and cycle number was reduced (20 cycles) to minimize PCR-mediated recombination[196]. We proceeded with library preparation according to manufacturer's instructions with a SMRT Bell 3.0 library kit (PacBio). Amplicons were sequenced with PacBio Sequel II 8M SMRT Cell. FASTQ files were analyzed using Unix and R scripts.

Splicing Assay and Illumina Library Prep. HEK cells were grown in "HEK media": Dulbecco's Eagle's medium supplemented with 10% fetal bovine serum, 1% non-essential amino acids, and 1% penicillin/streptomycin. For the HEK cell assay, 3 wells of a 6 well plate with HEK cells at 30-

40% confluency were independently transfected with the barcoded plasmid library using FuGENE 6 (Promega) per manufacturer's instructions. For iPSC-CM transfections, we grew and maintained the 'C2' line as previously described[169]. iPSCs were differentiated into a monolayer of iPSC-CMs using a chemical differentiation method[197]. Following 30 days of differentiation, iPSC-CMs were dissociated with TrypLE™ Select (ThermoFisher) and replated as single cells. Cells were cultured for an additional 2 days, after which the pooled library was transfected using Viafect (Promega)[198]. For both HEK293 and iPSC-CM experiments, RNA was isolated 24 hours post-transfection using the RNeasy Plus Mini kit (QIAGEN) and reverse transcribed using the primer mo38 (which targets a constant region of the transcribed minigene) with SuperScript III (Invitrogen). Libraries were prepared for Illumina sequencing by touchdown PCR with primers binding reference minigene exons, sequence containing Illumina i7 and i5 indexing sequences and Illumina NovaSeq dual indexing motifs. PCR amplicons were purified with a PCR purification kit (QIAGEN). Libraries were then sequenced using Illumina NovaSeq paired-end 150 base sequencing to ~50M reads/sample.

Computational Pipeline – Assembly. The full computational pipeline will be made available on GitHub upon publication. FASTQ files were processed and parsed for barcodes with specified prefixes and suffixes (details to be available on GitHub) using Unix and Python. Unique barcodes were identified, and then reads were parsed by barcode into new FASTQ files. For each barcode, the Unix *grep* command was used to count the number of reads containing each designed insert. To be included in the assembly, a barcode was required to be present in at least 50 reads in the subassembly sequencing. The barcode identity was assigned as the most frequently aligned insert if that insert represented more than 0.50 of the normalized read counts. After implementation of these quality control cutoffs, 284 of the 291 targeted plasmids were successfully detected in the subassembly.

Computational Pipeline – Assay. Barcodes were parsed as above. In all splicing assay studies, each barcode was required to be present in at least 25 reads in each replicate to be included. The percent spliced in (PSI) metric was conducted using *grep* searches for splice junctions corresponding to the WT exon splicing to the reference exons in the R1 and R2 reads:

$$PSI_{barcode} = \min(PSI_{barcode,\ R1},\ PSI_{barcode,\ R2})/\text{total reads}_{barcode}$$

For variants that would alter the coding sequence of the WT exon, a bespoke R1 or R2 junction was created and used for those specific variants. PSIs were then averaged across barcodes for each variant, and then compared among individual replicates. For each variant, a ΔPSI value was calculated:

$$\Delta PSI_{variant} = \text{mean}(PSI_{barcode,WT}) - \text{mean}(PSI_{barcode,variant})$$

To normalize these scores, a normalized ΔPSI was then calculated

$$\Delta\text{norm\_}PSI_{variant} = \Delta\text{\_}PSI_{variant} / \text{mean}(PSI_{barcode,WT})$$

We then generated a data frame that combined the lookup table of barcode assignments with the PSI values for each barcode.

Statistical Analysis and Data Availability. Statistical analyses were performed in R. Comparison of WT and variant PSI (using mean of three replicate samples) was completed using a 2-sided t-test in R. For each variant, false discovery rates (FDR) were calculated using the R command *p.adjust*. Variants with FDR < 0.1 and ΔPSI_norm > 50% were considered splice-altering. Variants with FDR > 0.1 and ΔPSI_norm =< 20% were considered non-splice altering. All other variants were labeled indeterminate in the assay.

ACMG OddsPath Calculation. We performed assay calibration according to the framework developed by Brnich et. al[47] and extended by Fayer et al[199]. For the high-throughput splicing assay, we calculate the likelihood ratio of pathogenicity (termed OddsPath) for both splice-altering

(pathogenic) and non-splice-altering (benign) assay results. We calculated a Benign and Pathogenic OddsPath using the equation:

$$OddsPath = \frac{P2\ x(1-P1)}{(1-P2)\ x\ P1}$$

Where, the Prior P1 is defined as,

$$P1 = \frac{\#P/LP\ Controls}{\#\ Total\ Controls}$$

And the two posteriors $P2_{Pathogenic}$ and $P2_{Benign}$ are defined as,

$$P2 = \begin{cases} \frac{\#\ Functionally\ abnormal}{\#\ Functionally\ abnormal + 1}, Pathogenic \\ \frac{1}{\#\ Functionally\ Normal + 1}, Benign \end{cases}$$

Each posterior is combined with the prior to derive an OddsPath and assign evidence for PS3 and BS3 criteria, respectively.

SpliceAI Implementation. SpliceAI was implemented on the Unix command line using an input VCF file consisting of HGVS GRCh38 coordinates for all variant positions in the splicing library[64]. Pre-computed SpliceAI scores hosted at Illumina BaseSpace were used to identify exonic variants with aggregate scores above a threshold of 0.8; we computed an aggregate SpliceAI score to incorporate contributions from each of the 4 predicted categories using a previously described formula[45]:

P(aberrant splicing) = 1 – ((1-(AG)) * (1-(AL)) * (1-(DG)) * (1-(DL))

Where AG=probability of acceptor gain, AL=probability of acceptor loss, DG=probability of donor gain, and DL=probability of donor loss. The Receiver operator characteristic curves were completed in R Studio using the library pROC, evaluating aggregate SpliceAI scores for the binary outcome of normal or abnormal splice assay result[200].

ACMG Variant Classification. We provided clinical interpretations of variants using the ACMG criteria using an online tool hosted by the University of Maryland[201]. We implemented functional evidence at the calibrated strength of evidence based on the results of the calibration described above (PS3 strong and BS3 strong). We applied PP3 for an aggregate SpliceAI >0.5, and BP3 for SpliceAI <0.2. We applied PM2 for variants with an allele frequency of < $2.5x10^{-5}$ per recommendations for BrS[82].

High-throughput Electrophysiology. We used an established workflow for generating automated patch clamp data for variants in HEK293T landing pad cells (gift of Kenneth Matreyek).[76,78,79] Briefly, we cloned SCN5A variants into cDNA constructs using Quikchange mutagenesis (Agilent), integrated these constructs into landing pad cells, used negative and positive selection with iCasp and blasticidin to generate stable lines, and studied the stable lines on the SyncroPatch 384PE automated patch clamping device[202] (Nanion). This cloning and data analysis process has been extensively described by our group in previous publications[51,76,202].

iPSC-CM Maintenance and Differentiation. iPSCs were cultured at 37°C in a humidified 95% air/5% CO_2 incubator on Matrigel coated plates (BD BioSciences). Cells were maintained in mTeSR plus media (STEMCELL) and passaged every 3-4 days. Stem cell experiments were carried out in the 'C2' line previously characterized by our laboratory. At 60-80% confluency, stem cells began CM differentiation using a previously described chemical method[44]. RNA analyses and patch-clamp were completed at days 35-40 of chemical differentiation.

CRISPR-Cas9 Gene editing. We designed CRISPR-Cas9 guide constructs based on the 20-bp NGG Sp Cas9 editor using the online tool CRISPOR[203]. PAM sites were altered in the repair template to a synonymous amino acid for exonic variants and a random SNV when intronic; in both cases, the disrupting variant was predicted to have minimum impact on splicing (aggregate

SpliceAI <0.05). CRISPR guides were cloned into (pX458)[20] plasmid (Addgene #48138, a gift of Feng Zhang) as previously described. The cloned guide plasmid and a 151-nucleotide repair template bearing the desired change and PAM site variant were co-electroporated into dissociated iPSCs using the Neon Transfection System (ThermoFisher MPK5000). After 72 hours, cells were dissociated, singularized, and sorted for GFP using a BD Fortessa 5-laser instrument. Single colonies were picked and genotyped for heterozygous edits.

Results

Methodology Overview and Feasibility. We envisioned a high-throughput minigene-based assay that would 1) build upon manually intensive single-variant minigene studies used for clinical variant interpretation, and 2) extend previous high-throughput assays by establishing a platform to investigate variants for potential clinical relevance. Variants are known to disrupt RNA-splicing through disruption of many splice regulatory elements (Figure 6.1A). Our design uses high-throughput sequencing to quantify splice alterations as changes in Percent Spliced In (PSI) of the

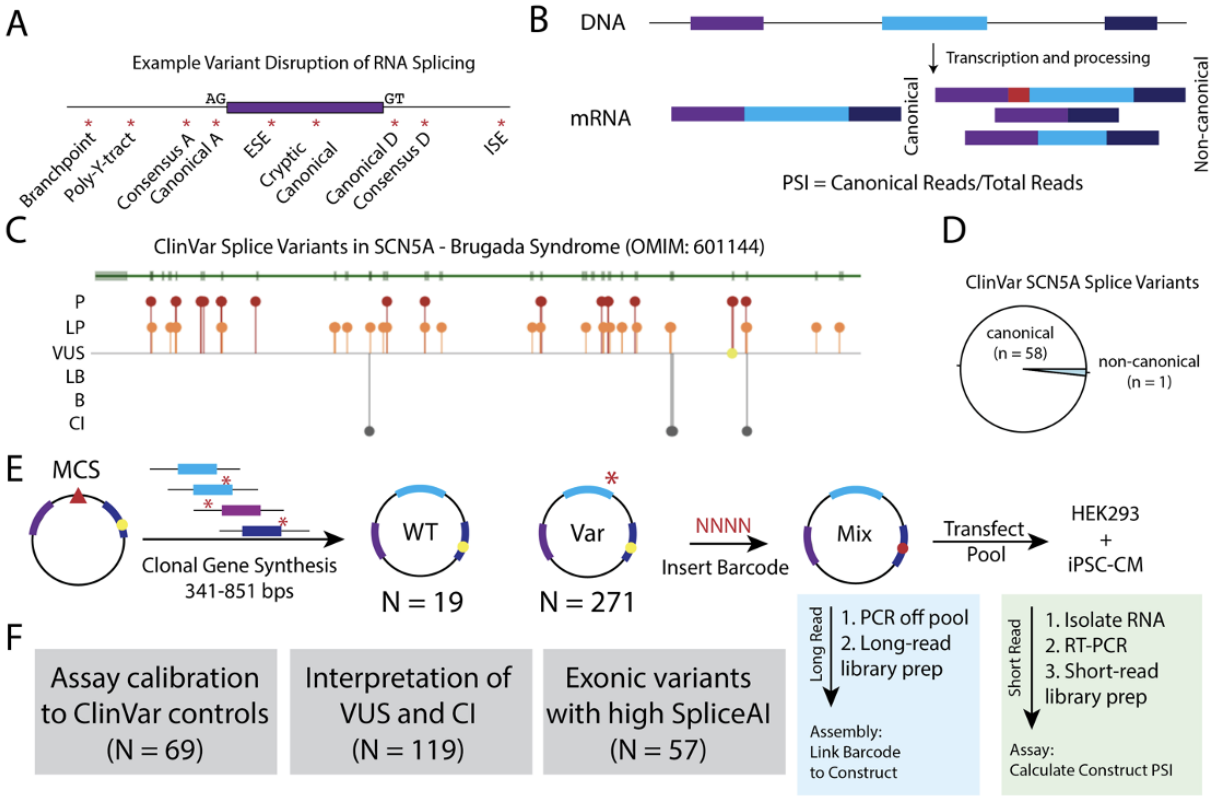


Figure 6.1. Splice-altering Variation and Assay Schematic.
A) Splicing regulatory sequences disrupted or introduced by *cis*-genetic variants. A – acceptor site. D – donor site. ESE – exonic splicing enhancer. ISE – intronic splicing enhancer. Y – pyrimidine.
B) Quantification of Percent Spliced In (PSI) from transcripts associated with WT- or variant-containing transcripts. Canonical reads are divided by the total amount of reads for a given exon triplet cassette.
C) ClinVar view of splice-altering variants for the gene *SCN5A*. Most known splice variants are P/LP.
D) Most *SCN5A* splice variants in ClinVar are associated with canonical splice sites, with rare non-canonical splice variants.
E) Schematic of ParSE-seq assay. A clonal gene library is cloned into a minigene vector, pooled, and barcoded. The barcode is assigned to the construct by long-read sequencing through assembly, and the splicing outcomes are determined with short-read sequencing after transfection into cells. Yellow circle represents restriction site for barcode in downstream exon.
F) The *SCN5A* variant library was chosen to evaluate 3 categories of variants – assay calibration, interpretation of clinically relevant variants, and to identify splice-altering exonic variation. VUS – Variant of Uncertain Significance. CI – Conflicting Interpretation.

canonical triplet cassette between variant and wild-type (WT) constructs (Figure 6.1B). We designed a library to test the splicing effect of variants in *SCN5A*. Most ClinVar annotated splice-altering variants in *SCN5A* are at the canonical splice site, motivating us to explore other variants that may affect splicing in this gene (Figures 6.1C and 6.1D). Towards this goal, we introduced two restriction sites into the 3' exon of the minigene vector pET01 to allow for a barcode to be read in spliced constructs[191] (Figure 6.1E; see Methods for full details). This design enabled the linkage of reads to variants even if the variant was not included in the spliced transcript.

