

Titin: an analysis of genetic variation and cardiac phenotype

Submitted to Imperial College London
for the degree of Doctor of Philosophy

by

Dr Angharad Roberts

Cardiovascular Genetics and Genomics
NIHR Royal Brompton Cardiovascular Biomedical Research Unit
Imperial College Faculty of Medicine

September 2014

Abstract

Non ischaemic dilated cardiomyopathy (DCM) is an important cause of heart failure leading to chronic morbidity and death and as such is a major health burden. DCM is familial in up to 50% of cases but is genetically heterogeneous, hindering both genotype-phenotype studies and the application of genetic information for stratified patient management.

TTN truncating variants (TTNtv) cause severe and familial DCM, but sometimes occur in healthy individuals, posing challenges for the interpretation of these variants. In this PhD thesis, the power of quantitative cardiac MRI (CMR) is integrated with targeted resequencing of TTN in order to assess the relationship between TTN genotype, cardiac phenotype and clinical outcomes.

A prospectively recruited DCM cohort was established following CMR assessment in 374 patients (88% Caucasian, 72% male, mean age 54 ± 35 years, mean left ventricular ejection fraction (LVEF) $38\% \pm 24.5\%$, mean indexed end-diastolic volume 129 ± 141 mm). Following several iterations of design, refinement and testing an NGS assay was produced that captures all TTN coding exons and splice sites. Optimal parameters for sequencing and analysis of variants were established and genotype data compiled from 374 prospective, unselected cases, 155 end-stage retrospective cases, and 308 MRI phenotyped healthy volunteers. These data were integrated with that from 3603 population subjects and together used to identify molecular signatures that aid interpretation of TTN truncations both in DCM and as incidental findings.

Overall, TTNtv were identified in 1.4% of controls, 13% of unselected and 22% of end-stage DCM cases (OR = 13, $P = 2.8 \times 10^{-43}$, DCM vs controls) confirming TTN as the commonest cause of genetic DCM in all patient groups. TTNtv-containing exons in DCM have higher usage than those in controls ($P = 2.5 \times 10^{-4}$) and these are estimated to have >93% probability of pathogenicity (likelihood ratio 14). Compared to TTNtv-ve DCM, TTNtv+ve patients had lower LVEF ($P = 0.02$), thinner LV walls ($P = 0.02$), and a higher incidence of sustained ventricular tachycardia ($P = 0.001$). C-terminus TTNtv were also associated with lower LVEF versus N-terminus ($\beta = -18 \pm 7\%$, $p = 0.006$) and were more common in end-stage disease.

Together these data provide the first insight into genotype-phenotype correlations and will be of benefit in variant interpretation and patient stratification in TTN-based DCM.

Declaration of Originality and Copyright

The work contained in this thesis is my own. Any contribution of others to work described herein is duly acknowledged, and all else appropriately referenced.

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non Commercial No Derivatives licence.

Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work

Contribution of others to work presented in this thesis

Stuart Cook and Sanjay Prasad assigned DCM patients to the Biobank database. Basic CMR measurements (biventricular volumes, LV mass, EF and LGE assessment) were in most cases performed by the reporting clinician. John Baksi performed additional ventricular volume measurements and LGE assessment. Ventricular wall thickness and LAV measurements were performed by AMR.

BRU research nurses administered the patient questionnaire. Clinical data were collected from patient notes, other hospital records, and patient questionnaire by AMR. Data entry was performed by AMR. The DCM cohort in the database was curated by AMR.

AMR and Leica Briskey extracted DNA. AMR prepared the sequencing libraries for the first 2 generations of assays. Rachel Buchan prepared subsequent sequencing libraries. RB assisted with AMR's first SOLiD sequencing run, and carried out subsequent SOLiD sequencing.

RNA was extracted, and RNA-seq libraries were prepared by Leanne Felkin, the sequence data was aligned and annotated by Sebastian Schaefer.

The TTN LRG was developed in conjunction with James Ware, Roddy Walsh and the EBI Hinxton.

Shibu John and Sungsam Gong established the sequence alignment, variant calling, and basic variant annotation pipelines used in the analysis described here.

Splice variant prediction in the UK cohorts was performed by AMR, and in the population cohorts by Daniel Herman. AMR and DH developed the threshold for splice variant prediction.

Samuel Wilkinson performed Sanger validation

In chapter 6, plots of exon usage in the heart, distribution of TTNtv, distribution of sites susceptible to truncation, usage of TTN exons across isoforms and cohorts, and correlation of truncated transcript length with indices of DCM were generated by JW in discussion with AMR. Linear modelling was performed by JW in discussion with AMR and DH.

Acknowledgements

I am indebted to Professor Stuart Cook, as my supervisor, for his support and guidance, and for the opportunity to pursue this project to its completion. I should also like to thank Dr Paul Barton for his support and encouragement. His door has always been open, calorific snacks in plentiful supply, and his mentorship invaluable.

A large number of people have contributed to the development of this work through discussion and practical help. I am grateful to all members of the Genetics and Genomics team, for all they have done to help me along the way. I am also grateful to members of the Royal Brompton and Harefield NHS Foundation Trust CMR unit, the Cardiovascular and Metabolic Sciences team at the Max Delbrück Center for Molecular Medicine, and Department of Genetics at Harvard Medical School. Particular thanks are due to Dr James Ware, for his clarity of thought and patient guidance through the wilderness of statistics and R language. I also wish to thank Rachel Buchan, for whipping me into shape in the lab, that I might earn her highest praise: she forgets I am a medic and considers me a scientist.

This research was principally supported by a Fellowship from the Wellcome Trust, with additional support from the Medical Research Council, British Heart Foundation, the NIHR Royal Brompton Cardiovascular Biomedical Research Unit and the Fondation Leducq.

Finally, I am indebted to my parents, for their generous and unconditional support.

Selected publications arising from this work

Papers

Eric Villard, Claire Perret, Françoise Gary, Carole Proust, Gilles Dilanian, Christian Hengstenberg, Volker Ruppert, Eloisa Arbustini, Thomas Wichter, Marine Germain, Olivier Dubourg, Luigi Tavazzi, Marie-Claire Aumont, Pascal DeGroote, Laurent Fauchier, Jean-Noël Trochu, Pierre Gibelin, Jean-François Aupetit, Klaus Stark, Jeanette Erdmann, Roland Hetzer, Angharad M. Roberts, Paul J.R. Barton, Vera Regitz-Zagrosek, Cardiogenics Consortium, Uzma Aslam, Laëtitia Duboscq-Bidot, Matthias Meyborg, Bernhard Maisch, Hugo Madeira, Anders Waldenström, Enrique Galve, John G. Cleland, Richard Dorent, Gerard Roizes G, Tanja Zeller, Stefan Blankenberg, Alison H. Goodall, Stuart Cook, David A. Tregouet, Laurence Tiret, Richard Isnard, Michel Komajda, Philippe Charron, François Cambien
A genome-wide association study identifies new loci associated with heart failure due to dilated cardiomyopathy
European Heart Journal, Volume 32, Issue 9, Pages 1065-1076

Ware JS, Roberts AM, Cook SA.

Next generation sequencing for clinical diagnostics and personalised medicine: implications for the next generation cardiologist. Invited review for Heart. 2012;98:276-281.

Also reprinted in Postgraduate Medical Journal. 2012;88:234-239.

Ware JS, Roberts AM, John S, Gong S, Buchan RJ, Robinson DO, Lucassen AM, Peters NS, Behr ER, Cook SA.

Next generation diagnostics for Long QT syndromes. Journal of Cardiovascular Translational Research. 2012 DOI: 10.1007/s12265-012-9401-8 (Online First)

James S. Ware, Shibu John, Angharad M. Roberts, Rachel Buchan, Sungsam Gong, Nicholas S. Peters, David O. Robinson, Anneke Lucassen, Elijah R. Behr, Stuart A. Cook

Next Generation Diagnostics in Inherited Arrhythmia Syndromes: A Comparison of Two Approaches

Journal of Cardiovascular Translational Research, 2013 Feb;6(1):94-103

Angharad M. Roberts, James S. Ware, Daniel S. Herman, Sebastian Schafer, John Baksi, Alexander G. Bick, Rachel J. Buchan, Roddy Walsh, Shibu John, Samuel Wilkinson, Francesco Mazzarotto, Leanne E. Felkin, Sungsam Gong, Jacqueline A.L. MacArthur, Fiona Cunningham, Jason Flannick, Stacey B. Gabriel, David M. Altshuler, Peter S. Macdonald, Matthias Heinig, Anne M. Keogh, Christopher S. Hayward, Nicholas R. Banner, Dudley J. Pennell, Declan O'Regan, Tan Ru San, Antonio de Marvao, Timothy J. W. Dawes, Ankur Gulati, Emma J. Birks, Magdi H. Yacoub, Michael Radke, Michael Gotthardt, James G. Wilson, Christopher J. O'Donnell, Sanjay K. Prasad, Paul J.R. Barton, Diane Fatkin, Norbert Hubner, J. G. Seidman, Christine E. Seidman and Stuart A. Cook

Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease

Science Translational Medicine, 14 January 2015: Vol. 7, Issue 270, p. 270ra6

Abstracts and oral presentations

A. Roberts, J. Ware, D. Herman, S. Schafer, J. Baksi, R. Buchan, R. Walsh, S. John, S. Wilkinson, L. Felkin, A. Bick, F. Mazzarotto, M. Radke, M. Gotthardt, P. Barton, N. Hubner, J. Seidman, C. Seidman, S. Cook. Titin truncations: dissection of genotype and cardiac phenotype. American Society of Human Genetics annual meeting 2015 (platform presentation)

Angharad M. Roberts, James S. Ware, Daniel S. Herman, Sebastian Shafer, John Baksi, Rachel J. Buchan, Roddy Walsh, Shibu John, Samuel Wilkinson, Leanne E. Felkin, Antonio de Marvao, Timothy J. W. Dawes, Michael Radke, Michael Gotthardt, Sanjay K. Prasad, Paul J.R. Barton, Norbert Hubner, J. G. Seidman, Christine E. Seidman, Stuart A. Cook. The Phenotypic Spectrum of Titin Truncations in Health and Disease. The Fourth Cardiff Symposium on Clinical Cardiovascular Genetics, 2013 (oral presentation)

Angharad M. Roberts, James S. Ware, Daniel S. Herman, Sebastian Shafer, John Baksi, Rachel J. Buchan, Roddy Walsh, Shibu John, Samuel Wilkinson, Leanne E. Felkin, Antonio de Marvao, Timothy J. W. Dawes, Michael Radke, Michael Gotthardt, Sanjay K. Prasad, Paul J.R. Barton, Norbert Hubner, J. G. Seidman, Christine E. Seidman, Stuart A. Cook. Phenotype and molecular mechanism of Titin cardiomyopathy. British Society of Genetic Medicine, 2013 (oral presentation)

Angharad M. Roberts, James S. Ware, Daniel S. Herman, Sebastian Shafer, John Baksi, Rachel J. Buchan, Roddy Walsh, Shibu John, Samuel Wilkinson, Leanne E. Felkin, Antonio de Marvao, Timothy J. W. Dawes, Michael Radke, Michael Gotthardt, Sanjay K. Prasad, Paul J.R. Barton, Norbert Hubner, J. G. Seidman, Christine E. Seidman, Stuart A. Cook. The phenotypic spectrum of Titin truncations in health and disease. EMBL Conference Myofibrillar Z-disk Structure and Dynamics, 2013 (oral presentation)

Angharad M Roberts, James S Ware, Rachel J Buchan, Shibu John, Roddy Walsh, John Baksi, Samuel Wilkinson, Leanne Felkin, Paul J Barton, Stuart A Cook. Titin: genetic variation and cardiac phenotype. Agilent user meeting 2013 Invited presentation

Angharad M Roberts, James S Ware, Rachel J Buchan, Shibu John, Roddy Walsh, John Baksi, Samuel Wilkinson, Leanne Felkin, Paul J Barton, Stuart A Cook. Titin mutations in dilated cardiomyopathy. Personal Genomes & Medical Genomics 2012, Cold Spring Harbour Labs (Poster presentation and published abstract)

James Ware*, Angharad Roberts*, Shibu John, Stuart Cook. Comparison of NGS technologies for molecular diagnosis of Inherited Cardiac Conditions. Genomic Disorders 2011 - The Genomics of Rare Diseases (Poster presentation and published abstract) *Joint 1st author