ParSE-seq Assay Implementation. We designed a library of variants to test three questions (Figure 6.1F) – 1) how well does the assay perform on reported ClinVar P/LP and B/LB control variants?; 2) can these results help interpret VUS or provide orthogonal information to resolve variants with conflicting interpretations (CI)?; and 3) what is the prospective accuracy of SpliceAI on exonic synonymous and missense variation? We generated a plasmid library consisting of 291 wild-type (WT) and variant constructs to address these questions. After introducing a random barcode, we completed assembly with long read sequencing to link each barcode to the insert. We identified 3,238 unique barcodes linked to 284 unique inserts (median 11 barcodes per insert). Barcode frequencies were highly correlated when counted by both PacBio and Illumina sequencing (R = 0.84, p-value <0.01).

We transfected HEK cells and iPSC-CM cells each in triplicate with the barcoded library, isolated RNA, performed RT-PCR, and used targeted RNA-seq to quantify the impact of variants on splicing. After quality control, we recovered splicing data in HEK cells for 2,218 barcodes associated with 263 unique inserts (244 variants, 19 WT exons, Figure 6.2A). Although the transfection efficiency in iPSC-CMs was low (<2%), we nonetheless recovered splicing data for a total of 927 unique barcodes associated with 257 unique constructs (238 variants, 19 WT exons; Figure 6.2A). We calculated a PSI for each insert and calculated the ΔPSI for each variant by

comparing its PSI to that of its corresponding WT insert. WT constructs generally had higher PSIs in iPSC-CMs than in HEK293T cells (Figure 6.2B).

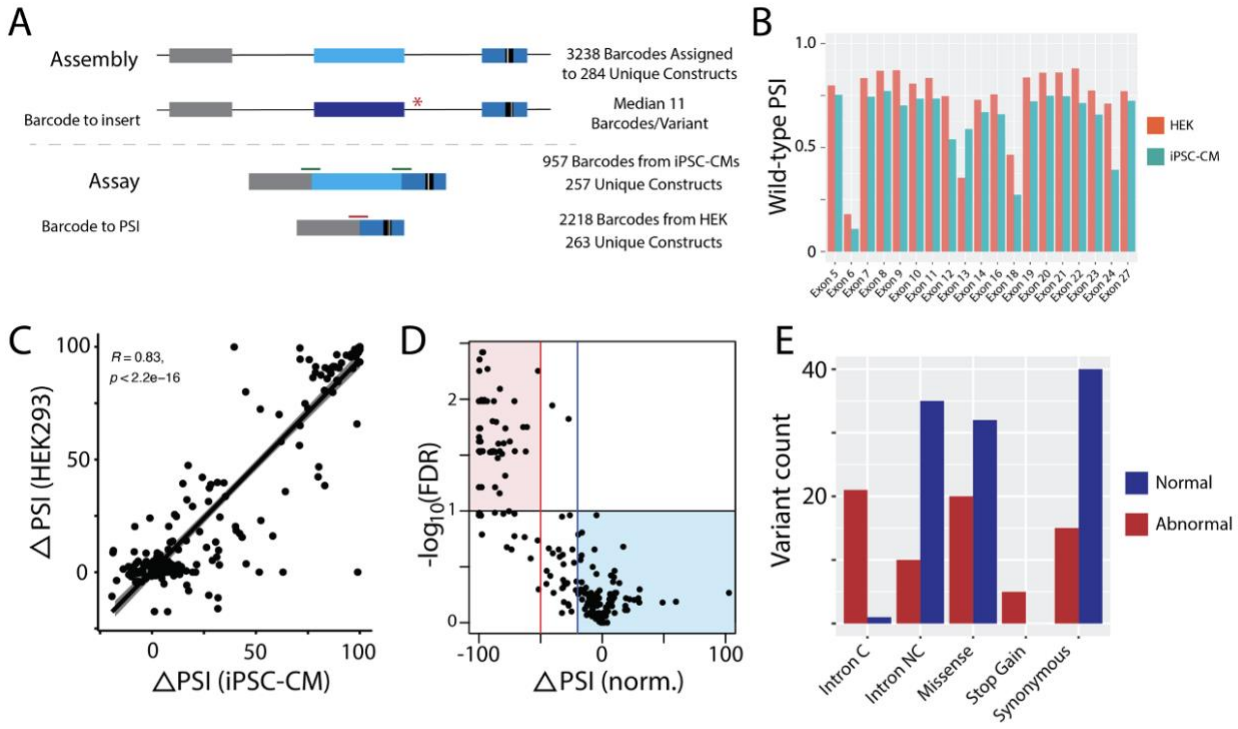


Figure 6.2. Results of assay in HEK cells and iPSC-CMs.
A) Schematic of assay and assembly with results for barcode recovery in HEK and iPSC-CMs.
B) Wild-type construct PSI in HEK and iPSC-CMs.
C) Correlation of the ΔPSI in HEK and iPSC-CMs.
D) Volcano plot of normalized ΔPSI in iPSC-CMs and -$\log_{10}$(FDR). Variants (dots) in red shading were functionally abnormal, and variants in blue functionally normal. All others were considered indeterminant.
E) Counts of normal and abnormal variants among different variant classes. C – canonical, NC – non-canonical.

We considered variants with a false discovery rate (FDR) < 0.1 and a normalized ΔPSI > 50% to be 'abnormal', and variants to be 'normal' with an FDR > 0.1 and change in normalized ΔPSI < 20%. We observed excellent correlation of normalized ΔPSI scores between the assay in HEK cells and iPSC-CMs (R^2 = 0.69, p < 2.2 * 10^{-16}). 105/94/64 variants had normal/abnormal/indeterminate splicing in the HEK293 assay, and 109/71/58 variants had normal/abnormal/indeterminate splicing in the iPSC-CM assay (Figure 6.2D). Figure 6.2E shows the results for variant class across the library, with 20/21 2-bp canonical splice sites causing abnormal splicing, and a mix of results among missense, synonymous, and non-canonical splice

site intronic variants. The 41 exonic splice-disrupting variants are spread throughout the protein, but often cluster in hotspots near exon boundaries.

Functional Studies and SpliceAI. To determine the concordance between our experimental and the *in silico* predictor SpliceAI[64], we calculated scores for all variants in our library. In Figure 6.3A, we plot the distributions of the aggregate SpliceAI scores calculated for each variant in unique ClinVar classes. Variants with higher SpliceAI scores were more likely be disease-associated, with an intermediate distribution for VUS and CI variants. In Figure 6.3B, we plot normalized ΔPSI scores against aggregate SpliceAI scores. We observe a high correlation between SpliceAI scores and negative changes in PSI (R^2 = 0.58, $p < 2.2 * 10^{-16}$). Assessing the dichotomous outcome of normal or abnormal splicing, we observed high sensitivity and specificity of SpliceAI predictions, achieving an AUC of 0.960 across the full library (Figure 6.3C). Notably, there was high concordance of scores and experimental data when selecting for intronic variants outside of the canonical splice sites (0.998) and exonic variants (0.954). Recent studies have explored the ability of SpliceAI to retrospectively predict variant effect based on small-medium sets of functionally profiled splicing variants[65]. To further explore the robustness of this *in silico* tool, we prospectively evaluated the ability of SpliceAI to predict splice-disrupting variants among exonic variants. As described above, 42 prospectively identified variants with pre-computed SpliceAI scores >0.80 were adjudicated in the iPSC-CM assay, of which two were observed in ClinVar, and none observed in gnomAD. Parse-seq outcomes of these variants are shown in Figure 6.3D, suggesting a high positive predictive value of SpliceAI-prioritized variants. Notably, we observe that 19 of 21 missense variants effected splicing.

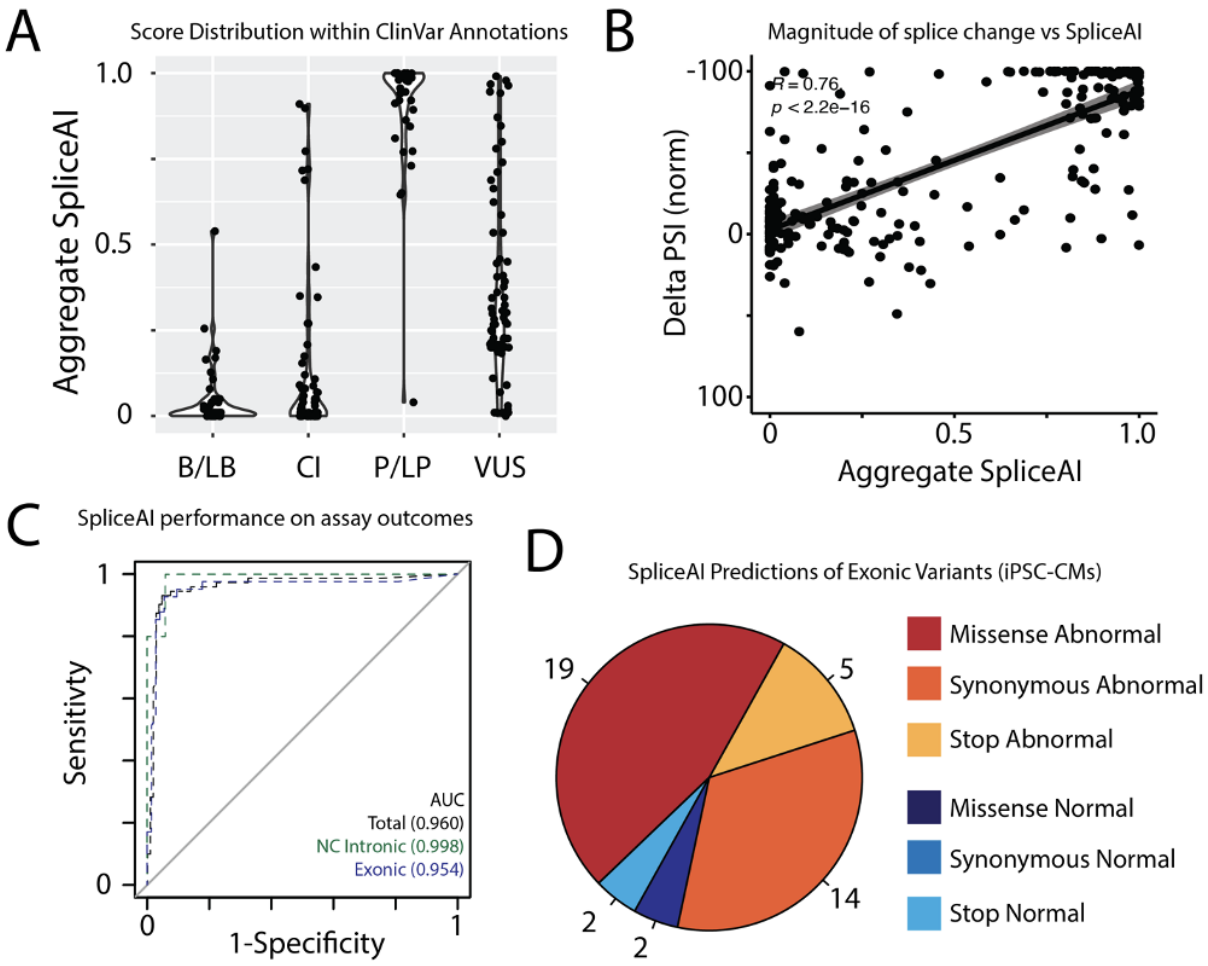


Figure 6.3. Comparison of experimental data and *in silico* SpliceAI scores.
A) Aggregate SpliceAI scores for each ClinVar variant class.
B) Correlation of change in PSI against aggregate SpliceAI score.
C) Receiver operating characteristic curve for SpliceAI applied to normal and abnormal variants.
D) Results of prospectively identified variants by SpliceAI score >0.8.

ACMG Assay Calibration. We tested known control variants to provide a calibrated strength of evidence for the assay. We followed the ClinGen Sequence Variant Interpretation Working Group scheme for converting control assay results into likelihood ratios of pathogenicity (OddsPath; Figure 6.4A)[47]. A total of 45 ClinVar B/LB and P/LP control variants were recovered from our splicing experiment in iPSC-CMs with clear normal or abnormal splicing results. We observed a near-perfect concordance of ClinVar classifications and assay outcomes: 22/23 B/LB variants had

normal splicing, and 21/22 P/LP variants had abnormal splicing (Figure 6.4B). Applying these values, the OddsPath of abnormal assay results was 21.2 and the OddsPath of normal assay results was 0.048. These values correspond to strong application of abnormal functional evidence towards pathogenicity (PS3) and strong application of normal functional evidence towards a benign classification (BS3).

Variant Interpretation and Reclassification. After identifying a calibrated strength of evidence to implement functional data (PS3 and BS3), we sought to provide updated clinical interpretations for 42 VUS and 34 CI variants for which we obtained clear results in the iPSC-CM assay. We identified 8 VUS that were abnormal in our splicing assay (5 exonic, 3 intronic). Applying these functional data and *in silico* predictions from SpliceAI, we were able to reclassify 7 to LP and

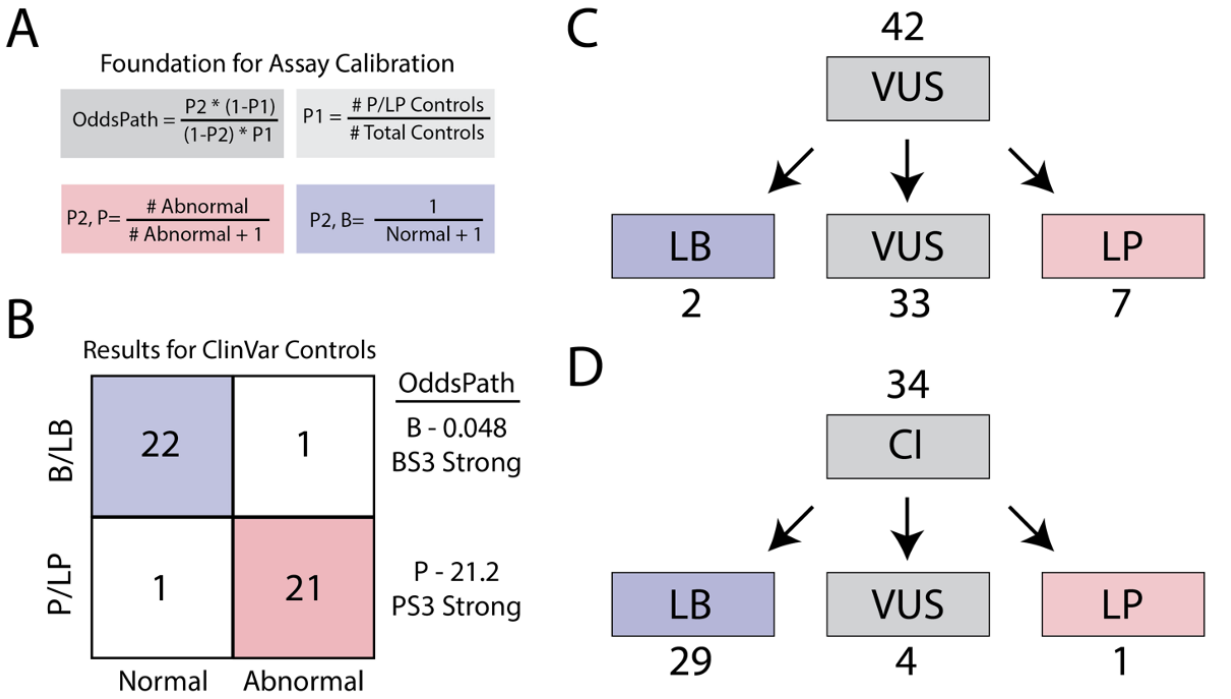


Figure 6.4. Assay Calibration and Variant Interpretation.
A) Framework for calibrating assay strength of evidence obtained from Brnich et. al.
B) ParSE-seq results for B/LB and P/LP controls in iPSC-CMs. We obtain 2 posteriors to implement BS3 and PS3, both at strong level of evidence.
C) VUS are reclassified using functional data from ParSE-seq at the strong level of evidence. BS3 evidence is applied exclusively to synonymous and intronic VUS. PS3 evidence is applied to all VUS.
D) Assertions for Conflicting Interpretation variants using functional evidence. BS3/PS3 applied as for VUS in C.

demote 2 variants to LB. We limited application of benign functional data to synonymous and intronic variants, as we were unable to rule out dysfunction of the translated protein product for missense variants from the ParSE-seq assay alone. Conflicting interpretation variants are a growing fraction of variants on ClinVar, often resulting from different thresholds applied to the same underlying data. We anticipated that functional data would help provide orthogonal data to adjudicate conflicting interpretation variants, especially for synonymous and intronic variants. Applying the ParSE-seq results from iPSC-CMs, we found that 29 of 34 CI variants could be classified as LB. We upgraded 1 CI variant to LP due to its splice-disrupting effect.

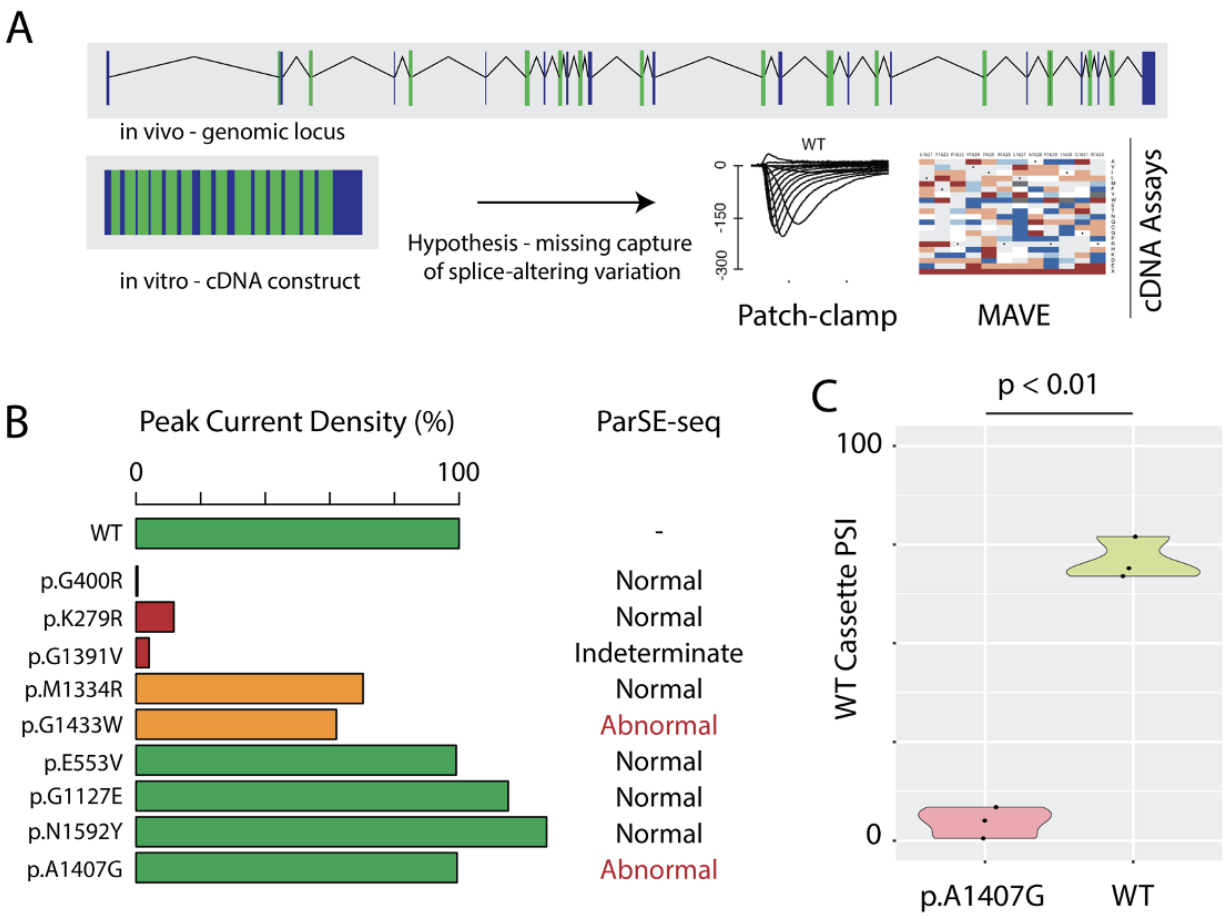


Figure 6.5: Test of the hypothesis that exonic variants can be pathogenic by altering splicing.
A) Schematic of cDNA based assays versus the *in vivo* gene locus. Opportunity to identify 'missing' fraction of variants in genotype-negative cases.
B) Normalized peak current for 9 missense variants with SpliceAI >0.35.
C) Data demonstrating a profound splicing defect (triplicate assays) for a missense variant that displayed normal electrophysiology. Variant and WT inserts.

Cryptic splicing effects of exonic variants. Missense variants that disrupt gene function and lead to disease are usually assumed to disrupt protein function. As such, functional assays of missense variants (e.g., electrophysiology of *SCN5A* variants) are generally performed using cDNA, which obscures effects on splicing. We hypothesized that some missense variants may cause Mendelian disorders such as BrS through an aberrant splicing mechanism (Figure 6.5A). To test this hypothesis, we identified 9 tested in ParSE-seq and then by automated-patchclamp that 1) had SpliceAI scores > 0.35, and 2) at least one BrS heterozygote reported[69]. Of these 9 variants, 4 had peak sodium currents in the normal range (75%-125% of WT), 3 had loss of function (<15% peak current), and 2 had minimal loss-of-function (50-75% of WT) (Figure 6.5B). Two missense variants, p.A1407G and p.R1433W, disrupted splicing but have normal, or close to normal, electrophysiologic function when tested using cDNA(Figure 6.5C). For these variants, cDNA-based assays of protein function would yield incorrect conclusions about variant effects on function.

Recapitulating Splice Effects at the Endogenous Locus. Modeling cardiac disease has been greatly advanced by joint developments in gene editing and iPSC-CM technologies[204]. To test how well our *in vitro* splicing assay predicted variant effect in a more native state, we used CRISPR-Cas9 to introduce the *SCN5A* intronic VUS c.1891-5A>G (ClinVar: 201463) and the synonymous VUS c.1338G>A (ClinVar: 451631) as edits in a healthy control iPSC ‘C2’ line. Stable edits in differentiated iPSC-CMs would allow us to query variant effect on transcript composition, and also allow us to link transcript-level defects to sodium channel function in a pseudoprimary tissue. We are currently differentiating these edited iPSC lines for RNA-seq and manual patch-clamp experiments.

Discussion

ParSE-seq Summary. In this study we developed a high-throughput method to assess the splicing consequences of hundreds of variants. We implemented barcoding of a minigene system to enable multiplexed splicing readouts using high-throughput sequencing and applied it to the cardiac sodium channel gene *SCN5A*. In disease-relevant iPSC-CMs, we determined variant-induced changes in splicing for 180 variants, detecting 71 variants with abnormal splicing. Using a large set of B/LB and P/LP variants, we determined that our assay could be applied at the strong level in the ACMG classification scheme (BS3 and PS3). Using these calibrated strengths of evidence, we interpreted VUS and CI *SCN5A* variants. We further determined that the *in silico* tool SpliceAI has excellent (but not perfect) concordance with experimentally measured splicing effects. We also demonstrated examples of missense variants that had normal electrophysiology using conventional heterologous expression cDNA-base approaches, but disrupt splicing. We envision that ParSE-seq will be applicable to any disease gene and will be accessible using openly available computational pipelines and democratized gene synthesis available to the community.

Incorporating Splicing Assays into the Genetic Testing Pipeline. Calibrated splicing assays can improve variant interpretation and support the implementation of personalized medicine[186]. Critically, our assay identified 41 exonic variants and 10 intronic variants outside the canonical splice site which disrupt splicing in iPSC-CMs. This data is of immediate clinical relevance to carriers of these variants, who should be screened for BrS or other conduction defects. In addition, we identify 31 VUS or CI variants that could be classified as LB after incorporating our splicing data, which reduces clinical uncertainty for these variant carriers[186].