Ware JS, John S, Roberts AM, Buchan RJ, Aitman TJ, Peters NS, Robinson DO, Lucassen AM, Behr ER, Cook SA. Molecular diagnosis of inherited cardiac conditions using next generation sequencing approaches. St Georges Cardiovascular Research Group Symposium 2010, Cardiac Arrhythmias and Sudden Death: From Bench to Bedside (poster presentation)

S John*, A Roberts*, J Ware, S Cook. SureSelect in-solution hybridization target enrichment for whole exome sequencing: coverage and efficacy. British Human Genetics Conference 2010 (Poster presentation and published abstract)
*joint 1st author

Glossary of terms and abbreviations

AHA	American Heart Association
arrayCGH	Array-comparative genomic hybridization – a technique to detect DNA copy number variation
ARVC	Arrhythmogenic right ventricular cardiomyopathy – an inherited cardiomyopathy
ATP	Adenosine triphosphate, a nucleoside triphosphate used in cells as a coenzyme.
BLAST	Basic Local Alignment Search Tool, finds regions of local similarity between sequences.
bp	Base pairs – the length of a (double-stranded) DNA sequence is reported in base pairs. Standard prefixes are used to indicate larger units, such as kbp and Mbp
BRU	Biomedical research unit
BSA	Body surface area
CAD	Coronary artery disease
CAPN3	Gene encoding Calpain-3
cDNA	Complementary DNA is DNA synthesized from a messenger RNA (mRNA) template
CM	Core-myopathies, the most common form of inherited non-dystrophic childhood muscle disorder, defined by areas of mitochondria depletion and sarcomere disorganisation in muscle fibres
CMD1G1	An autosomal dominant DCM locus, located on 2q13
CMR	Cardiovascular magnetic resonance imaging. A highly accurate and reproducible a medical imaging technology for the non-invasive assessment of the function and structure of the cardiovascular system
CNM	Centronuclear myopathies, a diverse group of clinically and genetically heterogeneous muscle disorders defined by the presence of internalised nuclei in myofibers without excessive regeneration or structural abnormalities
Consanguineous	Refers to a union between two related individuals
DCM	Dilated cardiomyopathy – cardiomyopathy characterised by increased ventricular cavity size without increased wall thickness, often inherited
DNA	Deoxyribonucleic acid
Dominant inheritance	A pattern of inheritance in which one aberrant copy of a gene is sufficient for the disease or trait to develop.
EBI	European Bioinformatics Institute (EMBL-EBI) is a centre for research and services in bioinformatics
ECG	Electrocardiogram – an ECG is a recording of the electrical

	activity of the heart taken from the body surface
EDMD	Emery-Dreifuss muscular dystrophy, a disease characterized by muscular weakness and atrophy, with early joint contractures and cardiomyopathy.
EDV	End diastolic volume is the volume of blood in a ventricle at end load or filling in diastole
ELM	The Eukaryotic Linear Motif Resource, a computational biology resource for investigating candidate functional sites in eukaryotic proteins
EOMFC	Early-onset myopathy with fatal cardiomyopathy, a sub-type of multi minicore disease
ePCR	Emulsion PCR, isolates individual DNA molecules along with primer-coated beads in aqueous droplets within an oil phase
ESC	European cardiology society
ESP	Exome sequencing project
EST	Expressed sequence tag
ESV	End-systolic volume is the volume of blood in a ventricle at the end of contraction, or systole
Expressivity	Variations in a phenotype among individuals carrying a particular genotype. It is determined by the proportion of individuals with a given genotype who also possess the associated phenotype
FDA	US Food and Drug Administration
FHS	Framingham heart study
FINmaj	A founder mutation within the Finnish population causing tibial muscular dystrophy in heterozygotes
FN-III	Fibrinectin-III domain, a repeated domain within the TTN protein
GATK	Genome Analysis Toolkit – open-source software for the analysis of next-generation sequencing data
GC	GC-content is the percentage of nitrogenous bases on a DNA molecule that are either guanine or cytosine
Genotyping	The process of identifying genetic variation in an individual genome
GERP	GERP score: A score used to calculate the conservation of each nucleotide in multi-species alignment
GWAS	Genome-wide association study
Haploinsufficiency	Occurs when a diploid organism has only a single functional copy of a gene (with the other copy inactivated by mutation) and the single functional copy does not produce enough of a gene product (typically a protein) to bring about a wild-type condition, leading to an abnormal or diseased state

HCM	Hypertrophic cardiomyopathy – cardiomyopathy characterised by increased ventricular wall thickness, without increased cavity size, usually inherited
Heterozygote	A diploid organism is heterozygous at a gene locus when its cells contain two different alleles of a gene on both homologous chromosomes. The cell or organism in question is called a heterozygote
HMERF	Hereditary myopathy with early respiratory failure, a disease which typically involves adult-onset proximal-distal weakness and respiratory failure in an ambulant patient
Homozygote	A diploid organism is homozygous for a particular gene when identical alleles of the gene are present on both homologous chromosomes. The cell or organism in question is called a homozygote
ICC	Inherited cardiac condition. A group of largely monogenic disorders affecting the heart, its conducting system and vasculature
Ig	Immunoglobulin domain, a repeated domain within the TTN protein
IHD	Ischaemic heart disease – “ischaemia” means inadequate blood supply to an organ. IHD encompasses a range of pathological processes from acute thrombosis suddenly occluding a coronary artery, to more gradual, progressive narrowing of large or small vessels supplying the heart
INR	International normalised ratio, a measure of the extrinsic pathway of coagulation.
ISFC	International Society and Federation of Cardiology
JHS	Jackson heart study
LA	Left atrium
LAV	Left atrial volume
LBBB	Left bundle branch block
LGE	Late gadolinium enhancement - detects accumulation of contrast in areas of infarction or fibrosis due to slower contrast kinetics and greater volume of contrast in the extracellular space
LGMD	Limb girdle muscular dystrophy
LMNA	Gene encoding LaminA/C, lamins are structural protein components of the nuclear lamina
LQT	Long QT syndrome – an inherited arrhythmia syndrome
LRG	Locus reference genomic sequence – a standardised framework for reporting the location of DNA variation
LV	Left ventricle

LVNC	Left ventricular non-compaction - a rare cardiomyopathy characterised by failure of compaction of the ventricles and a hypertrabeculated appearance
Mendelian	A Mendelian trait is one that is controlled by a single locus that is inherited according to Mendel's laws
MEx1/2	The six exons at the 3' end of the TTN gene encode the M-line section of titin and are referred to as Mex1 to Mex6.
MI	Myocardial infarction
MmD	Multi-minicore disease describes a recessive core myopathy (CM)
MRC	Medical research council
MRI	Magnetic resonance imaging, a type of scan that uses strong magnetic fields and radio waves to produce detailed images
mRNA	Messenger ribonucleic acid
MSA	Multiple sequence alignment
NGS	Next-generation sequencing
NHLBI	The National Heart, Lung, and Blood Institute, a division of the US National Institutes of Health
NHS	National Health Service (UK)
nsSNP	Non-synonymous SNP
PCR	Polymerase chain reaction, a biochemical technology used to amplify DNA
Penetrance	The proportion of individuals carrying a particular variant of a gene (allele or genotype) that also expresses an associated trait (phenotype)
PEVK	A region of Titin that is rich in proline, glutamate, valine and lysine
Phenotype	The composite of an organism's observable characteristics or traits, such as its morphology, development, biochemical or physiological properties, and behavior. A phenotype results from the expression of an organism's genes as well as the influence of environmental factors and the interactions between the two
Private mutation	A genetic variation that is unique to that individual or pedigree
PSI	Proportion spliced in - a score that provides the inclusion level of each exon
RBHT	Royal Brompton & Harefield NHS Foundation Trust
RCM	Restrictive cardiomyopathy
Recessive inheritance	A pattern of inheritance in which both copies of a gene must be aberrant in order for the disease or trait to develop.
RNA	Ribonucleic acid
RV	Right ventricle

SCD	Sudden cardiac death
SD	Standard deviations
Sequencing	Determining the primary structure of a biological polymer, particularly the sequence of nucleotides in DNA or RNA, or of amino acids in a protein molecule
SIFT	Sorts intolerant from tolerant. An online tool that predicts whether an amino acid substitution affects protein function, based on the degree of conservation of amino acid residues in sequence alignments derived from closely related sequences
SNP	Single nucleotide polymorphism
SV	Stroke volume is the volume of blood pumped from one ventricle of the heart with each beat
TK	Titin kinase
TMD	Tibial muscular dystrophy, an autosomal dominant distal myopathy characterised by weakness of the muscles of anterior compartment of lower limbs, appearing in the fourth to seventh decade of life. Caused by heterozygous mutations in the M-band of TTN
TR	Tricuspid regurgitation
UTR	untranslated region – DNA sequence at each end of a gene that is transcribed to RNA, but not translated to amino acid
VUS	Variant of unknown significance. A variation in a genetic sequence whose association with disease risk is unknown
WES	Whole exome sequencing
WGS	Whole genome sequencing
WHO	World health organisation
WT	Wall thickness

Table of contents

Abstract	2
Declaration of Originality and Copyright	3
Contribution of others to work presented in this thesis	4
Glossary of terms and abbreviations	9
1 Introduction	19
1.1 Inherited cardiovascular conditions and cardiomyopathy	19
1.2 Dilated cardiomyopathy	20
1.2.1 Epidemiology	20
1.2.2 Clinical presentation and consequences	21
1.2.3 Aetiology	21
1.2.4 Familial DCM and family screening	23
1.2.5 Genetic factors, molecular diagnosis and the impact of next generation sequencing (NGS)	24
1.3 Personalised medicine	27
1.3.1 Molecular diagnostics and stratified medicine	27
1.3.2 Stratifying risk in common diseases	28
1.3.3 Pharmacogenetics	28
1.4 Novel gene and variant discovery	29
1.4.1 Candidate gene studies	29
1.4.2 Linkage studies	29
1.4.3 Association studies	30
1.5 Sequencing strategies	30
1.5.1 Target enrichment	31
1.5.2 Whole exome sequencing	32
1.5.3 RNA-sequencing	34
1.5.4 Publically available control cohorts	34
1.5.5 Delineation and interpretation of variants	35
1.6 DCM as a disease of the cardiac myocyte	37
1.6.1 Overview of cardiac muscle and contractile function	37
1.6.2 Mechanisms underlying heart failure in DCM	38
1.7 Titin structure and function in the cardiomyocyte	40
1.8 Titinopathies	41
1.8.1 Dominant titinopathies are adult onset and affect either cardiac or skeletal muscle but not both	41
1.8.2 Recessive titinopathies have diverse phenotypes and typically involve the M-band.	45
1.8.3 Titin copy number variation	48
1.8.4 Titin animal models	48
1.9 Cardiovascular magnetic resonance (CMR) imaging	51
1.9.1 Standard CMR image acquisition	51
1.9.2 CMR most accurately measures ventricular morphology and function	52
1.9.3 CMR is highly reproducible and increases power to detect a true difference between patient populations	52
1.9.4 Gadolinium-chelated contrast agents aid discrimination of DCM from other causes of heart failure	55
2 Aims of this Thesis	56
3 Materials and methods	57
3.1 Patient and control cohort selection	57

3.1.1	Prospectively recruited unselected DCM subjects	57
3.1.2	End-stage DCM subjects	58
3.1.3	Healthy volunteers	60
3.1.4	Population Cohorts	60
3.2	CMR phenotyping	61
3.2.1	CMR of prospectively enrolled DCM subjects:	61
3.2.2	CMR of prospectively enrolled healthy volunteers:	61
3.3	DNA preparation and quality control	62
3.4	Target enrichment of TTN	62
3.4.1	Assay design	62
3.4.2	DNA library preparation	63
3.4.3	Next generation sequencing platforms	63
3.5	Whole genome sequencing	64
3.6	LV tissue studies	64
3.7	RNA sequencing	64
3.7.1	Exon expression estimation	65
3.8	Sanger sequencing	65
3.9	Bioinformatics	65
3.9.1	Bioinformatic analysis of targeted resequencing	65
3.9.2	Variant annotation	66
3.9.3	Splice variant prediction	83
3.10	Statistical analyses	83
3.10.1	Overview	83
3.10.2	Determining the likelihood that a TTNtv found in an individual with DCM is pathogenic	84
3.10.3	Stratification of DCM and linear modelling	84
4	Deep phenotyping of a DCM cohort	85
4.1	Introduction	85
4.2	Aims	86
4.3	Results	86
4.3.1	Establishing a DCM cohort using CMR phenotyping and comparison to the normalised mean for LVEF and LVEDVi	86
4.3.2	Exploration of CMR parameters in a DCM cohort and comparison to healthy volunteers	91
4.4	Discussion	98
4.4.1	Demographics	98
4.4.2	CMR parameters of morphology and function	99
4.4.3	Limitations of relying on CMR	100
4.5	Summary	101
5	Design and optimisation of the Titin sequencing assay	102
5.1	Introduction	102
5.2	Aims	103
5.3	Results	103
5.3.1	Definition of the target region and development of a TTN LRG	103
5.3.2	GC ratio affects hybridization efficiency	104
5.3.3	Sequencing depth affects overall assay performance	106
5.3.4	Assay design optimisation	107
5.3.5	Library preparation and NGS performance	110
5.3.6	Improving sequencing efficiency	112
5.4	Discussion	113
5.5	Summary	114