Recently, standardization of techniques and harmonization of interpretation has improved the clinical integration of splicing results from primary or peripheral tissue[205]. We imagine that lab-

based methods like ParSE-seq and clinical studies of peripheral tissue will be complementary. *In vitro* functional data will be especially useful when tissue is not available, when there are no other heterozygotes to segregate a phenotype, or expression of the relevant transcript is not sufficiently high in the clinically accessible tissue. Clinical potential will be further improved by the united integration of both improved functional assays as well as AI-based *in silico* tools for splicing predictions[64,206].

Comparison to Previous High-throughput Assays. ParSE-seq builds on pioneering work in the field of high-throughput splicing assays. For example, two excellent examples of multiplexed readouts were recently applied to *LMNA* and *MYBPC3*, and *TTN* variants using gene blocks and barcoding[146,147]. While multiplexed methods have been developed to measure variant splice effect, these have often been limited in read out or constructs. Although previous studies have examined thousands of multiplexed splice outcomes, these splicing libraries are typically limited to only small exons with minimal adjacent intronic sequence due to the limits of chip-based oligo synthesis[152,189,207,208]. This often precludes studying clinically relevant exons due to length and size requirements. ParSE-seq builds on this by considering a more complete set of aberrant splicing patterns through the triplet exon cassette design, as well as rendering the testing of much larger clinically relevant constructs feasible with commercially-available clonal gene synthesis technology. Furthermore, capturing larger segments of flanking introns has been shown to increase the assay validity due to more complete capture of *cis*-regulatory elements[152]. A major advantage of this framework is the standardization of library preparation, wherein a library can quickly be assembled using our provided scripts and genomic coordinates of variants of interest. To our knowledge, none of these previous methods have been calibrated using known B/LB and P/LP standards for clinical variant interpretation in the ACMG scheme. Finally, we perform the assay in a disease-relevant cell type (iPSC-CMs).

ParSE-seq may also complement deep mutational scans, high-throughput scans of coding variants[209]. These high-throughput assays subject libraries of many variants at each position of a protein to a fitness assay, which allows prospective functional evaluation of a variant even before it is clinically observed. This approach holds great promise for reducing the VUS problem[44]. One limitation is that many previous approaches have used cDNA-based libraries. In some cases, investigators have excluded their functional results for variants with SpliceAI scores above a certain threshold[210], or assumed deleterious splicing based on prediction-alone when applying results to clinical cohorts[211]. Here, we show the feasibility of obtaining prospective splicing results of many exonic variants prioritized by an *in silico* predictor. For example, obtaining splicing data for putative splice-altering variants in the channelopathy genes could complement cDNA-based MAVEs in progress for these genes[55,56]. Notably, this complementary approach would not be as feasible using constructs derived from chip-based oligonucleotide synthesis due to size restrictions.

Future Directions. We anticipate that ParSE-seq will be a useful method for many applications in clinical genetics. Given the plethora of VUS and CI variants that may act through splice disruption[186], investigators may use this method to quickly facilitate the interpretation of larger sets of variants in disease-associated genes, agnostic to disease category.

Limitations. Although the cost of gene synthesis is falling, it remains an economic barrier to genome-wide variant profiling. We anticipate with decreasing costs, larger scale implementations of ParSE-seq will be available for focused clinical libraries up to orders of magnitude more, but likely never feasible for saturation mapping of all possible variant splicing effects across the clinically-relevant genome. The library design is fundamentally constrained by the requirements of transfection, with more clinically relevant cell types often being recalcitrant to larger cargos. While not expected to alter the underlying biology, all variants in this study were purposely

designed in a single gene, *SCN5A.* Lastly, there may be examples of compensatory mechanisms within cells that overcome an *in vitro* aberration, where the assay may not fully capture all nuances of endogenous locus biology.

Conclusions. We describe ParSE-seq, an ACMG-calibrated high-throughput splicing assay. We anticipate that ParSE-seq will be a useful method for many applications in clinical genetics. Given the plethora of VUS and CI variants that may act through disruptions of splicing, investigators may use this method to quickly facilitate the interpretation of larger sets of variants in disease-associated genes.

Chapter 7

Synthesis of the Thesis

Summary

Understanding the clinical implications of genetic variation can be greatly enhanced by delineating the functional impact of such variation at the molecular level[44]. The goal of my thesis was to develop and apply functional methods to understand genetic variation in clinically relevant arrhythmia genes, and then integrate these findings into variant interpretation frameworks. This work spanned the spectrum of translational science, including experimental, computational, and clinical approaches.

My work began with improving our understanding of missense variation in the sodium channel-encoding gene *SCN5A* by using high-throughput APC to reveal widespread dominant negative effects among loss-of-function variants (Chapter 2). To integrate these APC functional data into quantitative variant interpretation frameworks, I extended Bayesian penetrance methods to capture incomplete penetrance among all possible SNV variation in *KCNQ1* and *SCN5A*, representing a departure from coarse categorical variant classification (Chapter 3). Next, I studied exonic and intronic VUS thought to affect splicing in the arrhythmia genes *SCN5A*, *KCNQ1*, *KCNH2*, and *FLNC.* First, I employed transcriptional studies of minigenes and CRISPR-Cas9 editing of iPSCs to study the impact of putative splice-altering variants in channelopathy genes, and then reclassified these variants using those data (Chapter 4). Next, I leveraged clinical observations of arrhythmogenic cardiomyopathy probands who were heterozygous for a recurrent *FLNC* intronic variant (Chapter 5). I used clinical co-segregation, peripheral blood splicing assays, iPSC-CMs transcriptional studies, and patch-clamp electrophysiology, to collectively reclassify the c.970-4A>G variant from VUS to P. Lastly, I developed a multiplexed minigene splicing assay in a new method called ParSE-seq (Chapter 6). I deployed the method on a library of *SCN5A* variants, and revealed many variants that affected splicing.

Synthesis of Efforts

Assay Development. Chapter 2 and Chapter 6 provided new assays to facilitate the functional interrogation of genetic variation. While individual cases of dominant negative SCN5A variants had been demonstrated before, the work in Chapter 2 provided a proof-of-concept for studying *SCN5A* variants in a *trans*-orientation from two independently, stably integrated loci. I used the Sleeping Beauty transposon system[80] in conjunction with a previously developed 'Landing Pad' cell line[79]. By optimizing integration efficiency, we attained nearly equivalent activity of $Na_V1.5$ from each locus. These experiments were an impetus for further exploring dominant negative effects in our lab in medium- to high-throughput, including an assay for comprehensively testing *KCNQ1* dominant negative effects of *KCNQ1* variants.

Similarly, ParSE-seq represents a new method that we anticipate will be of broad use to other investigators. Adoption will be facilitated by the democratization of library preparation by using clonal gene synthesis, availability of code used to analyze the datasets, and extensive examples of implementation. The method improves upon the previous state-of-the-art by incorporating clinically relevant exon/intron constructs through clonal gene synthesis rather than small inserts available through chip-based oligonucleotide synthesis[147,189]. Moreover, we demonstrate the feasibility of capturing the *trans*-acting splicing milieu by studying the splicing library in a disease-relevant cell type. Lastly, this work meets the high specifications of assay calibration proposed by the ClinGen Sequence Variant Interpretation group for the interpretation of variant effect[47].

Contributions of Splice-altering Variants in Inherited Arrhythmia Syndromes. Chapters 4, 5, and 6 collectively investigated putative splice-altering variants in cardiac disease genes. Together, these chapters revealed that many intronic and exonic variants may provoke disease through aberrant splicing. In Chapter 4, I evaluated 13 VUS in a reported clinical cohort. In Chapter 5, I

leveraged international clinical collaborations and multiple transcriptional and whole-cell functional studies to understand a recurrent VUS. In Chapter 6, I developed a general multiplexed method to study variant splice effects at larger scale. In aggregate, these investigations support the consideration of variant splice-effects when investigating genetic disease. Furthermore, these studies caution that some cDNA-based assays may return false positives (e.g. using electrophysiology approaches to study cDNAs in heterologous expression systems), as these variants may nonetheless drive disease through aberrant splicing.

Clinical actionability. A major goal of this work was to use *in vitro* functional studies to provide knowledge that can impact patient management. The inherited arrhythmia syndromes are an especially poignant example of the utility of genotype-first approaches to precision medicine[50]. Since sudden cardiac death may be the first manifestation of disease, genetics-guided diagnosis and care can help prevent the lethal arrhythmias associated with disease. In the future, individuals may be sequenced, and those with variants in disease-related genes provided with personalized care based on their genotypes. By providing variant functional data in Chapters 2, 4, 5, and 6, we contribute to helping realize this goal for those participants that will be discovered to harbor these variants. More broadly, in Chapter 3, I provide a framework for further enabling genotype-first precision medicine in a more quantitative manner that explicitly models the incomplete penetrance of variants.

Collectively, these studies improve our understanding of which genetic variation in inherited arrhythmia syndrome-related genes contribute to clinical phenotypes, and provide clinically-actionable knowledge for affected patients. This work will catalyze future studies that deploy these methods to characterize variant effects across a spectrum of diseases.

Limitations

Model Systems. As with many *in vitro* studies, we employed model systems that incompletely capture the tissue-level effects of variants in the human heart. For example, much of this work was performed in Human Embryonic Kidney (HEK293) cells, which do not express the same genes as mature cardiomyocytes. To address this, I performed some of the function studies in induced pluripotent stem cell-derived cardiomyocytes (iPSC-CM). These 'pseudo'-primary cells nonetheless resemble an 'immature' cardiomyocyte state, and fail to integrate complex signaling to the heart such as adrenergic feedback that modulates cardiac repolarization. Recent advances in cardiac organoid engineering may offer an opportunity to push these efforts forward to an increasingly realistic model of the adult human heart. Additional cellular engineering will still be required to enable multiplexed variant investigations in iPSC-CMs using the methods developed herein.

Recapitulating the Endogenous Locus. Most of the variants investigated throughout the functional studies were expressed outside of their endogenous locus. For example, the dominant negative Syncropatch studies used integration at a landing pad site and a random integration for the two copies of *SCN5A* cDNA, rather than editing of the endogenous locus of *SCN5A*. While this configuration allowed us to understand the interaction at the protein level, it may incompletely describe variant effect due to confounding from native compensatory mechanisms such as alternative splicing, enhancer interactions, or epigenetic regulation of one locus. The majority of splicing variants relied on transient transfection of minigene constructs bearing ~100-400 nucleotides of *cis*-genetic context on each side of the variant of interest. In this case, effects of additional *cis*-regulatory elements contained in the larger endogenous intronic segments may be missed, or a putative splice-altering event may act through a distinct mechanism incompletely captured by *in vitro* splice-assay. To address this, we studied the splicing effects of 6 variants in

the native context in iPSC-CMs using Cas9-gene editing. These results were largely consistent with minigene outcomes.

Incomplete Capture of Additional Modifiers of Penetrance. This thesis has necessarily focused largely on the molecular characterization of rare, large-effect size variants characteristic of 'monogenic' disease. Most of these variants were first observed in the literature or public databases such as ClinVar, and therefore few details of the patient's history and additional information were available. Genetics is fundamentally non-deterministic, and variants with normal or abnormal function do not act in isolation (see Figure 3.7 for incomplete penetrance of P/LP variants). For example, the penetrance of a variant can be modified by the common variant background, which can be quantified in part by polygenic risk scores[22] In addition, demographic and environmental exposures can have large influences on individual disease risk; these influences are often incompletely captured in published cohort data.. Current frameworks for variant interpretation such as the ACMG criteria do not account for these modifiers[43]. The studies in this thesis have not sought to capture polygenic background or compound mutations within the same or between different disease genes. Future work may be deployed that explicitly models the disease risk in individual patients using a combination of their rare genetic variants, polygenic risk scores, demographics, and environmental exposures.

Future Directions

Scaling up Variant Interpretation. An ideal genotype-first approach to precision medicine will require a molecular understanding of all possible genetic variation in multiple functional contexts, a seemingly Herculean task with current methods. One approach to undertake these systematic studies has been the development of Deep Mutational Scans (DMS) or Multiplexed Assays of Variant Effects (MAVEs), which create libraries of all possible variants at each amino acid position (usually cDNA libraries or saturation editing of the endogenous locus with CRISPR)[44]. Typically,

variant pools are sequenced to quantify variant abundance, subjected to a fitness or other assay, and then again sequenced to measure the change in variant abundance[194,209]. Our lab is currently deploying this technology to study functional effects of variants in *SCN5A* using cDNA-based libraries subjected to a multi-drug selection pressure. This presents a higher-throughput complement to the SyncroPatch studies described in Chapter 2, and may provide additional data for empowering penetrance predictions in Chapter 3. A main limitation is the incomplete capture of channel parameters that could be ascertained using detailed APC studies, and the inability to study splice-altering variant effects as a result of the cDNA configuration. In the future, editing of iPSCs across the coding and splice-regions of *SCN5A* may allow for the joint interrogation of translated channel function and splice-altering variants in parallel. These same approaches are being deployed for other arrhythmia ion channel genes such as *KCNH2* and *KCNQ1*, or ion channels in other diseases. Indeed, methods developed in this thesis could be applied to improve variant interpretation for a host of disease-relevant ion channels or other proteins.

Predictive Value of Assays. The predictive value of the assays utilized and developed within this thesis can be further refined by calibrating them to clinical events. These may be readouts such as validation with controls as I have done for ParSE-seq in Chapter 6. In addition, well-phenotyped cohorts of genotype-positive individuals may be used to test whether covariates in functional studies correlate with the frequency of clinical phenotypes – premature ventricular contractions, ventricular fibrillation, Implanted Cardioverter Defibrillator placement, etc. While some studies have provided evidence for functional features to correlate with these outcomes, replication in larger cohorts and across multiple assays will significantly improve the predictive value of these assays[88]. Future studies will need to be calibrated to clinical data obtained from individuals of many different genetic ancestries, and ideally would invoke functional data collected from different laboratories across the world to ensure assay robustness.

Enabling Sharing of Functional Data. All too often, efforts in the laboratory fail to translate into clinically-actionable information. For example, many providers in genetics clinics rely on resources such as ClinVar to interpret variants discovered in the patients they treat. Many laboratories have not included the outcomes of their reclassification efforts, or their primary functional evidence, in this resource – these include studies in *KCNQ1*, *KCNH2*, and *SCN5A*. Since this primary literature is often unapproachable for those without deep knowledge of molecular biology and patch-clamp electrophysiology, sharing data in an open way on these sites represents not only good stewardship of data, but also the fulfillment of the promise to provide meaningful data. This motivated the development of the Variant Browser in Chapter 3 to integrate clinical observations, *in silico* predictions, functional data, and channel structural data into an easily searched repository for variant effect in *SCN5A*, *KCNQ1*, and *KCNH2*. Increasing collaborations between clinical laboratories and experts in development and deployment of functional assays may provide a particularly rich opportunity for translating mechanistic insight into physician decision-making[186].

Functional Data to Improve Therapeutic Development. Bringing a successful drug to market to benefit patients may represent the highest pinnacle of translational science. Gleaning a sufficient understanding of a biological mechanism to rationally intervene upon it pharmacologically, and then substantiate that effect and interaction in increasingly complex systems (preclinical to Phase II/III) requires tremendous insight into fundamental biology. The primary future goal of mine is to use functional investigations like those described in this thesis to discover chemical matter that can alleviate the negative functional outcomes of such variation. As described throughout this thesis, splice-altering variation represents an eminently tractable class of genetic variation for therapeutic targeting. By investigating diverse therapeutic modalities including small molecules, antisense oligonucleotides, gene therapies, increasingly sophisticated read-through inhibition, and perhaps direct gene editing, it is reasonable to anticipate that pharmacological and gene-

therapy approaches may resolve pathogenic variant effects. Indeed, ‘N of 1’ drug discovery is being increasingly recognized as a tractable approach to targeting rare disease, and regulatory and research pipelines are further contributing to that goal[160]. Many of the variants and model systems developed in Chapters 4 and 5 would provide optimal starting points for investigating splice-altering antisense oligonucleotides, while diverse chemical matter could be investigated at scale using the libraries designed with ParSE-seq in Chapter 6. It is my hope that the combined use of clinical medicine, human genetics, and organic chemistry/chemical biology may move beyond advances in genetic diagnosis to intervening upon the relevant disease mechanisms.

REFERENCES

1 Claussnitzer, M. *et al.* A brief history of human disease genetics. *Nature* **577**, 179-189 (2020). https://doi.org:10.1038/s41586-019-1879-7

2 Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-753 (2009). https://doi.org:10.1038/nature08494

3 Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A. & McKusick, V. A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* **33**, D514-517 (2005). https://doi.org:10.1093/nar/gki033

4 Motulsky, A. G. & King, M. C. The Great Adventure of an American Human Geneticist. *Annu Rev Genomics Hum Genet* **17**, 1-15 (2016). https://doi.org:10.1146/annurev-genom-083115-022528

5 Schrott, H. G., Goldstein, J. L., Hazzard, W. R., McGoodwin, M. M. & Motulsky, A. G. Familial hypercholesterolemia in a large indred. Evidence for a monogenic mechanism. *Ann Intern Med* **76**, 711-720 (1972). https://doi.org:10.7326/0003-4819-76-5-711

6 Neel, J. V. The Inheritance of Sickle Cell Anemia. *Science* **110**, 64-66 (1949). https://doi.org:10.1126/science.110.2846.64

7 Rommens, J. M. *et al.* Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* **245**, 1059-1065 (1989). https://doi.org:10.1126/science.2772657

8 Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184-191 (2016). https://doi.org:10.1038/nbt.3437

9 Moretti, A. *et al.* Patient-specific induced pluripotent stem-cell models for long-QT syndrome. *N Engl J Med* **363**, 1397-1409 (2010). https://doi.org:10.1056/NEJMoa0908679

10 Doudna, J. A. The promise and challenge of therapeutic genome editing. *Nature* **578**, 229-236 (2020). https://doi.org:10.1038/s41586-020-1978-5

11 Wainwright, C. E. *et al.* Lumacaftor-Ivacaftor in Patients with Cystic Fibrosis Homozygous for Phe508del CFTR. *N Engl J Med* **373**, 220-231 (2015). https://doi.org:10.1056/NEJMoa1409547

12 Dugger, S. A., Platt, A. & Goldstein, D. B. Drug development in the era of precision medicine. *Nat Rev Drug Discov* **17**, 183-196 (2018). https://doi.org:10.1038/nrd.2017.226

13 Koblan, L. W. *et al.* In vivo base editing rescues Hutchinson-Gilford progeria syndrome in mice. *Nature* **589**, 608-614 (2021). https://doi.org:10.1038/s41586-020-03086-7

14 Frangoul, H. *et al.* CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. *N Engl J Med* **384**, 252-260 (2021). https://doi.org:10.1056/NEJMoa2031054

15 Reichart, D. *et al.* Efficient in vivo genome editing prevents hypertrophic cardiomyopathy in mice. *Nat Med* **29**, 412-421 (2023). https://doi.org:10.1038/s41591-022-02190-7

16 Musunuru, K. *et al.* In vivo CRISPR base editing of PCSK9 durably lowers cholesterol in primates. *Nature* **593**, 429-434 (2021). https://doi.org:10.1038/s41586-021-03534-y

17 Visscher, P. M. & Goddard, M. E. From R.A. Fisher's 1918 Paper to GWAS a Century Later. *Genetics* **211**, 1125-1130 (2019). https://doi.org:10.1534/genetics.118.301594

18 Abdellaoui, A., Yengo, L., Verweij, K. J. H. & Visscher, P. M. 15 years of GWAS discovery: Realizing the promise. *Am J Hum Genet* **110**, 179-194 (2023). https://doi.org:10.1016/j.ajhg.2022.12.011

19 Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-678 (2007). https://doi.org:10.1038/nature05911

20 Torkamani, A., Wineinger, N. E. & Topol, E. J. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet* **19**, 581-590 (2018). https://doi.org:10.1038/s41576-018-0018-x

21 Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* **50**, 1219-1224 (2018). https://doi.org:10.1038/s41588-018-0183-z

22 Fahed, A. C. *et al.* Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. *Nat Commun* **11**, 3635 (2020). https://doi.org:10.1038/s41467-020-17374-3

23 Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications. *Nat Genet* **47**, 856-860 (2015). https://doi.org:10.1038/ng.3314

24 Cooper, D. N., Krawczak, M., Polychronakos, C., Tyler-Smith, C. & Kehrer-Sawatzki, H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum Genet* **132**, 1077-1130 (2013). https://doi.org:10.1007/s00439-013-1331-2

25 Green, E. D. *et al.* Strategic vision for improving human health at The Forefront of Genomics. *Nature* **586**, 683-692 (2020). https://doi.org:10.1038/s41586-020-2817-4

26 Schwartz, P. J. *et al.* Inherited cardiac arrhythmias. *Nat Rev Dis Primers* **6**, 58 (2020). https://doi.org:10.1038/s41572-020-0188-7

27 O'Hara, T., Virág, L., Varró, A. & Rudy, Y. Simulation of the undiseased human cardiac ventricular action potential: model formulation and experimental validation. *PLoS Comput Biol* **7**, e1002061 (2011). https://doi.org:10.1371/journal.pcbi.1002061

28 Brugada, J., Campuzano, O., Arbelo, E., Sarquella-Brugada, G. & Brugada, R. Present Status of Brugada Syndrome: JACC State-of-the-Art Review. *J Am Coll Cardiol* **72**, 1046-1059 (2018). https://doi.org:10.1016/j.jacc.2018.06.037

29 Kapplinger, J. D. *et al.* An international compendium of mutations in the SCN5A-encoded cardiac sodium channel in patients referred for Brugada syndrome genetic testing. *Heart Rhythm* **7**, 33-46 (2010). https://doi.org:10.1016/j.hrthm.2009.09.069

30 Hosseini, S. M. *et al.* Reappraisal of Reported Genes for Sudden Arrhythmic Death: Evidence-Based Evaluation of Gene Validity for Brugada Syndrome. *Circulation* **138**, 1195-1205 (2018). https://doi.org:10.1161/circulationaha.118.035070

31 Schwartz, P. J. 1970-2020: 50 years of research on the long QT syndrome-from almost zero knowledge to precision medicine. *Eur Heart J* **42**, 1063-1072 (2021). https://doi.org:10.1093/eurheartj/ehaa769

32 Roden, D. M. Drug-induced prolongation of the QT interval. *N Engl J Med* **350**, 1013-1022 (2004). https://doi.org:10.1056/NEJMra032426

33 Antzelevitch, C. & Burashnikov, A. Overview of Basic Mechanisms of Cardiac Arrhythmia. *Card Electrophysiol Clin* **3**, 23-45 (2011). https://doi.org:10.1016/j.ccep.2010.10.012

34 Schwartz, P. J. & Crotti, L. in *Circulation* Vol. 124 2181-2184 (2011).

35 Kapplinger, J. D. *et al.* Spectrum and prevalence of mutations from the first 2,500 consecutive unrelated patients referred for the FAMILION long QT syndrome genetic test. *Heart Rhythm* **6**, 1297-1303 (2009). https://doi.org:10.1016/j.hrthm.2009.05.021
36 Adler, A. *et al.* An International, Multicentered, Evidence-Based Reappraisal of Genes Reported to Cause Congenital Long QT Syndrome. *Circulation* **141**, 418-428 (2020). https://doi.org:10.1161/circulationaha.119.043132
37 Roden, D. M. Clinical practice. Long-QT syndrome. *N Engl J Med* **358**, 169-176 (2008). https://doi.org:10.1056/NEJMcp0706513
38 Lukas Laws, J. *et al.* Arrhythmias as Presentation of Genetic Cardiomyopathy. *Circ Res* **130**, 1698-1722 (2022). https://doi.org:10.1161/circresaha.122.319835
39 Gigli, M. *et al.* Phenotypic Expression, Natural History, and Risk Stratification of Cardiomyopathy Caused by Filamin C Truncating Variants. *Circulation* **144**, 1600-1611 (2021). https://doi.org:10.1161/circulationaha.121.053521
40 Gaita, F. *et al.* Short QT Syndrome: a familial cause of sudden death. *Circulation* **108**, 965-970 (2003). https://doi.org:10.1161/01.cir.0000085071.28695.c4
41 Leenhardt, A., Denjoy, I. & Guicheney, P. Catecholaminergic polymorphic ventricular tachycardia. *Circ Arrhythm Electrophysiol* **5**, 1044-1052 (2012). https://doi.org:10.1161/circep.111.962027
42 Sun, B. *et al.* Cardiac ryanodine receptor calcium release deficiency syndrome. *Sci Transl Med* **13** (2021). https://doi.org:10.1126/scitranslmed.aba7287
43 Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405-424 (2015). https://doi.org:10.1038/gim.2015.30
44 Starita, L. M. *et al.* Variant Interpretation: Functional Assays to the Rescue. *Am J Hum Genet* **101**, 315-325 (2017). https://doi.org:10.1016/j.ajhg.2017.07.014
45 Landrum, M. J. *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* **42**, D980-985 (2014). https://doi.org:10.1093/nar/gkt1113
46 Rehm, H. L. *et al.* ClinGen--the Clinical Genome Resource. *N Engl J Med* **372**, 2235-2242 (2015). https://doi.org:10.1056/NEJMsr1406261
47 Brnich, S. E. *et al.* Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med* **12**, 3 (2019). https://doi.org:10.1186/s13073-019-0690-2
48 Tavtigian, S. V., Harrison, S. M., Boucher, K. M. & Biesecker, L. G. Fitting a naturally scaled point system to the ACMG/AMP variant classification guidelines. *Hum Mutat* **41**, 1734-1737 (2020). https://doi.org:10.1002/humu.24088
49 Tavtigian, S. V. *et al.* Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet Med* **20**, 1054-1060 (2018). https://doi.org:10.1038/gim.2017.210
50 Parikh, V. N. Promise and Peril of Population Genomics for the Development of Genome-First Approaches in Mendelian Cardiovascular Disease. *Circ Genom Precis Med* **14**, e002964 (2021). https://doi.org:10.1161/circgen.120.002964