6 Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease	116
6.1 Introduction	116
6.2 Aims	116
6.3 Results	116
6.3.1 Comparison of sequencing capture between cohorts to assess validity of comparison	116
6.3.2 Burden of TTNtv in health and disease	118
6.3.3 Distribution of TTNtv in health and disease	118
6.3.4 Interpretation of TTNtv found in controls	133
6.3.5 Diagnostic interpretation of TTNtv in DCM	140
6.3.6 Stratification of DCM by TTN genotype	140
6.4 Discussion	147
6.4.1 Limitations of sequence data	147
6.4.2 Limitation of novex-3 interpretation	147
6.4.3 Limitations of deriving and applying exon usage values	148
6.4.4 Clinical stratification of DCM by TTN genotype	149
6.4.5 Interpretation of TTNtv in individuals without DCM	149
6.4.6 Insights into TTN DCM disease mechanism	150
6.5 Summary	151
7 Concluding remarks	152
7.1 Overview of thesis	152
7.2 Novelty	153
7.3 Future work	153
References	155

Table of Figures

Figure 1.1	Next generation sequencing (NGS) workflow.	33
Figure 1.2	Acquisition of the cine images and short axis stack	54
Figure 3.1	Schematic representation of study design	59
Figure 3.2	A unified representation of TTN transcripts.	67
Figure 4.1	Relation of LVEF and LVEDVi for the four most frequent diagnoses on initial CMR analysis	87
Figure 4.2	The effect of age and sex on continuous CMR parameters for the DCM cohort	90
Figure 4.3	Frequency density plots of CMR parameters for males and females in the DCM and healthy volunteer cohorts alongside probability density histograms with raw and smoothed density curves for the DCM cohort	93
Figure 4.4	Frequency density plots of CMR parameters for males and females in the DCM cohort alongside probability density histograms with raw and smoothed density curves	96
Figure 4.5	Overview comparison of continuous CMR parameters between DCM and healthy volunteer cohorts stratified by gender.	97
Figure 5.1	Callability and GC content of TTN exons.....	105
Figure 5.2	Relationship between GC content and callability for TTN exons.....	105

Figure 5.3	Relationship between sequencing depth and callability for target enrichment assays.....	106
Figure 5.4	Callability per exon for iterative assay designs	108
Figure 5.5	Beeswarm plot of callability for each assay design measured against the meta-transcript LRG_391_t1.	111
Figure 5.6	Beeswarm plot of callability stratified by sequencing run, measured against the meta-transcript LRG_391_t1.	111
Figure 5.7	Relationship between sequencing depth and callability for TTN.....	112
Figure 5.8	Relationship between sequencing depth and enrichment for TTN.....	113
Figure 6.1	TTN sequencing capture for each cohort.....	117
Figure 6.2	TTN exon usage in the heart and distribution of TTNtv in healthy volunteers, FHS and JHS participants, unselected and end-stage DCM patients.....	128
Figure 6.3	Sites susceptible to truncating events are non-uniformly distributed within the TTN gene but do not influence clustering effects in the A band.....	129
Figure 6.4	Usage of TTN exons across isoforms.....	131
Figure 6.5	Usage of TTN exons containing TTNtv across all subject groups	131
Figure 6.6	TTNtv and survival in DCM.....	141
Figure 6.7	Truncated transcript length is correlated with indices of cardiac impairment severity in DCM.	144
Figure 6.8	TTN exon usage in DCM and control hearts.....	148

Table of Tables

Table 1.1	Overview of factors that may cause secondary or syndromic DCM	22
Table 1.2	Putative DCM genes, their protein product and its cellular location/function.....	25
Table 3.1	Reference transcript and protein identifiers for TTN.....	68
Table 3.2	Overview of TTN transcripts and exon usage	69
Table 3.3	Minimum absolute score change for each prediction algorithm applied as threshold across all cohorts.....	83
Table 4.1	Full demographic data and results for categorical variables.....	89
Table 4.2	Comparison of means assesses the difference between DCM and healthy volunteer cohorts.	92
Table 5.1	Exons not fully captured by TTN single gene assays TTNv1 and TTNv2.....	109
Table 6.1	TTN truncating variants identified in UK prospective DCM cohort	119
Table 6.2	TTN truncating variants identified in the FHS Offspring cohort.	122
Table 6.3	TTN truncating variants identified in the JHS cohort	123
Table 6.4	TTN truncating variants identified in end-stage DCM.....	125
Table 6.5	TTN truncating variants identified in healthy volunteers.....	127
Table 6.6	Burden of TTN truncating variants in DCM patients and controls.....	132

Table 6.7	TTN truncating variants in publicly available control populations	134
Table 6.8	Burden, type and distribution of <i>TTN</i> truncating variants in publicly available control populations.....	139
Table 6.9	Clinical characteristics in DCM patients with and without TTNtv	142
Table 6.10	Linear Modelling of the relationship between TTN genotype and phenotype for 14 continuous variables in the unselected DCM cohort	145
Table 6.11	Full Linear Model describes impact of multivariate TTN genotype on phenotype for 14 continuous variables in the unselected DCM cohort	146

1 Introduction

1.1 Inherited cardiovascular conditions and cardiomyopathy

Inherited cardiac conditions (ICCs) are genetic conditions that affect the heart and circulation. These diseases are the commonest cause of sudden death in the young and are also a major cause of death and disability across all age groups. ICCs can be broadly grouped into those affecting the heart only (e.g. cardiomyopathies or channelopathies), and those affecting the heart and circulation (e.g. Marfan syndrome or familial hypercholesterolaemia).

Cardiomyopathy is an umbrella term describing a heterogeneous group of disorders characterised by structural and functional abnormalities of the myocardium, in the absence of coronary artery disease (CAD), hypertension, valvular disease or congenital heart disease sufficient to cause the observed myocardial abnormality.

Cardiomyopathies can be grouped into five forms: dilated (DCM), hypertrophic (HCM), restrictive (RCM), arrhythmogenic right ventricular cardiomyopathy (ARVC) and left ventricular non-compaction (LVNC). However, the precise classification of these cardiomyopathies and their subtypes has remained an area of contention owing to extensive genotypic and phenotypic overlap and our evolving understanding of the complex disease processes involved. One recent area of consensus is the departure from the WHO/ISFC classification and exclusion of left ventricular (LV) dysfunction as a direct consequence of another cardiac abnormality including hypertension, valve disease, congenital heart disease and CAD, which were historically termed 'specific heart muscle diseases' or 'specific cardiomyopathies'.

In 2006 the American Heart Association (AHA) revised their definition and classification of cardiomyopathies¹ and proposed their overarching division into primary cardiomyopathy (genetic, non-genetic and acquired) where disease is solely or predominantly confined to the heart muscle, and secondary cardiomyopathy showing myocardial involvement as part of a large number and variety of generalised systemic disorders (previously described as specific cardiomyopathies).

Interestingly they included ion channel disorders (including Long QT and Brugada syndromes) as cardiomyopathies 'on the basis of the scientifically reasonable (but largely hypothetical) assertion that ion channel mutations are responsible for altering the biophysical properties and protein structure, thereby creating structurally

abnormal ion channel interfaces and architecture'. The organ specific distinction proposed by the AHA is not ideal as many 'primary' cardiomyopathies have extra-cardiac manifestations, and pathology in many 'secondary' cardiomyopathies can predominantly or exclusively involve the heart.

Latterly, the European Society of Cardiology (ESC) published a clinically-orientated classification system in which heart muscle disorders are grouped according to ventricular morphology and function, and sub-classified as familial or non-familial². The rationale being that this is the most useful method for diagnosing patients and managing families. In this system familial disease is defined as the occurrence of the same disorder or phenotype in more than one family member which is or could be caused by the same genetic variant, or the occurrence of as yet sporadic disease due to a de novo genetic variant which could be transmitted to offspring. Non-familial cardiomyopathies are sporadic and may be sub-divided into idiopathic or acquired cardiomyopathy in which cardiac dysfunction is a complication and not a direct consequence of the disease.

1.2 Dilated cardiomyopathy

DCM is defined by the ESC as left ventricular dilatation and systolic dysfunction in the absence of abnormal loading conditions (hypertension, valve disease) or CAD sufficient to cause the level of systolic impairment observed. Typically the LV wall is thinned and there may be right ventricular dilatation and dysfunction but these are neither invariable nor required for the diagnosis.

1.2.1 Epidemiology

The prevalence of DCM remains uncertain and the only formal estimate is based on a study conducted from 1975 to 1984 which estimated DCM prevalence at approximately 1:2,700³. However, given that the same study underestimated the prevalence of HCM by tenfold, it is likely that the same may be true for DCM and it is now generally accepted that DCM prevalence may be as high as 1:500. Since subjects may remain asymptomatic until late stage disease, it is likely that DCM has been under-diagnosed. A recent review put forward further persuasive arguments that the prevalence is closer to 1 in 250⁴: extrapolating prevalence of non-ischaemic cardiomyopathy from heart failure clinical trial data, deriving the prevalence of

idiopathic systolic dysfunction identified by echocardiography from cross-sectional study data, and using population-based prevalence estimates of asymptomatic LV dysfunction as a surrogate for DCM. Given recent guidelines highlighting the importance of family screening, coupled with the implementation of more sensitive diagnostic imaging such as CMR, the frequency with which DCM is diagnosed is set to rise. The disorder may manifest at any age, in either sex and in any ethnic origin although it is more common in males and in blacks and typically presents in adulthood ⁵⁻⁹.

1.2.2 Clinical presentation and consequences

DCM is a major cause of morbidity and mortality and is the leading cause for heart transplantation¹⁰. Presentation is typically with symptomatic heart failure with features of congestion (oedema, orthopnoea, paroxysmal dyspnoea) and/or reduced cardiac output (fatigue, dyspnoea on exertion). Other complications include arrhythmias and/or conduction system disease and thromboembolic disease including stroke. At least half of all cases of heart failure have systolic dysfunction ¹¹, a substantial proportion of which results from DCM ¹⁰.

In the early stages various compensatory mechanisms act to allow the patient to remain asymptomatic, sometimes for years. Ultimately these homeostatic responses fail and a common path of decompensation and dilatation ensues leading to symptomatic heart failure ¹². Therefore symptomatic DCM represents late disease. Screening and identification of DCM before the onset of symptoms enables the initiation of medical therapy that may delay disease progression. However, cardiac transplantation remains the definitive treatment for progressive DCM and heart failure refractory to medical or device therapy.

1.2.3 Aetiology

DCM is the most clinically and genetically heterogeneous of the cardiomyopathies. Although regarded as a single entity characterised by LV dilatation and dysfunction, myocyte cell death, and fibrosis, it may be more accurately thought of as a final common phenotype for a diverse range of cellular processes that cause myocardial damage ¹³. Many common conditions such as ischaemic heart disease, and environmental factors such as alcohol, can lead to heart failure, and it is important to

asses these when making a diagnosis (Table 1.1) ^{1,2}. Unless there are characteristic histopathological or imaging findings, it can be difficult to prove that DCM is secondary to a particular factor or whether that factor is merely concurrent or modifying expression. The term idiopathic DCM was coined before genetic testing became routine and has in the past included DCM which is purely idiopathic as well as where a genetic cause has been identified.