51 Glazer, A. M. *et al.* Arrhythmia Variant Associations and Reclassifications in the eMERGE-III Sequencing Study. *Circulation* (2021). https://doi.org:10.1161/circulationaha.121.055562
52 Hille, B. *Ion channels of excitable membranes*. 3rd edn, (Sinauer, 2001).
53 Obergrussberger, A., Friis, S., Brüggemann, A. & Fertig, N. Automated patch clamp in drug discovery: major breakthroughs and innovation in the last decade. *Expert Opin Drug Discov* **16**, 1-5 (2021). https://doi.org:10.1080/17460441.2020.1791079
54 Ng, C. A. *et al.* A massively parallel assay accurately discriminates between functionally normal and abnormal variants in a hotspot domain of KCNH2. *Am J Hum Genet* (2022). https://doi.org:10.1016/j.ajhg.2022.05.003
55 Glazer, A. M. *et al.* Deep Mutational Scan of an SCN5A Voltage Sensor. *Circ Genom Precis Med* **13**, e002786 (2020). https://doi.org:10.1161/circgen.119.002786
56 Kozek, K. A. *et al.* High-throughput discovery of trafficking-deficient variants in the cardiac potassium channel K(V)11.1. *Heart Rhythm* **17**, 2180-2189 (2020). https://doi.org:10.1016/j.hrthm.2020.05.041
57 The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-585 (2013). https://doi.org:10.1038/ng.2653
58 Chelly, J., Concordet, J. P., Kaplan, J. C. & Kahn, A. Illegitimate transcription: transcription of any gene in any cell type. *Proc Natl Acad Sci U S A* **86**, 2617-2621 (1989). https://doi.org:10.1073/pnas.86.8.2617
59 Kato, K. *et al.* LMNA Missense Mutation Causes Nonsense-Mediated mRNA Decay and Severe Dilated Cardiomyopathy. *Circ Genom Precis Med* **13**, 435-443 (2020). https://doi.org:10.1161/circgen.119.002853
60 Kremer, L. S. *et al.* Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat Commun* **8**, 15824 (2017). https://doi.org:10.1038/ncomms15824
61 Aicher, J. K., Jewell, P., Vaquero-Garcia, J., Barash, Y. & Bhoj, E. J. Mapping RNA splicing variations in clinically accessible and nonaccessible tissues to facilitate Mendelian disease diagnosis using RNA-seq. *Genet Med* **22**, 1181-1190 (2020). https://doi.org:10.1038/s41436-020-0780-y
62 Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281-2308 (2013). https://doi.org:10.1038/nprot.2013.143
63 Wuriyanghai, Y. *et al.* Complex aberrant splicing in the induced pluripotent stem cell-derived cardiomyocytes from a patient with long QT syndrome carrying KCNQ1-A344Aspl mutation. *Heart Rhythm* **15**, 1566-1574 (2018). https://doi.org:10.1016/j.hrthm.2018.05.028
64 Jaganathan, K. *et al.* Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **176**, 535-548.e524 (2019). https://doi.org:10.1016/j.cell.2018.12.015
65 Rowlands, C. *et al.* Comparison of in silico strategies to prioritize rare genomic variants impacting RNA splicing for the diagnosis of genomic disorders. *Sci Rep* **11**, 20607 (2021). https://doi.org:10.1038/s41598-021-99747-2
66 Mizusawa, Y. & Wilde, A. A. Brugada syndrome. *Circ Arrhythm Electrophysiol* **5**, 606-616 (2012). https://doi.org:10.1161/circep.111.964577

67 Gui, J. *et al.* Multiple loss-of-function mechanisms contribute to SCN5A-related familial sick sinus syndrome. *PLoS One* **5**, e10985 (2010). https://doi.org:10.1371/journal.pone.0010985
68 Bezzina, C. R. *et al.* Compound heterozygosity for mutations (W156X and R225W) in SCN5A associated with severe cardiac conduction disturbances and degenerative changes in the conduction system. *Circ Res* **92**, 159-168 (2003). https://doi.org:10.1161/01.res.0000052672.97759.36
69 Kroncke, B. M., Glazer, A. M., Smith, D. K., Blume, J. D. & Roden, D. M. SCN5A (Na(V)1.5) Variant Functional Perturbation and Clinical Presentation: Variants of a Certain Significance. *Circ Genom Precis Med* **11**, e002095 (2018). https://doi.org:10.1161/circgen.118.002095
70 Walsh, R. *et al.* Enhancing rare variant interpretation in inherited arrhythmias through quantitative analysis of consortium disease cohorts and population controls. *Genet Med* **23**, 47-58 (2021). https://doi.org:10.1038/s41436-020-00946-5
71 Pan, X. *et al.* Comparative structural analysis of human Na(v)1.1 and Na(v)1.5 reveals mutational hotspots for sodium channelopathies. *Proc Natl Acad Sci U S A* **118** (2021). https://doi.org:10.1073/pnas.2100066118
72 Clatot, J. *et al.* Voltage-gated sodium channels assemble and gate as dimers. *Nat Commun* **8**, 2077 (2017). https://doi.org:10.1038/s41467-017-02262-0
73 Clatot, J. *et al.* Mutant voltage-gated Na(+) channels can exert a dominant negative effect through coupled gating. *Am J Physiol Heart Circ Physiol* **315**, H1250-h1257 (2018). https://doi.org:10.1152/ajpheart.00721.2017
74 Keller, D. I. *et al.* Brugada syndrome and fever: genetic and molecular characterization of patients carrying SCN5A mutations. *Cardiovasc Res* **67**, 510-519 (2005). https://doi.org:10.1016/j.cardiores.2005.03.024
75 Doisne, N. *et al.* In vivo Dominant-Negative Effect of an SCN5A Brugada Syndrome Variant. *Front Physiol* **12**, 661413 (2021). https://doi.org:10.3389/fphys.2021.661413
76 Glazer, A. M. *et al.* High-Throughput Reclassification of SCN5A Variants. *Am J Hum Genet* **107**, 111-123 (2020). https://doi.org:10.1016/j.ajhg.2020.05.015
77 Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443 (2020). https://doi.org:10.1038/s41586-020-2308-7
78 Matreyek, K. A., Stephany, J. J., Chiasson, M. A., Hasle, N. & Fowler, D. M. An improved platform for functional assessment of large protein libraries in mammalian cells. *Nucleic Acids Res* **48**, e1 (2020). https://doi.org:10.1093/nar/gkz910
79 Matreyek, K. A., Stephany, J. J. & Fowler, D. M. A platform for functional assessment of large variant libraries in mammalian cells. *Nucleic Acids Res* **45**, e102 (2017). https://doi.org:10.1093/nar/gkx183
80 Kowarz, E., Löscher, D. & Marschalek, R. Optimized Sleeping Beauty transposons rapidly generate stable transgenic cell lines. *Biotechnol J* **10**, 647-653 (2015). https://doi.org:10.1002/biot.201400821
81 Mátés, L. *et al.* Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nat Genet* **41**, 753-761 (2009). https://doi.org:10.1038/ng.343

82 Whiffin, N. *et al.* Using high-resolution variant frequencies to empower clinical genome interpretation. *Genet Med* **19**, 1151-1158 (2017). https://doi.org:10.1038/gim.2017.26
83 Clatot, J. *et al.* Dominant-negative effect of SCN5A N-terminal mutations through the interaction of Na(v)1.5 α-subunits. *Cardiovasc Res* **96**, 53-63 (2012). https://doi.org:10.1093/cvr/cvs211
84 Poelzing, S. *et al.* SCN5A polymorphism restores trafficking of a Brugada syndrome mutation on a separate gene. *Circulation* **114**, 368-376 (2006). https://doi.org:10.1161/circulationaha.105.601294
85 Baroudi, G., Napolitano, C., Priori, S. G., Del Bufalo, A. & Chahine, M. Loss of function associated with novel mutations of the SCN5A gene in patients with Brugada syndrome. *Can J Cardiol* **20**, 425-430 (2004).
86 Veitia, R. A. & Birchler, J. A. Dominance and gene dosage balance in health and disease: why levels matter! *J Pathol* **220**, 174-185 (2010). https://doi.org:10.1002/path.2623
87 Wang, Z. *et al.* Calmodulin binds to the N-terminal domain of the cardiac sodium channel Na(v)1.5. *Channels (Austin)* **14**, 268-286 (2020). https://doi.org:10.1080/19336950.2020.1805999
88 Ishikawa, T. *et al.* Functionally validated SCN5A variants allow interpretation of pathogenicity and prediction of lethal events in Brugada syndrome. *Eur Heart J* **42**, 2854-2863 (2021). https://doi.org:10.1093/eurheartj/ehab254
89 Meregalli, P. G. *et al.* Type of SCN5A mutation determines clinical severity and degree of conduction slowing in loss-of-function sodium channelopathies. *Heart Rhythm* **6**, 341-348 (2009). https://doi.org:10.1016/j.hrthm.2008.11.009
90 Vanoye, C. G. *et al.* High-Throughput Functional Evaluation of KCNQ1 Decrypts Variants of Unknown Significance. *Circ Genom Precis Med* **11**, e002345 (2018). https://doi.org:10.1161/circgen.118.002345
91 Ng, C. A. *et al.* High-throughput phenotyping of heteromeric human ether-à-go-go-related gene potassium channel variants can discriminate pathogenic from rare benign variants. *Heart Rhythm* **17**, 492-500 (2020). https://doi.org:10.1016/j.hrthm.2019.09.020
92 Bezzina, C. R. *et al.* Common variants at SCN5A-SCN10A and HEY2 are associated with Brugada syndrome, a rare disease with high risk of sudden cardiac death. *Nat Genet* **45**, 1044-1049 (2013). https://doi.org:10.1038/ng.2712
93 Wright, C. F. *et al.* Assessing the Pathogenicity, Penetrance, and Expressivity of Putative Disease-Causing Variants in a Population Setting. *Am J Hum Genet* **104**, 275-286 (2019). https://doi.org:10.1016/j.ajhg.2018.12.015
94 McClellan, J. & King, M. C. Genetic heterogeneity in human disease. *Cell* **141**, 210-217 (2010). https://doi.org:10.1016/j.cell.2010.03.032
95 Priori, S. G., Napolitano, C. & Schwartz, P. J. Low penetrance in the long-QT syndrome: clinical impact. *Circulation* **99**, 529-533 (1999). https://doi.org:10.1161/01.cir.99.4.529
96 Lahrouchi, N. *et al.* Transethnic Genome-Wide Association Study Provides Insights in the Genetic Architecture and Heritability of Long QT Syndrome. *Circulation* **142**, 324-338 (2020). https://doi.org:10.1161/circulationaha.120.045956

97 Wijeyeratne, Y. D. *et al.* SCN5A Mutation Type and a Genetic Risk Score Associate Variably With Brugada Syndrome Phenotype in SCN5A Families. *Circ Genom Precis Med* **13**, e002911 (2020). https://doi.org:10.1161/circgen.120.002911

98 Nauffal, V. *et al.* Monogenic and Polygenic Contributions to QTc Prolongation in the Population. *medRxiv*, 2021.2006.2018.21258578 (2021). https://doi.org:10.1101/2021.06.18.21258578

99 van Rees, J. B. *et al.* Inappropriate implantable cardioverter-defibrillator shocks: incidence, predictors, and impact on mortality. *J Am Coll Cardiol* **57**, 556-562 (2011). https://doi.org:10.1016/j.jacc.2010.06.059

100 Schwartz, P. J. *et al.* Who are the long-QT syndrome patients who receive an implantable cardioverter-defibrillator and what happens to them?: data from the European Long-QT Syndrome Implantable Cardioverter-Defibrillator (LQTS ICD) Registry. *Circulation* **122**, 1272-1282 (2010). https://doi.org:10.1161/circulationaha.110.950147