Table 1.1 Overview of factors that may cause secondary or syndromic DCM

Toxins	Alcohol, amphetamines, cocaine Antiretroviral agents Carbon monoxide Chemotherapeutic drugs (e.g. anthracyclines) Chloroquine, phenothiazine Cobalt, lead, mercury
Infectious Agents	Bacterial Brucellosis, diphtheria, psittacosis, typhoid fever Fungal Mycobacterial Parasitic Chagas disease, schistosomiasis, toxoplasmosis Rickettsial Viral Adenovirus, Coxsackie A and B, cytomegalovirus, Epstein-Barr, human herpes virus 6, HIV, parvovirus B19, varicella
Metabolic Abnormalities	Electrolyte disturbances Hypocalcaemia, Hypophosphataemia, Uraemia Endocrine abnormalities Cushing's disease, Diabetes mellitus, Acromegaly Hypo/hyperthyroidism, Pheochromocytoma Nutritional deficiencies Carnitine, thiamine, selenium
Inflammatory/ infiltrative conditions	Hypersensitivity myocarditis Infiltrative diseases Haemochromatosis, Sarcoidosis

	Vasculitis Churg-Strauss, Kawasaki disease, Polyarteritis nodosa
Connective tissue diseases	Dermatomyositis Marfan syndrome Scleroderma Systemic lupus erythematosus
Neuromuscular Diseases	Duchenne /Becker muscular dystrophy Erb limb girdle muscular dystrophy Fascioscapulohumeral muscular dystrophy Friedreich's ataxia Myotonic dystrophy
Pregnancy	
Tachyarrhythmia	

1.2.4 Familial DCM and family screening

DCM has been estimated to be familial in 20% to 50% of cases ¹⁴. A medical history, physical examination, echocardiogram, and electrocardiogram (ECG) should be used to evaluate a proband's first-degree relatives to determine if any have asymptomatic DCM. It is recommended that first-degree relatives with a normal echocardiogram and ECG be re-screened every three to five years to fully assess their risk. However, screening of relatives based on phenotype is unreliable and requires long-term and often invasive follow up. The variable penetrance typical of DCM can considerably complicate family screening, which is stressful for patients, costly and time-consuming. Molecular diagnosis could aid both treatment and prognostication for the proband, and accurate diagnosis of family members. Genetic testing is available for diagnostic use by clinicians in the UK, primarily for familial DCM. However, although genetic testing of probands and cascade screening of families with ICCs is recommended by the National Service Framework for Coronary Heart Disease ¹⁵ and the National Institute of Clinical Excellence ¹⁶ these tests are not often used.

Genetic testing plays an important role in the management of inherited diseases, but there are many caveats. One area of concern is genetic testing in children. A negative test may free a child from follow-up or invasive investigation but a positive result may have complex ramifications. For instance, many ICCs have low penetrance, so a positive result could result in lifestyle restrictions and anxiety for a

child who may never become symptomatic. On the other hand, delayed testing for childhood-onset conditions may lead to a delay in appropriate management. These potential benefits and burdens must be considered carefully.

1.2.5 Genetic factors, molecular diagnosis and the impact of next generation sequencing (NGS)

Disease aetiology lies on a continuous spectrum from genetic to environmental. A simple and purely genetic disease arises when a fully penetrant mutation in a single gene is sufficient to produce disease and is inherited in a Mendelian fashion.

Mutations are typically in protein-coding regions of the genome, and cause substantial changes in protein function by altering splicing, truncating the protein, altering the reading frame, or substituting amino acids within critical functional domains.

Most genetic DCM is inherited in an autosomal dominant fashion although autosomal recessive, X-linked and mitochondrial inheritance all occur. DCM is the end phenotype of diverse mutations in heterogeneous pathways ranging from components of the membrane-scaffolding apparatus, sarcomeric proteins, nuclear envelope proteins, calcium-handling proteins, transcription cofactors, RNA splicing and cell energy-generating machinery¹³. Table 1.2 shows known DCM genes and their cellular location/function.

Table 1.2 Putative DCM genes, their protein product and its cellular location/function

Gene name	Protein	Location/function
ABCC9	ATP-binding cassette C9	Ion channel
ACTC1	Cardiac Actin	Sarcomere
ACTN2	Actinin, alpha 2	Z-disk
ALMS1	Alstrom syndrome1	Cytoskeleton/centrosome/cilia
ANKRD1	Ankyrin repeat protein 1	Transcription factor
APOA1	Apolipoprotein A-I	Lipoprotein
CAV3	Caveolin-3	Plasma membrane
CRYAB	Alpha B crystallin	Z-disk
CSRP3	Cys and gly-rich protein 3	Z-disk
CTF1	Cardiotrophin 1	Cytokine
DES	Desmin	Desmosome
DMD	Dystrophin	Dystrophin associated protein complex
DNAJC19	HSP40 Homolog, C19	Mitochondrial membrane
DSG2	Desmoglein 2	Desmosome
DSP	Desmoplakin	Desmosome
DTNA	Alpha-Dystrobrevin	Dystrophin associated protein complex
EMD	Emerin	Nuclear envelope
EYA4	Eyes absent 4	Transcription factor
FHL2	Four and half LIM protein 2	Z-disk
FKRP	Fukutin-related protein	Golgi/glycosyl
FKTN	Fukutin	Golgi/glycosyl
FLT1	Fms-related tyrosine kinase 1	Kinase
FOXD4	Forkhead box D4	Transcription factor
FRDA	Frataxin	Mitochondrial Fe chaperone
HADHA	L. Chain 3-OH-Ac dehydrogenase.	Fatty acid synthesis
HFE	Hemochromatosis gene	Fe transport
HOPX	Homeobox only protein	Transcription factor
ILK	Integrin-linked kinase	Kinase
LAMA2	Laminin alpha2	Extra cellular matrix
LAMA4	Laminin alpha4	Extra cellular matrix
LAMP2	Lysosome-Associated Membrane Protein 2	Lysosome
LDB3	Cypher/ZASP	Z-disk
LMNA	Lamin A/C	Nuclear envelope
MYBPC3	Myosin-binding protein C3	Sarcomere
MYH6	Alpha myosin heavy chain	Sarcomere
MYH7	Alpha myosin heavy chain 7	Sarcomere
NEXN	Nexilin	Z-disk
NFKB1	NF-Kappa B1	Transcription factor
PDLIM3	PDZ/LIM domain protein 3	Z-disk
PLEC1	Plectin-1	Inter filament
PLN	Phospholamban	Ca ²⁺ cycling
PSEN1	Presenilin 1	Membrane protein/protease
PSEN2	Presenilin 2	Membrane protein/protease
RBM20	RNA-Binding motif protein 20	RNA binding

SCN5A	Voltage gated Na channel V alpha subunit	Ion channel
SDHA	Succinate dehydrogenase	Mitochondrial
SGCA	Alpha-Sarcoglycan	Dystrophin associated protein complex
SGCB	Beta-Sarcoglycan	Dystrophin associated protein complex
SGCD	Delta-Sarcoglycan	Dystrophin associated protein complex
SGCG	Gamma-Sarcoglycan	Dystrophin associated protein complex
SOD2	Superoxide dismutase	Mitochondrial matrix
SYNE1	Nesprin-1	Nuclear envelope
SYNM	Desmulin	Z-disk
TAZ	Tafazzin	Mitochondrial
TBX20	T-box 20	Transcription factor
TCAP	Titin-cap	Z-disk
TMPO	Thymopoietin	Nuclear envelope
TNNC1	Troponin C	Sarcomere
TNNI3	Troponin I	Cardiac sarcomere
TNNT2	Troponin T	Cardiac sarcomere
TPM1	Tropomyosin, alpha1	Sarcomere
TTN	Titin	Sarcomere
VCL	(meta)Vinculin	Z-disk

Within the 50+ genes reputed to cause DCM, mutations are often private (unique to one pedigree). At the time of commencing this project only 6 DCM genes were available for clinical genetic testing in the UK (<http://www.ukgtn.nhs.uk> accessed 8th March 2012) and according to published frequencies these genes account for 12-17% of DCM. It is likely that the remainder of the genetic burden is derived from a large number of genes. However, the frequency of genetic causation in persons with sporadic DCM remained largely unknown, and mutation frequencies provided for most known DCM genes should be considered preliminary, as usually only 1 or 2 primary reports focusing on single genes are available from which to estimate frequencies. Previously reported in 3 pedigrees, Titin (TTN) mutations were recently reported in 18-25% of selected DCM cases in a large multicenter study¹⁷: a discovery which could significantly change the landscape of genetic testing and family screening in DCM.

NGS diagnostics for ICCs will most likely considerably reduce costs and increase availability if implemented in high-volume centres. Accredited NGS analysis of cardiomyopathy genes has been available abroad for several years (e.g. GeneDx (www.genedx.com) offered a panel of 23 DCM genes in October 2009 and now offers a comprehensive cardiomyopathy panel including 36 DCM genes). During

the course of this project an NGS cardiomyopathy panel became available for the first time in the UK (<http://www.ukgtn.nhs.uk>, <http://www.ouh.nhs.uk/services>). The benefits of NGS, in terms of a more comprehensive genetic analysis at reduced cost, are likely to be striking for genetically heterogeneous conditions such as DCM.

Determining whether a particular variant causes disease is one of the greatest challenges of contemporary clinical genetics, largely due to our limited understanding of the spectrum of rare but non-pathogenic genetic variation. This will become an increasing challenge as we are inundated with NGS data from disease-causing genes from research and diagnostics laboratories. Geneticists have traditionally relied on identifying co-segregation of a candidate variant with disease phenotype in a sufficiently large family, but at the population level this is not possible and we will increasingly need large and well-annotated databases to ascribe probabilistic disease scores to variants. Assigning causality to a variant is further complicated as many ICCs are not completely penetrant and variants in several genes may act together to cause disease or modify expressivity.¹⁸⁻²⁰ NGS analysis of comprehensive panels of genes associated with a phenotype represents a timely advance, providing the tools to address these shortcomings, but also presenting major informatic challenges.

Consequent upon the lack of comprehensive studies of DCM genetics, limited data is available in the literature for genotype/phenotype correlation in DCM. In the new era of NGS far more longitudinal data can be gathered on these types of individuals, and together with an increased understanding of the role of disease modifying genes, may allow clinicians to better advise and treat patients.

1.3 Personalised medicine

1.3.1 Molecular diagnostics and stratified medicine

Human disease is heterogeneous and can be usefully classified into subtypes based on features that predict disease progression and treatment response—for example, type 1 versus type 2 diabetes or ST elevation versus non-ST elevation myocardial infarction. Identification of the specific molecular mechanism underlying an inherited disease provides another level of information for stratified medicine approaches. Molecular classification has enabled subtype-specific treatment for the three

commonest LQT syndromes,²¹⁻²⁶ genotype-specific sudden cardiac death (SCD) risk prediction in HCM²⁷ and the identification of rapidly progressing subtypes of DCM requiring early transplant assessment²⁸⁻³⁰. Until now, no individual gene has been identified as a common enough cause of DCM to allow systematic genotype-phenotype correlation. As we gather more phenotype-genotype profiles in rare diseases (e.g. LQT 4-13^{21,24}), genotype-specific risk calculators may be developed. Treatment may also be targeted to correct a specific molecular defect, as is already the case for rare cardiovascular diseases including monogenic hypertension and monogenic type II diabetes.^{31,32}

1.3.2 Stratifying risk in common diseases

Recent genetic studies have identified a number of genetic variants that are associated with common cardiovascular diseases, including coronary artery disease, hypertension, diabetes, and arrhythmia.^{33,34} A number of studies have assessed the utility of genetic risk stratification by comparing the predictive value of conventional risk factors against risk factors plus genotype, but the additional power from current genetic markers is marginal.³⁵ This is because the effect size associated with individual genetic variants is typically very small. Conventional risk factors often also assess family history, and a family history effectively interrogates many more genetic loci than can be meaningfully tested in a laboratory³⁶. In the near future, deep sequencing with NGS will probably identify (rare) variants of larger effect size than those currently known, which may be more informative of risk. In situations where conventional risk stratification is limited, such as assessing risk of sudden cardiac death, NGS assessment of risk may be particularly informative.³⁷

1.3.3 Pharmacogenetics

Of particular interest to clinical cardiologists are genotype-specific drug responses for commonly used drugs such as clopidogrel, warfarin and statins. Common genetic variants in CYP2C19 reduce conversion of clopidogrel to its active metabolite,³⁸ and studies have suggested that poor metabolisers treated with clopidogrel after acute myocardial infarction suffer higher rates of subsequent cardiovascular events than normal metabolisers,^{39,40} leading the FDA to issue a black box warning advising physicians of the availability of CYP2C19 genotype testing.⁴¹ While subsequent studies^{42,43} have led to divergent opinions on the

interpretation of these results, it remains a powerful illustration of the application of genomic information in routine clinical practice. In the case of warfarin therapy it is clear that genotype can predict the dose required to achieve target INR but there is not yet evidence of associated improved clinical outcome.^{44,45} For statins, genotype identifies a subpopulation at increased risk of specific statin-induced myopathy, for whom an alternative lipid-lowering therapy may be more appropriate.⁴⁶ It is clear that the next generation cardiologist will increasingly rely on pharmacogenetics to guide therapy, often at the point of care.