101 Larrea-Sebal, A. *et al.* MLb-LDLr: A Machine Learning Model for Predicting the Pathogenicity of LDLr Missense Variants. *JACC Basic Transl Sci* **6**, 815-827 (2021). https://doi.org:10.1016/j.jacbts.2021.08.009

102 Draelos, R. L. *et al.* GENESIS: Gene-Specific Machine Learning Models for Variants of Uncertain Significance Found in Catecholaminergic Polymorphic Ventricular Tachycardia and Long QT Syndrome-Associated Genes. *Circ Arrhythm Electrophysiol*, 101161circep121010326 (2022). https://doi.org:10.1161/circep.121.010326

103 Kozek, K. *et al.* Estimating the Post-Test Probability of Long QT Syndrome Diagnosis for Rare KCNH2 Variants. *Circ Genom Precis Med* (2021). https://doi.org:10.1161/circgen.120.003289

104 Kroncke, B. M. *et al.* A Bayesian method to estimate variant-induced disease penetrance. *PLoS Genet* **16**, e1008862 (2020). https://doi.org:10.1371/journal.pgen.1008862

105 Fletcher Mercaldo, S. & Blume, J. D. Missing data and prediction: the pattern submodel. *Biostatistics* **21**, 236-252 (2020). https://doi.org:10.1093/biostatistics/kxy040

106 Kroncke, B. M. *et al.* Protein structure aids predicting functional perturbation of missense variants in SCN5A and KCNQ1. *Comput Struct Biotechnol J* **17**, 206-214 (2019). https://doi.org:10.1016/j.csbj.2019.01.008

107 Schwartz, P. J., Crotti, L. & Insolia, R. Long-QT syndrome: from genetics to management. *Circ Arrhythm Electrophysiol* **5**, 868-877 (2012). https://doi.org:10.1161/CIRCEP.111.962019

108 Priori, S. G. *et al.* HRS/EHRA/APHRS expert consensus statement on the diagnosis and management of patients with inherited primary arrhythmia syndromes: document endorsed by HRS, EHRA, and APHRS in May 2013 and by ACCF, AHA, PACES, and AEPC in June 2013. *Heart Rhythm* **10**, 1932-1963 (2013). https://doi.org:10.1016/j.hrthm.2013.05.014

109 Glazer, A. M. *et al.* Arrhythmia variant associations and reclassifications in the eMERGE-III sequencing study. *medRxiv*, 2021.2003.2030.21254549 (2021). https://doi.org:10.1101/2021.03.30.21254549

110 Sun, J. & MacKinnon, R. Cryo-EM Structure of a KCNQ1/CaM Complex Reveals Insights into Congenital Long QT Syndrome. *Cell* **169**, 1042-1050.e1049 (2017). https://doi.org:10.1016/j.cell.2017.05.019
111 Zhang, X. *et al.* Disease-specific variant pathogenicity prediction significantly improves variant interpretation in inherited cardiac conditions. *Genet Med* **23**, 69-79 (2021). https://doi.org:10.1038/s41436-020-00972-3
112 Adzhubei, I. A. *et al.* in *Nat Methods* Vol. 7 248-249 (2010).
113 Choi, Y. & Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**, 2745-2747 (2015). https://doi.org:10.1093/bioinformatics/btv195
114 Ioannidis, N. M. *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet* **99**, 877-885 (2016). https://doi.org:10.1016/j.ajhg.2016.08.016
115 Wilde, A. A. M. & Amin, A. S. Clinical Spectrum of SCN5A Mutations: Long QT Syndrome, Brugada Syndrome, and Cardiomyopathy. *JACC Clin Electrophysiol* **4**, 569-579 (2018). https://doi.org:10.1016/j.jacep.2018.03.006
116 Kapa, S. *et al.* Genetic testing for long-QT syndrome: distinguishing pathogenic mutations from benign variants. *Circulation* **120**, 1752-1760 (2009). https://doi.org:10.1161/CIRCULATIONAHA.109.863076
117 Schwartz, P. J. *et al.* Mutation location and IKs regulation in the arrhythmic risk of long QT syndrome type 1: the importance of the KCNQ1 S6 region. *Eur Heart J* **42**, 4743-4755 (2021). https://doi.org:10.1093/eurheartj/ehab582
118 Sun, J. & MacKinnon, R. Structural Basis of Human KCNQ1 Modulation and Gating. *Cell* **180**, 340-347.e349 (2020). https://doi.org:10.1016/j.cell.2019.12.003
119 Li, Z. *et al.* Structure of human Na(v)1.5 reveals the fast inactivation-related segments as a mutational hotspot for the long QT syndrome. *Proc Natl Acad Sci U S A* **118** (2021). https://doi.org:10.1073/pnas.2100069118
120 Napolitano, C. *et al.* Genetic testing in the long QT syndrome: development and validation of an efficient approach to genotyping in clinical practice. *Jama* **294**, 2975-2980 (2005). https://doi.org:10.1001/jama.294.23.2975
121 Barc, J. *et al.* Genome-wide association analyses identify new Brugada syndrome risk loci and highlight a new mechanism of sodium channel regulation in disease susceptibility. *Nat Genet* **54**, 232-239 (2022). https://doi.org:10.1038/s41588-021-01007-6
122 Yagi, N. *et al.* A challenge for mutation specific risk stratification in long QT syndrome type 1. *J Cardiol* **72**, 56-65 (2018). https://doi.org:10.1016/j.jjcc.2017.12.011
123 Lane, C. M. *et al.* Long QT syndrome type 5-Lite: Defining the clinical phenotype associated with the potentially proarrhythmic p.Asp85Asn-KCNE1 common genetic variant. *Heart Rhythm* **15**, 1223-1230 (2018). https://doi.org:10.1016/j.hrthm.2018.03.038
124 Giudicessi, J. R., Roden, D. M., Wilde, A. A. M. & Ackerman, M. J. Classification and Reporting of Potentially Proarrhythmic Common Genetic Variation in Long QT Syndrome Genetic Testing. *Circulation* **137**, 619-630 (2018). https://doi.org:10.1161/circulationaha.117.030142

125 Schwartz, P. J., Crotti, L. & George, A. L., Jr. Modifier genes for sudden cardiac death. *Eur Heart J* **39**, 3925-3931 (2018). https://doi.org:10.1093/eurheartj/ehy502
126 Wada, Y. *et al.* Common Ancestry-specific Ion Channel Variants Predispose to Drug-induced Arrhythmias. *Circulation* (2022). https://doi.org:10.1161/circulationaha.121.054883
127 Crotti, L. *et al.* Calmodulin mutations and life-threatening cardiac arrhythmias: insights from the International Calmodulinopathy Registry. *Eur Heart J* **40**, 2964-2975 (2019). https://doi.org:10.1093/eurheartj/ehz311
128 Kang, P. W. *et al.* Calmodulin acts as a state-dependent switch to control a cardiac potassium channel opening. *Sci Adv* **6** (2020). https://doi.org:10.1126/sciadv.abd6798
129 Clerx, M., Heijman, J., Collins, P. & Volders, P. G. A. Predicting changes to I(Na) from missense mutations in human SCN5A. *Sci Rep* **8**, 12797 (2018). https://doi.org:10.1038/s41598-018-30577-5
130 Heyne, H. O. *et al.* Predicting functional effects of missense variants in voltage-gated sodium and calcium channels. *Sci Transl Med* **12** (2020). https://doi.org:10.1126/scitranslmed.aay6848
131 Gnecchi, M., Sala, L. & Schwartz, P. J. Precision Medicine and cardiac channelopathies: when dreams meet reality. *Eur Heart J* **42**, 1661-1675 (2021). https://doi.org:10.1093/eurheartj/ehab007
132 Backman, J. D. *et al.* Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* **599**, 628-634 (2021). https://doi.org:10.1038/s41586-021-04103-z
133 Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290-299 (2021). https://doi.org:10.1038/s41586-021-03205-y
134 Weile, J. & Roth, F. P. Multiplexed assays of variant effects contribute to a growing genotype-phenotype atlas. *Hum Genet* **137**, 665-678 (2018). https://doi.org:10.1007/s00439-018-1916-x
135 Renaux, A. *et al.* ORVAL: a novel platform for the prediction and exploration of disease-causing oligogenic variant combinations. *Nucleic Acids Res* **47**, W93-w98 (2019). https://doi.org:10.1093/nar/gkz437
136 Sibley, C. R., Blazquez, L. & Ule, J. Lessons from non-canonical splicing. *Nat Rev Genet* **17**, 407-421 (2016). https://doi.org:10.1038/nrg.2016.46
137 Pertea, M., Lin, X. & Salzberg, S. L. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res* **29**, 1185-1190 (2001). https://doi.org:10.1093/nar/29.5.1185
138 Yeo, G. & Burge, C. B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* **11**, 377-394 (2004). https://doi.org:10.1089/1066527041410418
139 Hanses, U. *et al.* Intronic CRISPR Repair in a Preclinical Model of Noonan Syndrome-Associated Cardiomyopathy. *Circulation* **142**, 1059-1076 (2020). https://doi.org:10.1161/circulationaha.119.044794
140 Bastarache, L. *et al.* Phenotype risk scores identify patients with unrecognized Mendelian disease patterns. *Science* **359**, 1233-1239 (2018). https://doi.org:10.1126/science.aal4043

141 Laššuthová, P. *et al.* Clinical, in silico, and experimental evidence for pathogenicity of two novel splice site mutations in the SH3TC2 gene. *J Neurogenet* **26**, 413-420 (2012). https://doi.org:10.3109/01677063.2012.711398
142 Cho, S. Y. *et al.* Novel PPOX exonic mutation inducing aberrant splicing in a patient with homozygous variegate porphyria. *Clin Chim Acta* **512**, 117-120 (2021). https://doi.org:10.1016/j.cca.2020.10.033
143 Bardai, A. *et al.* Sudden cardiac arrest associated with use of a non-cardiac drug that reduces cardiac excitability: evidence from bench, bedside, and community. *Eur Heart J* **34**, 1506-1516 (2013). https://doi.org:10.1093/eurheartj/eht054
144 Burset, M., Seledtsov, I. A. & Solovyev, V. V. Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Res* **28**, 4364-4375 (2000). https://doi.org:10.1093/nar/28.21.4364
145 Frisso, G. *et al.* Functional Studies and In Silico Analyses to Evaluate Non-Coding Variants in Inherited Cardiomyopathies. *Int J Mol Sci* **17** (2016). https://doi.org:10.3390/ijms17111883
146 Patel, P. N. *et al.* Contribution of Noncanonical Splice Variants to TTN Truncating Variant Cardiomyopathy. *Circ Genom Precis Med* **14**, e003389 (2021). https://doi.org:10.1161/circgen.121.003389
147 Ito, K. *et al.* Identification of pathogenic gene mutations in LMNA and MYBPC3 that alter RNA splicing. *Proc Natl Acad Sci U S A* **114**, 7689-7694 (2017). https://doi.org:10.1073/pnas.1707741114
148 Vaz-Drago, R., Custódio, N. & Carmo-Fonseca, M. Deep intronic mutations and human disease. *Hum Genet* **136**, 1093-1111 (2017). https://doi.org:10.1007/s00439-017-1809-4
149 Gaildrat, P. *et al.* Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. *Methods Mol Biol* **653**, 249-257 (2010). https://doi.org:10.1007/978-1-60761-759-4_15
150 Gao, D. *et al.* A deep learning approach to identify gene targets of a therapeutic for human splicing disorders. *Nat Commun* **12**, 3332 (2021). https://doi.org:10.1038/s41467-021-23663-2
151 Gergics, P. *et al.* High-throughput splicing assays identify missense and silent splice-disruptive POU1F1 variants underlying pituitary hormone deficiency. *Am J Hum Genet* (2021). https://doi.org:10.1016/j.ajhg.2021.06.013
152 Cheung, R. *et al.* A Multiplexed Assay for Exon Recognition Reveals that an Unappreciated Fraction of Rare Genetic Variants Cause Large-Effect Splicing Disruptions. *Mol Cell* **73**, 183-194.e188 (2019). https://doi.org:10.1016/j.molcel.2018.10.037
153 Holliday, M. *et al.* Transcriptome Sequencing of Patients With Hypertrophic Cardiomyopathy Reveals Novel Splice-Altering Variants in MYBPC3. *Circ Genom Precis Med* **14**, e003202 (2021). https://doi.org:10.1161/circgen.120.003202
154 Lopes, L. R. *et al.* Cryptic Splice-Altering Variants in MYBPC3 Are a Prevalent Cause of Hypertrophic Cardiomyopathy. *Circ Genom Precis Med* **13**, e002905 (2020). https://doi.org:10.1161/circgen.120.002905
155 Singer, E. S., Ingles, J., Semsarian, C. & Bagnall, R. D. Key Value of RNA Analysis of MYBPC3 Splice-Site Variants in Hypertrophic Cardiomyopathy. *Circ Genom Precis Med* **12**, e002368 (2019). https://doi.org:10.1161/circgen.118.002368

156 Crotti, L. *et al.* A KCNH2 branch point mutation causing aberrant splicing contributes to an explanation of genotype-negative long QT syndrome. *Heart Rhythm* **6**, 212-218 (2009). https://doi.org:10.1016/j.hrthm.2008.10.044
157 Tobert, K. E. *et al.* Genome Sequencing in a Genetically Elusive Multi-Generational Long QT Syndrome Pedigree Identifies a Novel LQT2-Causative Deeply Intronic KCNH2 Variant. *Heart Rhythm* (2022). https://doi.org:10.1016/j.hrthm.2022.02.004
158 Lord, J. *et al.* Pathogenicity and selective constraint on variation near splice sites. *Genome Res* **29**, 159-170 (2019). https://doi.org:10.1101/gr.238444.118
159 Crooke, S. T., Baker, B. F., Crooke, R. M. & Liang, X. H. Antisense technology: an overview and prospectus. *Nat Rev Drug Discov* **20**, 427-453 (2021). https://doi.org:10.1038/s41573-021-00162-z
160 Kim, J. *et al.* Patient-Customized Oligonucleotide Therapy for a Rare Genetic Disease. *N Engl J Med* **381**, 1644-1652 (2019). https://doi.org:10.1056/NEJMoa1813279
161 Brierley, J., Aylett, S. & Archard, D. Framework for "N-of-1" Experimental Therapies. *N Engl J Med* **382**, e7 (2020). https://doi.org:10.1056/NEJMc1915778
162 Dotzler, S. M. *et al.* Suppression-Replacement KCNQ1 Gene Therapy for Type 1 Long QT Syndrome. *Circulation* (2021). https://doi.org:10.1161/circulationaha.120.051836
163 Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat Biotechnol* **38**, 824-844 (2020). https://doi.org:10.1038/s41587-020-0561-9
164 Towbin, J. A. *et al.* 2019 HRS expert consensus statement on evaluation, risk stratification, and management of arrhythmogenic cardiomyopathy. *Heart Rhythm* **16**, e301-e372 (2019). https://doi.org:10.1016/j.hrthm.2019.05.007
165 Ortiz-Genga, M. F. *et al.* Truncating FLNC Mutations Are Associated With High-Risk Dilated and Arrhythmogenic Cardiomyopathies. *J Am Coll Cardiol* **68**, 2440-2451 (2016). https://doi.org:10.1016/j.jacc.2016.09.927
166 Holliday, M. *et al.* Transcriptome Sequencing of Patients with Hypertrophic Cardiomyopathy Reveals Novel Splice-altering Variants in MYBPC3. *Circ Genom Precis Med* (2021). https://doi.org:10.1161/circgen.120.003202
167 Begay, R. L. *et al.* FLNC Gene Splice Mutations Cause Dilated Cardiomyopathy. *JACC Basic Transl Sci* **1**, 344-359 (2016). https://doi.org:10.1016/j.jacbts.2016.05.004
168 Miller, D. T. *et al.* ACMG SF v3.0 list for reporting of secondary findings in clinical exome and genome sequencing: a policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet Med* **23**, 1381-1390 (2021). https://doi.org:10.1038/s41436-021-01172-3
169 Bersell, K. R. *et al.* Transcriptional Dysregulation Underlies Both Monogenic Arrhythmia Syndrome and Common Modifiers of Cardiac Repolarization. *Circulation* (2022). https://doi.org:10.1161/circulationaha.122.062193
170 Dowey, S. N., Huang, X., Chou, B. K., Ye, Z. & Cheng, L. Generation of integration-free human induced pluripotent stem cells from postnatal blood mononuclear cells by plasmid vector expression. *Nat Protoc* **7**, 2013-2021 (2012). https://doi.org:10.1038/nprot.2012.121
171 Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890 (2018). https://doi.org:10.1093/bioinformatics/bty560

172 Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907-915 (2019). https://doi.org:10.1038/s41587-019-0201-4
173 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009). https://doi.org:10.1093/bioinformatics/btp352
174 Robinson, J. T. *et al.* in *Nat Biotechnol* Vol. 29 24-26 (2011).
175 Lykke-Andersen, S. & Jensen, T. H. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nat Rev Mol Cell Biol* **16**, 665-677 (2015). https://doi.org:10.1038/nrm4063
176 Chen, S. N. *et al.* Activation of PDGFRA signaling contributes to filamin C-related arrhythmogenic cardiomyopathy. *Sci Adv* **8**, eabk0052 (2022). https://doi.org:10.1126/sciadv.abk0052
177 Jordan, E. *et al.* Evidence-Based Assessment of Genes in Dilated Cardiomyopathy. *Circulation* **144**, 7-19 (2021). https://doi.org:10.1161/circulationaha.120.053033
178 Brun, F. *et al.* FLNC truncations cause arrhythmogenic right ventricular cardiomyopathy. *J Med Genet* **57**, 254-257 (2020). https://doi.org:10.1136/jmedgenet-2019-106394
179 Brodehl, A. *et al.* Mutations in FLNC are Associated with Familial Restrictive Cardiomyopathy. *Hum Mutat* **37**, 269-279 (2016). https://doi.org:10.1002/humu.22942
180 Tucker, N. R. *et al.* Novel Mutation in FLNC (Filamin C) Causes Familial Restrictive Cardiomyopathy. *Circ Cardiovasc Genet* **10** (2017). https://doi.org:10.1161/circgenetics.117.001780
181 Begay, R. L. *et al.* Filamin C Truncation Mutations Are Associated With Arrhythmogenic Dilated Cardiomyopathy and Changes in the Cell-Cell Adhesion Structures. *JACC Clin Electrophysiol* **4**, 504-514 (2018). https://doi.org:10.1016/j.jacep.2017.12.003
182 Carruth, E. D. *et al.* Loss-of-Function FLNC Variants Are Associated With Arrhythmogenic Cardiomyopathy Phenotypes When Identified Through Exome Sequencing of a General Clinical Population. *Circ Genom Precis Med*, 101161circgen121003645 (2022). https://doi.org:10.1161/circgen.121.003645
183 Kurosaki, T., Popp, M. W. & Maquat, L. E. Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat Rev Mol Cell Biol* **20**, 406-420 (2019). https://doi.org:10.1038/s41580-019-0126-2
184 Thompson, D., Easton, D. F. & Goldgar, D. E. A full-likelihood method for the evaluation of causality of sequence variants from family data. *Am J Hum Genet* **73**, 652-655 (2003). https://doi.org:10.1086/378100
185 Corrado, D. *et al.* Diagnosis of arrhythmogenic cardiomyopathy: The Padua criteria. *Int J Cardiol* **319**, 106-114 (2020). https://doi.org:10.1016/j.ijcard.2020.06.005
186 Truty, R. *et al.* Spectrum of splicing variants in disease genes and the ability of RNA analysis to reduce uncertainty in clinical interpretation. *Am J Hum Genet* **108**, 696-708 (2021). https://doi.org:10.1016/j.ajhg.2021.03.006
187 Gergics, P. *et al.* High-throughput splicing assays identify missense and silent splice-disruptive POU1F1 variants underlying pituitary hormone deficiency. *medRxiv*, 2021.2002.2004.21249469 (2021). https://doi.org:10.1101/2021.02.04.21249469

188 Cortés-López, M. *et al.* High-throughput mutagenesis identifies mutations and RNA-binding proteins controlling CD19 splicing and CART-19 therapy resistance. *Nat Commun* **13**, 5570 (2022). https://doi.org:10.1038/s41467-022-31818-y

189 Adamson, S. I., Zhan, L. & Graveley, B. R. Vex-seq: high-throughput identification of the impact of genetic variation on pre-mRNA splicing efficiency. *Genome Biol* **19**, 71 (2018). https://doi.org:10.1186/s13059-018-1437-x

190 Peters, S. *et al.* Arrhythmic Phenotypes Are a Defining Feature of Dilated Cardiomyopathy-Associated SCN5A Variants: A Systematic Review. *Circ Genom Precis Med* **15**, e003432 (2022). https://doi.org:10.1161/circgen.121.003432

191 O'Neill, M. J. *et al.* Functional Assays Reclassify Suspected Splice-Altering Variants of Uncertain Significance in Mendelian Channelopathies. *Circ Genom Precis Med*, 101161circgen122003782 (2022). https://doi.org:10.1161/circgen.122.003782

192 Hong, K. *et al.* Cryptic 5' splice site activation in SCN5A associated with Brugada syndrome. *J Mol Cell Cardiol* **38**, 555-560 (2005). https://doi.org:10.1016/j.yjmcc.2004.10.015

193 Jain, P. C. & Varadarajan, R. A rapid, efficient, and economical inverse polymerase chain reaction-based method for generating a site saturation mutant library. *Anal Biochem* **449**, 90-98 (2014). https://doi.org:10.1016/j.ab.2013.12.002

194 Fowler, D. M., Stephany, J. J. & Fields, S. Measuring the activity of protein variants on a large scale using deep mutational scanning. *Nat Protoc* **9**, 2267-2284 (2014). https://doi.org:10.1038/nprot.2014.153

195 Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133-138 (2009). https://doi.org:10.1126/science.1162986

196 Potapov, V. & Ong, J. L. Examining Sources of Error in PCR by Single-Molecule Sequencing. *PLoS One* **12**, e0169774 (2017). https://doi.org:10.1371/journal.pone.0169774

197 Burridge, P. W. *et al.* Chemically defined generation of human cardiomyocytes. *Nat Methods* **11**, 855-860 (2014). https://doi.org:10.1038/nmeth.2999

198 Bodbin, S. E., Denning, C. & Mosqueira, D. Transfection of hPSC-Cardiomyocytes Using Viafect™ Transfection Reagent. *Methods Protoc* **3** (2020). https://doi.org:10.3390/mps3030057

199 Fayer, S. *et al.* Closing the gap: Systematic integration of multiplexed functional data resolves variants of uncertain significance in BRCA1, TP53, and PTEN. *Am J Hum Genet* **108**, 2248-2258 (2021). https://doi.org:10.1016/j.ajhg.2021.11.001

200 Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77 (2011). https://doi.org:10.1186/1471-2105-12-77

201 Kleinberger, J., Maloney, K. A., Pollin, T. I. & Jeng, L. J. in *Genet Med* Vol. 18 1165 (2016).

202 O'Neill, M. J. *et al.* Dominant negative effects of SCN5A missense variants. *Genet Med* **24**, 1238-1248 (2022). https://doi.org:10.1016/j.gim.2022.02.010

203 Concordet, J. P. & Haeussler, M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res* **46**, W242-w245 (2018). https://doi.org:10.1093/nar/gky354

204 Sayed, N., Liu, C. & Wu, J. C. Translation of Human-Induced Pluripotent Stem Cells: From Clinical Trial in a Dish to Precision Medicine. *J Am Coll Cardiol* **67**, 2161-2176 (2016). https://doi.org:10.1016/j.jacc.2016.01.083

205 Bournazos, A. M. *et al.* Standardized practices for RNA diagnostics using clinically accessible specimens reclassifies 75% of putative splicing variants. *Genet Med* **24**, 130-145 (2022). https://doi.org:10.1016/j.gim.2021.09.001

206 Zeng, T. & Li, Y. I. Predicting RNA splicing from DNA sequence using Pangolin. *Genome Biol* **23**, 103 (2022). https://doi.org:10.1186/s13059-022-02664-4

207 Soemedi, R. *et al.* Pathogenic variants that alter protein code often disrupt splicing. *Nat Genet* **49**, 848-855 (2017). https://doi.org:10.1038/ng.3837

208 Chiang, H. L. *et al.* Mechanism and modeling of human disease-associated near-exon intronic variants that perturb RNA splicing. *Nat Struct Mol Biol* **29**, 1043-1055 (2022). https://doi.org:10.1038/s41594-022-00844-1

209 Fowler, D. M. & Fields, S. Deep mutational scanning: a new style of protein science. *Nat Methods* **11**, 801-807 (2014). https://doi.org:10.1038/nmeth.3027

210 Jia, X. *et al.* Massively parallel functional testing of MSH2 missense variants conferring Lynch syndrome risk. *Am J Hum Genet* **108**, 163-175 (2021). https://doi.org:10.1016/j.ajhg.2020.12.003

211 Scott, A. *et al.* Saturation-scale functional evidence supports clinical variant interpretation in Lynch syndrome. *Genome Biol* **23**, 266 (2022). https://doi.org:10.1186/s13059-022-02839-z