1.4 Novel gene and variant discovery

Various different approaches have been employed to understand the genetics of human disease. Historically candidate gene studies and linkage studies have been the mainstays of Mendelian gene discovery. These genetic approaches follow the rare disease-rare variant hypothesis that rare diseases are caused by variants of large effect in one or a few genes⁴⁷. In contrast genomic approaches, such as association studies, are driven by the common disease-common variant hypothesis, that predisposition to common diseases such as hypertension⁴⁸ is due to many common variants of modest effect acting in concert {Holm, 2007 #277;Lander, 1996 #282;Reich, 2001 #283}.

1.4.1 Candidate gene studies

Candidate gene studies use sequencing of a small number of genes in a well-characterised disease group. The frequency of rare variants (mutation burden) identified is compared with that of the control group and these are further classified in terms of likely pathogenicity. Limitations of this kind of study are the sequencing capacity required and that the candidate genes are predetermined based on existing knowledge of disease pathobiology {Altshuler, 2008 #287;Moulik, 2009 #273;Meyer, 2013 #274;Watkins, 2003 #275;Seidman, 2001 #276}.

1.4.2 Linkage studies

Linkage studies rely on linkage disequilibrium, recombination frequencies and genetic maps⁵⁶. Genetic variants spread throughout the genome are used as markers for haplotypes and thus the sections of genome that segregate with the

condition can be followed⁵². Once the locus linked to the condition is identified it is sequenced to identify the causative variant. This approach requires large pedigrees and/or distantly-related affected individuals, as large regions of the genome are likely to co-segregate by chance in small pedigrees⁵⁷. Sanger sequencing of areas of this magnitude is prohibitively costly and time-consuming, but is more accessible with NGS.

1.4.3 Association studies

Following completion of the human genome project in 2003⁵⁸ and the HapMap in 2005⁵⁹ a new approach became available. The HapMap project was an international effort to identify and catalogue genetic variation across the genome between individuals in four distinct populations. Imputation of untyped variants can be used to minimise the number of SNPs directly tested while maximising the information gained⁵⁶. Genome wide association studies (GWAS) use genotyping chip technology, such as Affy6 SNP chip (Affymetrix), to catalogue genome wide variation in thousands of individuals cost-effectively and ascertain associations between variants and disease⁶⁰. The effect sizes of associations between variant and disease tend to be small, so initial hopes of predicting risk of heart disease by genotyping a handful of variants have not been realised³³. Notwithstanding, a variant with a small odds ratio can still identify a new disease pathway and offer new insights into pathobiology⁶¹. There have been many GWAS studies in common cardiac conditions and cardiovascular quantitative traits^{34,62-71} but due to the large cohorts required very few in ICCs and only in two DCM^{72,73}. As yet many loci have been mapped but few causative genes have been pinned down^{74,75}. NGS will facilitate re-sequencing these loci looking for the causative variants.

1.5 Sequencing strategies

Since the 1970s the principal technique used for DNA sequencing was the Sanger method,⁷⁶ and this remains the gold standard for sequencing accuracy today. The current generation of automated high-throughput Sanger sequencing machines read ~2 million bases of sequence per day and were used to sequence the first human genome^{77,78}. This throughput is very low compared with the size of the human genome (haploid = 3×10^9 base pairs), to overcome this, NGS technologies parallelise the sequencing process, thereby massively increasing throughput (up to

~50 billion bases per day) at a reduced cost per base (Figure 1.1). In the short term, Sanger sequencing will continue to be used as a complementary strategy to sequence genomic regions that are difficult to target in these ways (e.g., regions with high GC content or repetitive sequence) and also to validate NGS findings. As the accuracy and efficiency of targeted sequencing improves, it is anticipated that NGS will increasingly be used as a stand-alone technique⁷⁹.

Clinical laboratories face major challenges as a consequence of the high throughput of novel NGS approaches: optimising the use of capacity, processing and storing the large volumes of data and analysing the results, requiring a team of informaticians, statisticians, clinical scientists and clinicians^{80,81}.

1.5.1 Target enrichment

While NGS per-base sequencing costs are relatively low, the cost of sequencing entire human genomes remains high and is unnecessary for many diagnostic and research applications. Unlike Sanger sequencing, NGS has no innate specificity. Such specificity can be achieved by preparing a library of DNA fragments for sequencing, typically either by pooling many PCR amplicons or by fragmenting genomic DNA and using sequence-specific hybridisation to select genomic regions of interest (Figure 1.1)⁷⁹. Neither approach is straightforward on the scales appropriate for NGS applications, though off-the-shelf target enrichment products that integrate well with NGS workflows are now available. Sequencing capacity can also be shared between several samples, by tagging each sample with a unique DNA “barcode”. In this way it is possible to use a single NGS run to sequence either a handful of genes from many patients, or many genes from fewer patients.

In brief, a library of DNA molecules for sequencing is first prepared, followed by ligation of a platform-specific adaptor sequence that provides the sequencing start point, independent of the sequence-content of target molecules. Molecules are clonally amplified, immobilised on a solid surface, and sequenced in parallel by synthesising DNA complementary to the immobilised template while observing the incorporation of each base by fluorescence imaging⁷⁹. Clonal amplification occurs bound to beads in emulsion PCR (ePCR) wherein fragment library, PCR reagent and

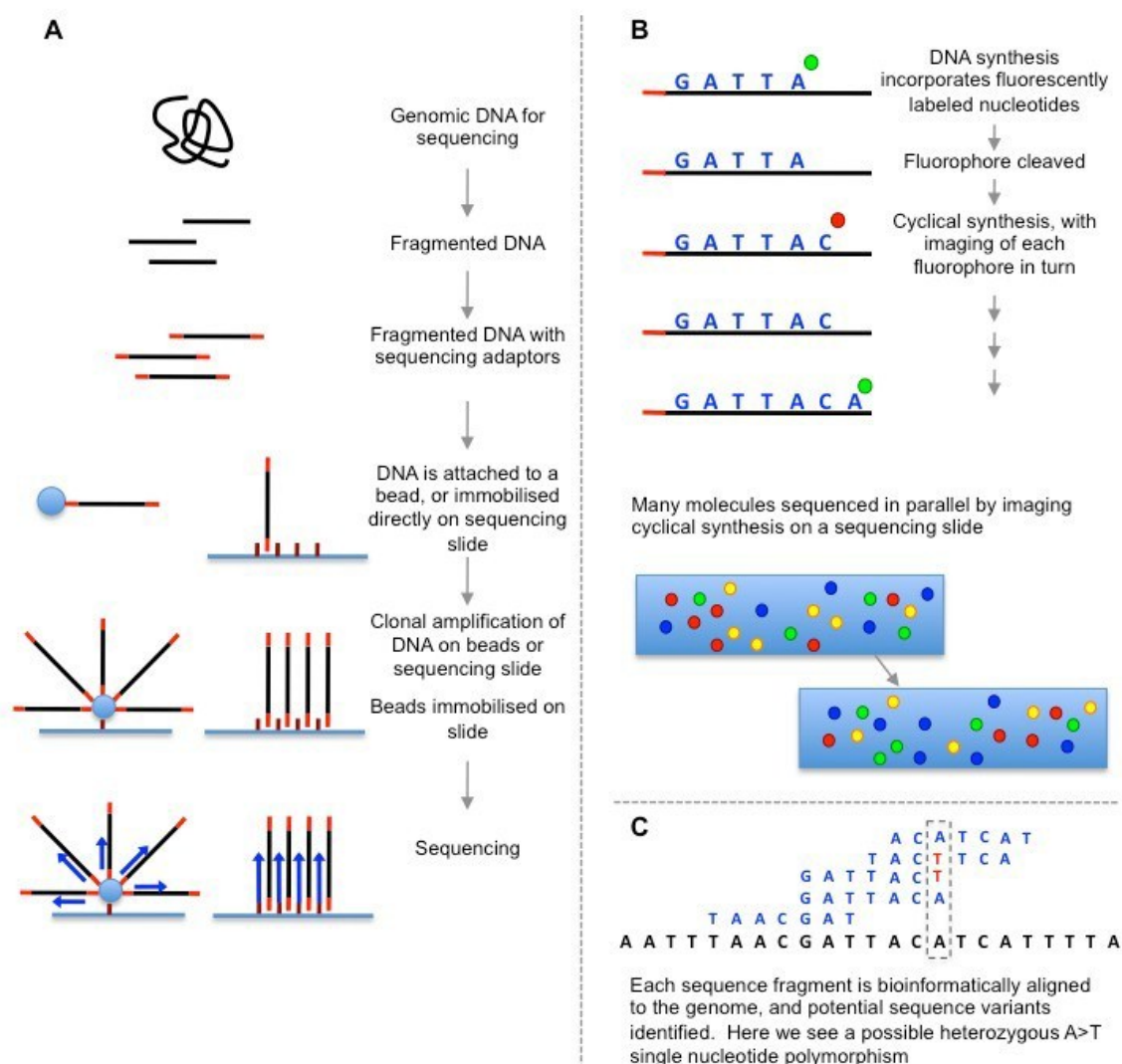
beads are emulsified with oil in a ratio that promotes a single fragment binding to each bead.

1.5.2 Whole exome sequencing

It is estimated that protein-coding genes constitute only approximately 1% of the human genome but harbour 85% of the mutations with large effects on disease-related traits. This makes targeted re-sequencing of all exons for an individual, whole exome sequencing (WES), an attractive and practical option. Off-the-shelf, whole exome target enrichment products lead to a flurry of publications identifying novel disease genes {Ng, 2009 #19;Choi, 2009 #2;Musunuru, 2010 #93;Haack, 2010 #115;Ng, 2010 #100;Wang, 2010 #98}. Direct sequencing of exomes has been shown to be an elegant and efficient way of searching for the causative gene for monogenic disorders ⁸⁸. This approach combines the benefits of the candidate gene approach in its ability to detect all variants, including rare and novel variants, with the unbiased nature of whole genome association studies. Exome sequencing is not without limitations. Since by its very design it targets only the coding region of the genome it does not afford interrogation of many variants that may be important for controlling gene transcriptional regulation or splicing. Neither is it currently able to reliably and comprehensively represent structural variation.

Figure 1.1 Next generation sequencing (NGS) workflow.

(A) Genomic DNA is fragmented and platform specific 'adaptors' are attached. The DNA is then either attached to a bead or directly to the sequencing slide. In either case, the DNA is clonally amplified in this location to provide a cluster of molecules with identical sequences. If beads are used they are then immobilised on a sequencing slide. Different NGS platforms employ different sequencing chemistries. (B) One approach to sequencing by synthesis, as employed by the Genome Analyser system (Illumina). The sequence of each fragment is read by decoding the sequence of fluorophores imaged at each physical position on a sequencing slide. Advanced optics allow for massively parallel sequencing. (C) Each DNA molecule yields one or two sequence fragments depending on whether it is sequenced from one or both ends. These sequence fragments are computationally aligned with a reference sequence and mismatches identified. Reproduced from Ware, Roberts & Cook⁷⁹



1.5.3 RNA-sequencing

Sequenced based approaches to determine the cDNA sequence initially involved Sanger sequencing of cDNA or expressed sequence tag (EST) libraries which was low throughput, costly and generally not quantitative. RNA-sequencing has benefited from the advent of NGS technologies in line with DNA sequencing, and NGS experiments follow the same principle as for DNA. In brief, a population of RNA (typically poly(A)+ enriched) is converted to a library of cDNA fragments with platform specific adaptors attached. Each molecule, with or without amplification, is then sequenced on an NGS platform to obtain sequences from both ends (pair-end sequencing). The resulting sequence reads are typically aligned with the reference genome or transcriptome, and classified as three types: exonic reads, junction reads and poly(A) end-reads. These three types are used to generate a base-resolution expression profile for each gene⁸⁹. Current RNA-seq technologies and analysis tools provide the ability to assess novel genes, alternatively spliced transcripts, post-transcriptional modifications, gene fusion, mutations/SNPs and changes in gene expression without relying on prior annotations⁹⁰.

1.5.4 Publically available control cohorts

Large publicly available control cohorts are an invaluable resource for investigating the differences in variant burden and type in case control studies. These data can be used to prioritise a list of candidate genes or to estimate the probability of pathogenicity of variants. Two well known examples are detailed below.

1.5.4.1 NHLBI GO Exome Sequencing Project (ESP)

The National Heart, Lung and Blood Institute (NHLBI)-funded ESP (esp.gs.washington.edu) is one of the largest medical sequencing studies ever undertaken. The group includes some of the largest well-phenotyped populations in the United States, representing more than 200,000 individuals altogether from the: Women's Health Initiative (WHI), Framingham Heart Study (FHS), Jackson Heart Study (JHS), Multi-Ethnic Study of Atherosclerosis (MESA), Atherosclerosis Risk in Communities (ARIC), Coronary Artery Risk Development in Young Adults (CARDIA), Cardiovascular Health Study (CHS), Lung Health Study (LHS), COPD Genetic Epidemiology (COPDGene), Severe Asthma Research Project (SARP), Pulmonary Arterial Hypertension population (PAH), Acute Lung Injury cohort (ALI), Cystic Fibrosis cohort (CF), PennCATH, Cleveland Clinic Genebank, Massachusetts

General Hospital Premature Coronary Artery Disease Study (MGH PCAD), Heart Attack Risk in Puget Sound (HARPS), and Translational Research Underlying Disparities in Myocardial Infarction Patients' Health Status (TRIUMPH). The current version dataset is comprised of a set of 2203 African-American and 4300 European-American unrelated individuals, totalling 6503 samples (13,006 chromosomes). The frequency counts of specific variants are released without linked phenotype information.

1.5.4.2 1000 Genomes project (1KG)

The goal of the 1000 Genomes Project (www.1000genomes.org) is to identify the majority of genetic variants, including structural variation, that have frequencies of at least 1% in the populations studied. The samples are mostly anonymous and have no associated medical or phenotype data. The pilot study included three projects: low-read depth whole-genome sequencing of 179 individuals from four populations; high-read depth sequencing of two mother–father–child trios; and exon-targeted sequencing of 697 individuals from seven populations. The plan for the full project is to sequence about 2,500 samples at 4x read depth^{91,92}.

1.5.5 Delineation and interpretation of variants

To assign pathogenicity to a novel variant, particularly in a gene or pathway not previously associated with a given disease requires two levels of evidence. First, is the variant predicted or proven to alter structure, function, expression or stability of the gene product? Secondly if there is a discernable effect on the protein, does this lead to disease? The following in silico tools can aid in prediction of the impact of a variant, but confirmation requires additional tissue or animal model studies.

1.5.5.1 Tools useful in assessing pathogenicity of novel variants in known disease genes

SIFT (sift.jcvi.org), Polyphen2 (genetics.bwh.harvard.edu/pph2) and Condel (bg.upf.edu/group/projects/condel.php) can be used to predict the effect of variation on protein function by amino acid substitution. SIFT (sorts intolerant from tolerant) predicts whether an amino acid substitution affects protein function, based on the degree of conservation of amino acid residues in sequence alignments derived from

closely related sequences. PolyPhen-2 (Polymorphism Phenotyping v2) is a tool which predicts possible impact of an amino acid substitution on the structure and function of a human protein using sequence, phylogenetic and structural information. Condel (Consensus deleteriousness score of missense SNVs) computes a weighed average of the scores of five known computational tools (SIFT, Polyphen2, MAPP, LogR Pfam E-value and Mutation Assessor) aimed at classifying missense mutations as likely deleterious or likely neutral.

1.5.5.2 Tools useful in assessing novel candidate genes

GeneCards (www.genecards.org) is a searchable, integrated, database of human genes that provides concise genomic related information, on all known and predicted human genes. The GeneCards human gene database extracts and integrates a carefully selected subset of gene related transcriptomic, genetic, proteomic, functional and disease information, from dozens of relevant sources

UniProt (www.uniprot.org) is a comprehensive, high-quality and freely accessible resource of protein sequence and functional information. It includes complete and reference proteome sets. The protein knowledgebase, UniProtKB, consists of two sections: Swiss-Prot, which is manually annotated and reviewed; and TrEMBL, which is automatically annotated and is not reviewed.

The Basic Local Alignment Search Tool (BLAST)

(<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

IntAct (www.ebi.ac.uk/intact) provides a freely available, open source database system and analysis tools for protein interaction data. All interactions are derived from literature curation or direct user submissions and are freely available.

ELM (www.elm.eu.org) is a computational biology resource for investigating candidate functional sites in eukaryotic proteins. ELM provides core functionality

including filtering by cell compartment, phylogeny, globular domain clash (using the SMART/Pfam databases) and structure. In addition, both the known ELM instances and any positionally conserved matches in sequences similar to ELM instance sequences are identified and displayed.

Coexpresdb (<http://coexpresdb.jp>) uses microarray expression data to provide co-regulated gene relationships to estimate gene functions.

STRING (www.string-db.org) is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations; they are derived from four sources: genomic context, high-throughput experiments, (conserved) co-expression, and previous knowledge. STRING quantitatively integrates interaction data from these sources for a large number of organisms, and transfers information between these organisms where applicable.

1.6 DCM as a disease of the cardiac myocyte

1.6.1 Overview of cardiac muscle and contractile function

The myocardium is involuntary striated muscle composed primarily of cardiomyocytes which exist in a complex network of endothelial cells, vascular smooth muscle, fibroblasts and immune cells. Gap junctions electrochemically coordinate contraction at the level of the cardiomyocyte and through connection to the extracellular matrix transduce force and coordinate contraction at the organ level⁹³. The proteins of the cardiomyocyte can be broadly categorised as contractile, sarcomeric skeleton, membrane associated and proteins of the intercalated disk⁹⁴. The sarcomere is the basic repeating contractile unit of muscle and is formed by an array of overlapping thin and thick filaments which respond to calcium to coordinate contraction and relaxation of the heart^{95 96}.

The sarcomere is demarcated at each end by a Z-disk and in the centre by the M-line. The Z-disk is a complex of ordered filaments, crosslinks and anchored proteins with diverse functions including signalling and transmission of tension during contraction⁹⁷⁻⁹⁹. Similarly the M-line is responsible both for the regular packing of thick filaments and the uniform distribution of tension^{100 101}. The region of the sarcomere containing the Z-disk and thin filaments is the I-band, and the remaining

region containing the M-line is the A-band. The primary component of the thin filament is actin, monomers of which polymerise into an elongated filament of double helical strands constructed upon the filamentous protein nebulin. Tropomyosin, a coiled-coil protein, polymerises along the actin filament and situates in the groove between the actin strands. A troponin complex comprised of Troponins C, I and T associates with each tropomyosin subunit ¹⁰². The main component of the thick filament is myosin. The complete myosin molecule is composed of six polypeptide chains consisting of two heavy chains, and two pairs of different light chains, the essential and the regulatory light chains. Myosin aligns with titin and myosin binding protein C to form the thick filament ¹⁰². Myosin has a long tail and a globular head, the direction of the myosin heads change polarity at the M-line, allowing interaction with the thin filaments anchored from the next adjacent Z-disk ⁹⁵.

T- tubules, deep invaginations of the sarcolemma, track laterally to the Z-disks and pair with a terminal cisterna of the sarcoplasmic reticulum at the Z-disk to form a diad, facilitating excitation-contraction coupling upon which cardiac contraction is dependant. The initial action potential at the sarcolemmal membrane stimulates an influx of ions through L-type calcium channels, in turn triggering calcium-induced calcium release from the sarcoplasmic reticulum and a marked rise in intracellular calcium. This calcium is bound by troponin C, thus altering thin filament protein interactions and allowing formation of strong cross bridges between actin and myosin. The myosin head acts as a lever; hydrolysis of ATP produces force and the relative movement of myosin along actin, thus shortening the sarcomere. Subsequent release from Troponin C and re-sequestration of calcium in the sarcoplasmic reticulum allows relaxation ¹⁰².

1.6.2 Mechanisms underlying heart failure in DCM

DCM is associated with ventricular remodelling, a complex multimechanistic process characterised by alteration in ventricular size, shape and function following cardiac injury. The process is linked to the activation of various neurohormonal, paracrine and autocrine factors and influenced by haemodynamic load and wall stress. Cardiac remodelling is driven on a histologic level by myocyte hypertrophy and apoptosis, fibroblast proliferation and interstitial fibrosis ¹⁰³.

The histological changes associated with DCM are frequently non-specific and not all features may be present, but classically include myocyte hypertrophy, myocyte atrophy, and increased fibrosis ¹⁰⁴⁻¹⁰⁷. Immunohistological examination may reveal interstitial and perivascular inflammatory cell infiltrates ^{108,109}. Ultrastructural changes include myofilament depletion, enlarged and bizarrely-shaped nuclei, proliferation of T-tubules, and accumulation of lipid droplets and glycogen ¹⁰⁵.

Energy deficiency is a key concept in heart failure. The heart requires a large amount of energy to sustain both cellular homeostasis and contraction. Energy in the form of ATP synthesised in the mitochondria can be derived from a variety of substrates, such as glucose, fatty acids, amino acids and lactate. In the healthy heart, the preferred energy substrate is free fatty acids and phosphocreatine is the main reserve source of ATP during acute stress. During the initial compensatory stages of cardiomyopathy, substrate utilization continues to favour free fatty acids, which yield about three times more ATP than glucose but require more oxygen to metabolize ^{110 111}. As pathogenesis progresses in cardiomyopathy, the demand for ATP increases, and the levels of phosphocreatine are progressively depleted ¹¹². During progression to decompensated heart failure energy substrate utilization shifts from fatty acid oxidation to glucose utilization, a less efficient means of producing energy ¹¹³⁻¹¹⁵. When the failing heart becomes energy deficient it can no longer sustain cellular homeostasis; the disturbance of energy metabolism, contractile function and gene expression can trigger programmed cell death ^{111 116,117}.

All three types of cell death are important in heart failure: apoptosis, necrosis and autophagic cell death ^{118 119}. Apoptosis is characterized by cell shrinkage and fragmentation into membrane enclosed apoptotic bodies that are phagocytosed. Plasma membrane integrity and organelle (e.g. mitochondrial) morphology are maintained until late in the process, and inflammation is avoided. Apoptosis is rare in normal human myocardium but marginally elevated in failing hearts ¹²⁰. Given the limited ability for cardiac myocytes to proliferate it has been postulated that the rate of apoptosis seen in failing hearts can be an important component of HF pathogenesis ¹²¹.

Necrosis has traditionally been regarded as passive and unregulated, however studies indicate that it is also regulated in some instances. The defining features of

necrosis are defective plasma and organelle membranes, cell and organelle swelling, severe ATP depletion, and marked inflammation¹²². Distinct, but highly overlapping central pathways mediate apoptosis and necrosis. The extrinsic pathway involves cell surface death receptors and the intrinsic pathway uses the mitochondria and endoplasmic reticulum. These pathways, which mediate both apoptosis and necrosis, are linked by multiple biochemical and functional connections¹²³. Although not extensively explored, initial studies suggest a role for programmed necrosis in the pathogenesis of heart failure^{124 125}.

Autophagy is a cell survival mechanism that involves degradation and recycling of cytoplasmic components. This provides energy during periods of starvation and stress, a means for the disposal of long-lived proteins, and a mechanism for protein quality control¹²⁶. Autophagy may mediate cell death under specific circumstances. A recent study indicated that upregulation of autophagy in the failing heart may also represent an adaptive response¹²⁷.

1.7 Titin structure and function in the cardiomyocyte

TTN is the largest protein in the human body and plays a key role in the sarcomere. Several isoforms ranging in size from 5,604 to 34,350 aa are produced from among the 364 exons of the single TTN gene which undergoes extensive alternative splicing. TTN is highly modular with ~90% of its mass composed of repeating immunoglobulin (Ig) and fibronectin-III (FN-III) domains. The remaining ~10% consists of 17 unique, non-repetitive sequence motifs (situated between the Ig and FN-III modules) that contain several phosphorylation sites, PEVK motifs and a carboxy terminal kinase domain¹²⁸. Two TTN filaments with opposite polarity span each sarcomere along its longitudinal axis forming a contiguous system within myofibrils. At its N terminus TTN is embedded in the Z-disk of the sarcomere and is important in myofibril assembly, stabilization and maintenance¹²⁹. The elastic I-band region behaves as a bidirectional spring, restoring sarcomeres to their resting length after systole and limiting their stretch in early diastole¹³⁰. The inextensible, thick filament binding A-band region is thought to be critical for biomechanical sensing and signalling and contains the TTN kinase domain^{131,132}. The kinase domain in particular may play a role in strain sensitive signalling and communication with the nucleus, which may effect gene expression and cardiac remodelling in DCM^{133,134}.

The major isoforms expressed in the adult myocardium are the long N2BA isoform, which contains the N2A and N2B unique segments, and the short N2B isoform that contains the N2B unique segment only. The two isoforms differ primarily in the I-band region; N2B TTN being shorter due to fewer Ig domains, a much smaller PEVK segment and absence of the N2A unique segment. TTN N2B and N2BA are co-expressed at the level of the half sarcomere (Z-disk to M-band). The force required to stretch a TTN molecule is non-linearly related to its fractional extension¹³⁵. For a given sarcomere length N2B will have greater fractional extension than the longer N2BA and thus N2B is stiffer¹³⁴. The co-expression ratio of TTN within the heart is not fixed but varies within and between the chambers. The atria typically express a greater proportion of N2BA, and transmural gradients exist across the LV wall with higher expression of N2BA toward the endocardium¹³⁶. Titin isoform shifts have been reported in several diseases. Isoform shift toward a higher N2BA/N2B ratio is typically seen in DCM¹³⁷.

1.8 Titinopathies

1.8.1 Dominant titinopathies are adult onset and affect either cardiac or skeletal muscle but not both

Titin was first associated with human disease in 1998 with the assignment of the tibial muscular dystrophy (TMD) locus to chromosome 2q31^{138 139} by linkage analysis in four Finnish kindreds who all shared a common core haplotype. However, the first pathogenic variants in the TTN M-band were not identified for another three years¹⁴⁰. TMD is an autosomal dominant late-onset distal myopathy with weakness and atrophy usually confined to the anterior muscles of the lower leg. The typical symptom of ankle dorsiflexion weakness usually presents in the 4th to 7th decade and walking ability is preserved. The vast majority of cases occur in the Finnish population due to a founder mutation (FINmaj) in Mex6, the last exon in TTN. This complex 11bp variant close to the C-terminus changes four well conserved amino acids in the hydrophobic core of the M10 domain of M-line titin and disrupts calpain-3 binding in the M-band¹⁴⁰. It is noteworthy that Calpain-3 is not expressed in the mature myocardium¹⁴¹. The calpains, or calcium-activated neutral proteases, are nonlysosomal intracellular cysteine proteases. Calpain-3 is a muscle specific form and appears to have a very rapid turnover mediated by autocatalysis, possibly

reflecting the need for precise regulation of its activity. Mutations on CAPN3, the gene encoding Calpain-3, lead to limb girdle muscular dystrophy LGMD2A. Histochemical analysis demonstrated specific loss of the C-terminal TTN epitopes in TMD muscle samples. A French family with TMD was also found to have a single base change in the same Mex6 exon co-segregating with disease^{140 142}. Interestingly one of the first large families used for linkage analysis was consanguineous and a minority of patients in that family had a severe limb-girdle muscular dystrophy (LGMD2J) and were homozygous for the FINmaj variant. Although initially classified as recessive these rare cases are best described as the homozygous manifestation of a dominant variant which manifests in heterozygotes with full penetrance as TMD.

Shortly after the TMD locus was mapped to 2q31, this locus was also reported as the 7th familial DCM locus (CMD1GI) in a large Native American kindred with AD transmission of progressive, early onset DCM without rhythm disturbance¹⁴³. DCM was non-penetrant in two obligate carriers. Although TTN was the obvious candidate gene, only the cardiac-specific N2B unique sequence was examined and no variants were found to segregate with disease. The authors concluded that 'although a mutation in other regions of this massive sarcomere protein have not been excluded, a defect located in titin domains that are expressed in all muscles would most likely cause both skeletal and cardiac dysfunction [and] that the likelihood of a disease causing mutation elsewhere in titin sequences is low'. In 2002 the causative variant in this family was identified as a missense mutation in exon 18 of TTN leading to substitution of tryptophan for arginine in the hydrophobic core of the Ig-Z4 region which, being located in the Z-disc – I-band transition zone, was hypothesised to be a key region for organisation of the Z-disc^{144 145}. In a second large pedigree, the authors also reported a 2bp insertion in TTN exon 326 leading to a stop codon 4 aa downstream of the frameshift co-segregating with DCM without conduction disease following linkage mapping to 2q31.

Itoh-Satoh et al used a candidate gene approach and sequencing limited to Z-disc coding exons of TTN (exons 2-16) and the cardiac specific N2B unique sequence in 120 unrelated Japanese patients with DCM. Seven TTN coding variants were identified, three of which were also seen in healthy controls and deemed benign¹⁴⁶. Two of the DCM variants were found in the Z-disc: a missense variant in exon 14 co-

segregated with arrhythmogenic DCM in one family and a second missense variant in exon 3 was found in a patient with familial DCM, but without affected living relatives for segregation analysis. The Z-disc missense variants were found to reduce the binding of alpha-actinin and T-cap respectively in yeast 2 hybrid assays. Within the N2B sequence two patients were found to have missense variants not seen in controls but familial samples were not available for segregation analysis.

It was not until the advent of high throughput sequencing and target capture technologies that a comprehensive study of *TTN* in human disease was undertaken. In 2012, Herman et al published a report detailing the sequencing of *TTN* in an unprecedented 312 DCM patients, 231 HCM patients and 249 control subjects¹⁷, Filter-based hybridization target capture followed by NGS or Sanger sequencing was used to analyse *TTN* coding exons corresponding to the UniProt Q8WZ42-1 transcript (equivalent to N2BA). Three DCM cohorts (group A, n= 92; group B, n=71; & group C, n=149) enriched for severe and familial DCM, and cases where pathogenic mutation in other DCM genes had been excluded, were studied. Variants predicted to truncate *TTN* were identified in up to 25% of DCM subjects. Interestingly the proportion of subjects with variants predicted to truncate the *TTN* protein was higher in subjects which were studied by NGS (groups A and B, 23% and 20% respectively) than among the subjects in group C studied by Sanger sequencing (8%), leading the authors to postulate that the detection of mutations was better with the NGS strategy than with Sanger sequencing. In this study the frequency of *TTN* truncating variants identified in subjects with HCM and in controls was similar (1% and 3%, respectively; $P=0.34$). In addition, each subject with HCM who had a *TTN* variant also had a pathogenic mutation in an established HCM gene, suggesting that *TTN* truncations are at most a rare cause of HCM. In DCM cases, no significant differences were observed between subjects with and those without *TTN* truncating mutations with respect to the age at diagnosis, left ventricular end-diastolic dimensions, ejection fraction, or rates of cardiac transplantation, implantation of a left ventricular assist device, and death from cardiac causes. However, men with *TTN* mutations had adverse events at significantly earlier ages than did women.

Spurred on by the mounting advances in DCM genetics a combination of exome sequence filtering and linkage analysis was used to prioritise a genetic cause for

DCM in 48 individuals from 17 families with at least three affected members, depleted for pathogenic variants in 16 known DCM genes¹⁴⁷. Missense variants were filtered against ESP exome data (frequency >0.5%, and either a PhastCons score >0.4 or a GERP score of >2). A total of six TTN truncating variants (two frameshift, three nonsense, and one splice site variant that occurred in two families) were identified in seven families (41%). Family B and D also carried rare missense TTN variants that also segregated with disease and although phasing was not explicitly discussed it is likely that the truncating and missense variants are in phase in these families given their co-segregation in affected members. Families A, B, C and D had at least one obligate carrier with a non-penetrant or intermediate phenotype. Two additional families carried TTN truncations that did not segregate with DCM: A frameshift mutation in family 14 was present in only one of three affected members but a PSEN variant previously published with functional data supporting pathogenicity fully segregated with disease in this family; a nonsense variant in family 17 was present in two of three affected family members, but 43 other coding variants were found to segregate fully with disease. The authors surmised that these non-segregating TTN truncations may be 'susceptibility variants', however the presence of another pathogenic variant does not preclude pathogenicity. In this study TTN was the only gene implicated with rare variants in multiple DCM families replicating previous findings that TTN truncations are a frequent cause of DCM. No functional data was presented.

Satoh et al have not limited their studies to DCM. The group also identified a missense variant in exon 14 in the TTN Z-disc that increased binding affinity of alpha-actinin in a patient with HCM¹⁴⁸, and a N2B missense variant co-segregating with HCM in two sisters. Lopes et al¹⁴⁹ took advantage of NGS capacity and the recent availability of sequence data sets for large control cohorts to statistically compare the burden of rare variants in a panel of HCM genes between 180 Caucasian HCM cases and 1,287 Caucasian controls from the UK10K cohort. While this approach cannot speak to the pathogenicity of a given variant, it provides insight into the role of a given gene or genes in disease. The panel comprised 20 genes, including TTN, known to be associated with HCM and DCM and 17 genes associated with other cardiomyopathies and arrhythmia syndromes. 11 patients (6%) carried potential loss of function TTN variants (10 truncations and one previously published missense) of which four were compound with likely pathogenic

HCM gene variants, two were compound with a variant of unknown significance (VUS) in desmosomal or ion channel genes, and five were the sole candidate variant (3%) giving a frequency comparable to previous publications¹⁷. 13% of patients carried rare missense TTN variants in isolation, 10% compound with a desmosomal or ion channel VUS, and 77% compound with a likely pathogenic HCM gene variant, making the frequency of these rare missense TTN variants lower in HCM than in controls. The authors postulate this may be because 'the annotation of titin variants is made extremely complex by the large number of possible isoforms/transcripts which are not accounted for in existing databases'.

In a series of 38 ARVC families depleted for desmosomal gene variants seven families (18%) carried novel missense variants in TTN¹⁵⁰, a frequency below that of controls. One variant, in exon 37, predicted to disrupt a highly conserved Ig-10 domain in the I-band segregated with disease in a large family comprising eight affected individuals. Gel electrophoresis, nuclear magnetic resonance, intrinsic fluorescence and proteolysis assays suggested the variant reduces the structural stability and increases proteolysis of the Ig-10 domain. Full segregation data and functional studies were not reported for the remaining six variants.

Hereditary myopathy with early respiratory failure (HMERF) was described in 7 families from Sweden in 1990 and typically involves adult-onset proximal-distal weakness and respiratory failure in an ambulant patient. Muscle weakness is variable, but MRI shows a pathognomic pattern of muscle involvement. Linkage initially mapped the locus to 2q31. Currently identified families all carry missense mutations in TTN exon 343 encoding the fibronectin-like FN3 119 domain of the A-band. The most frequent mutation occurs on a common haplotype in pedigrees from across Europe suggesting a founder effect. The founder mutation shows a semi-dominant inheritance pattern with subclinical myopathy in the heterozygous parents

¹⁵³.

1.8.2 Recessive titinopathies have diverse phenotypes and typically involve the M-band.

Core-myopathies (CM), the most common form of inherited non-dystrophic childhood muscle disorder, are defined by areas of mitochondria depletion and sarcomere

disorganisation in muscle fibres. They are phenotypically and genetically heterogeneous and may be autosomal dominant or recessive; in the latter case they are known as multi-minicore disease (MmD). Early-onset myopathy with fatal cardiomyopathy (EOMFC), a sub-type of MmD, represents the first titinopathy reported involving both skeletal and cardiac muscle. EOMFC was localised to 2q31 in two consanguineous families each carrying homozygous frameshift variants in the M-band downstream of the kinase domain (exon 358, Mex1 and exon 360, Mex3)¹⁵¹. Skeletal muscle biopsies showed minicore lesions, centrally located nuclei and/or dystrophic lesions. All children presented with congenital muscle weakness and childhood onset DCM with marked rhythm disturbance. Sudden death occurred in four of the five children before age 20, a fifth child underwent heart transplantation at age 15 years. Immunofluorescence data suggested the truncated TTN protein incorporated into the sarcomere despite complete loss of the C-terminal epitopes located downstream of both deletions. All heterozygote parents were asymptomatic.

Most recently, homozygous and compound heterozygous TTN mutations were identified in four families in a cohort of 23 families with congenital MmD of varying severity in conjunction with a range of primary heart disease. Family 1 was consanguineous and both affected siblings were homozygous for a frameshift mutation in Mex2, which leads to deletion of the terminal 522 amino acids. Both siblings presented at two years with moderate muscle weakness. Echocardiography detected mild DCM in one sibling age 19 years but was normal in his younger brother age 16 years. In family two the proband presented at birth with an atrioseptal defect which resolved spontaneously, then developed moderate DCM with rhythm disturbance aged 16 years. Her skeletal phenotype resembled Emery-Dreifuss muscular dystrophy (EDMD) but screening of the LMNA and FHL1 genes associated with EDMD was negative. She was found to carry compound heterozygous nonsense mutations in Mex1. In family three, the proband presented with atrioseptal and ventriculoseptal defects requiring surgical repair and developed end-stage DCM requiring transplantation and rhythm disturbance aged 13. His skeletal muscle phenotype is mild and stable. He was found to carry a nonsense mutation within the TTN kinase domain leading to loss of 100 amino acids from the kinase domain along with 1,963 C-terminal amino acids. He additionally carries a missense variant in exon 316 predicted to impact an Fn3 domain at the C-terminal end of the I-band. At the most severe end of the spectrum, in family four, the

proband presented with distal arthrogryposis multiplex congenita, congenital muscle weakness, kyphosis and neonatal heart failure due to left ventricular non-compaction (LVNC) and a ventriculoseptal defect. He carried a single nucleotide deletion in the splice donor site 5' of exon 38 in the I-band that was demonstrated to lead to use of an alternative splice-donor site (36% of reads) leading to truncation of TTN after the 2950 N-terminal amino acids. A truncated TTN protein of the size predicted was confirmed on gel electrophoresis. In addition he carried a missense mutation in the highly conserved core of the TTN kinase active site. In vitro studies suggest this mutation causes complete loss of function of the TK, thus the patient has no functioning TK, the other allele being truncated proximally. In this small series all heterozygous parents were asymptomatic suggesting truly recessive disease, and the cardiac phenotype seems to correlate with distance of the mutation from the C-terminus such that, within the M-band, the more C-terminal mutations had a milder phenotype.

Centronuclear myopathies (CNM) are a diverse group of clinically and genetically heterogeneous muscle disorders defined by the presence of internalised nuclei in myofibers without excessive regeneration or structural abnormalities. Classic CNM genes encode proteins involved in excitation-contraction coupling at the triad.

Exome sequencing in 29 unrelated CNM patients without mutations in classic CNM genes identified compound heterozygote truncating TTN mutations and in-frame insertions/deletions in five patients (17%)¹⁵². Biochemical analyses demonstrated increased TTN degradation and truncated TTN protein in patients' muscles. All patients presented by the age of three years with muscle weakness. One patient had a mild muscular ventriculoseptal defect, there were no other cardiac abnormalities but the age at last follow up ranged from five to 19 years.

Interestingly, mutations were dispersed throughout the I-band, A-band and M-band, whereas previously reported recessive titinopathies were restricted to the M-band.

Only one patient carried 2 truncating variants, an A-band nonsense mutation (exon275) and an M-band frameshift (exon 312). Four patients carried an inframe insertion/deletion or predicted splice variant as the second variant. Two mutations were previously described: one French TMD mutation inherited from the patients father in whom no abnormality was found on detailed cardiac and muscle examination age 44 years, and one I-band splice variant previously described in DCM, inherited from the probands mother who was asymptomatic but found to have

mild skeletal and cardiac myopathy on examination. Parents carrying the remaining variants were not examined and so it remains to be seen if these are truly recessive mutations or the manifestation of compound heterozygote dominant mutations similar to LGMD2J.

1.8.3 Titin copy number variation

Manolakos et al¹⁵⁴ described a novel microdeletion of 2q31.2 – 2q31.3 encompassing the whole TTN gene and reviewed the literature including four other microdeletions encompassing TTN {Manolakos, 2011 #435;Mitter, 2010 #437;Prontera, 2009 #439;Rifai, 2010 #438}. Patient ages ranged from two to 36 years and common features included premature delivery (3/5 cases), hypotonia (2/5 cases), growth retardation, developmental delay and dysmorphic facies in all cases. Cardiac phenotype was notably absent, although only one adult patient was reported.

1.8.4 Titin animal models

Invertebrates such as drosophila and C.elegans do not possess a single protein equivalent to full length TTN that span the Z-disk, I-band, A-band and M-line regions of the sarcomere. Rather, a TTN-like gene encodes the Z-disk and I-band regions and other genes such as stretchin, twitchin and projectin encode the A-band and M-line regions.

Zebrafish have two TTN orthologues, *ttna* and *ttnb*, located in a tandem array on chromosome 9. Both *ttna* and *ttnb* are highly orthologous to human TTN and encode peptide sequences corresponding to all regions of the sarcomere. As such it is believed that zebrafish represent the simplest model organism that contains TTN molecules equivalent to full length human TTN. The primary site of divergence of *ttna* and *ttnb* from human TTN is in the I-band which is heavily alternatively spliced. *ttna* but not *ttnb* is required for cardiac sarcomere assembly and establishment of cardiac contractility¹⁵⁸. *ttna* is not required for assembly of the Z-bodies or thick filaments during early sarcomereogenesis but is required for the lateral growth of Z-bodies to form Z-disks and the registration of thick filaments to form the A-bands.

The zebrafish is an excellent model for TTN cardiomyopathy. As oxygen is delivered via diffusion rather than the cardiovascular system the zebrafish embryo can tolerate absence of circulation, enabling dissection of the effect of the gene from the secondary effect of hypoxia.

Mutagenesis screening identified TTN as the gene for pickwick, a group of zebrafish mutants wherein the heart develops normally but is poorly contractile from the first beat¹⁵⁹. pickwickm171 (pik^{m171}) is a recessive lethal mutation in the N2B exon of ttna. Three other alleles (pik^{m242}, pik^{m740}, pik^{m186}) share the cardiac phenotype and one (pik^{mVO62H}) has additional skeletal muscle involvement. The pik^{m171} cardiomyocytes are thin but do contract. Transplantation of mutant cells into wildtype hearts confirmed that cellular thinning is cell-autonomous and not secondary to factors such as dysregulation of hypertrophic signals associated with low systolic pressure. Nascent myofibrils form in pik mutants but normal sarcomeres are absent.

There are several murine titinopathy models. Muscular dystrophy with myositis (mdm) is a complex rearrangement comprising an inframe deletion within the N2A sequence, and an inframe insertion of a LINE element¹⁶⁰. Heterozygote Ttn^{mdm} mice are phenotypically normal while homozygotes display progressive muscle degeneration with wasting of proximal and distal muscles. The mutation disrupts a putative Calpain-3 binding site. Since the FINmaj mutation in TMD and LGMD2J disrupts a Calpain-3 binding site, and mutations in CAPN3, the gene encoding Calpain-3, cause limb girdle muscular dystrophy LGMD2A in humans it is proposed that the Ttn^{mdm} mutation may cause disease via disruption of Calpain-3 binding.

The molecular pathogenesis of the FINmaj mutation was explored through production of a mouse model via a knock-in strategy¹⁶¹. The model recapitulates human FINmaj pathology in that in heterozygotes develop dystrophic features limited to specific muscles at a late stage similar to TMD, whereas homozygotes have earlier onset of a more severe phenotype affecting a larger number of muscles similar to LGMD2J. However unlike humans the homozygous mice exhibit a dilated cardiomyopathy with left ventricular dysfunction and fibrosis. Loss of the C-terminal end of the protein in this model is associated with secondary Calpain-3 instability in both heterozygotes and homozygotes. Reduction of the quantity of calpain-3 by

interstrain crossing with a calpain-3 deficient model revealed that the heterozygote FINmaj phenotype was rescued by reduction of calpain-3 but in the homozygote neither the cardiac nor the skeletal phenotype was ameliorated suggesting different pathophysiologies for these two allelic conditions.

Also exploring the effects of M-line mutations, Gotthard et al used a conditional knock-out approach to selectively delete exons MEx1 (encoding the kinase domain) and MEx2 (to maintain the reading frame) in cardiac and skeletal muscle during development ¹⁶². Excision of MEx1–2 from cardiac Ttn during development leads to absence of homozygous offspring, suggesting a critical role for MEx1–2 in heart development. Ultrastructural analyses indicate incorporation of mutant Ttn into sarcomeres and subsequent sarcomere disassembly. Sarcomeres of knock out animals had widened, pale M-lines devoid of M-line bridges and MURF-1. In contrast, the skeletal muscle specific knock out generated live homozygous offspring at the expected Mendelian ratios, affected by early and progressive muscle weakness. Progressive sarcomere disassembly and loss of MURF-1 and myomesin binding was observed but unlike cardiac sarcomeres M-line widening was absent ¹⁶³. To distinguish a role in sarcomere assembly from a role in stabilizing pre-existing sarcomeres and to address potential non-muscle functions, the same authors converted the conditional M-line Ttn knockout into a complete knockout using germline recombination ¹⁶⁴. They demonstrated that initial assembly of the sarcomere does not depend on the M-line region of Ttn nor the phosphorylation of T-cap by the Ttn kinase. Rather, M-line Ttn is required to form a continuous Ttn filament and to provide mechanical stability of the embryonic sarcomere. Even without integration of Ttn into the M band, sarcomeres show proper spacing and alignment of Z discs and M bands but fail to grow laterally, and ultimately disassemble.

Functional effects of the dominant 2bp insertion (c.43628insAT) co-segregating with DCM in a large pedigree following linkage mapping to 2q31 ¹⁴⁴ were explored in a knock-in mouse model ¹⁶⁵. The mutation in Ttn exon 326 leading to a stop codon 4 aa downstream of the frameshift truncates Ttn within the A-band region. Loss of Ttn in homozygous mice lead to early embryonic lethality. Electron microscopy of the myocardium demonstrated severe defects in sarcomere assembly with absence of striations. Heterozygous mice showed normal cardiac structure and function,

however they developed a severe stress-induced cardiomyopathy when exposed to angiotensin II or isoproterenol. Although both Ttn alleles were stably expressed at the mRNA level in heterozygote hearts, truncated Ttn protein was only detected at very low levels (1%) suggesting posttranslational degradation or modification of the mutant protein. Incorporation of the mutant protein into the sarcomere was not confirmed or excluded. Interestingly the authors also observed elevated transcription of the wild type Ttn allele in heterozygotes leading to incomplete compensation of expected full length Ttn protein levels.

1.9 Cardiovascular magnetic resonance (CMR) imaging

CMR creates images of the heart by radiofrequency excitation of hydrogen nuclei in a powerful magnetic field ^{166 167}. Upon relaxation, excited hydrogen nuclei emit energy which is converted into a reconstructed image. The orientation of the image is controlled by varying the main magnetic field using gradient coils. An imaging sequence consists of the combination of radiofrequency excitation pulses and applied gradients used to generate the image. The two main sequences in clinical use are spin echo and gradient echo. Image contrast may be weighted to demonstrate different anatomical structures or pathologies. Contrast between tissues on CMR images relies on the fact that different soft tissue types each return to their equilibrium state after excitation by the independent processes of T1 (spin-lattice) and T2 (spin-spin) relaxation and have distinct T1 and T2 relaxation times.

Cine-CMR is accepted as the gold standard for accurate and reproducible quantification of left and right ventricular volumes, biventricular EF, and LV mass obtained through manual planimetry or the use of semi-automated software. It also provides a powerful means to assess tissue characteristics, viability, perfusion and blood flow.

1.9.1 Standard CMR image acquisition

CMR is performed in a standardised manner. The body planes, orientated orthogonally to the long axis of the body, are used to guide scout images which in turn are used to establish the standard cardiac planes. Axial, coronal and sagittal scout images are acquired first. Axial images run perpendicular to the long axis of the body; coronal images bisect the back and front sides of the body; sagittal images

bisect the left and right sides of the body. Long axis images rotate about the LV longitudinal axis. Short axis images are taken perpendicular to the LV longitudinal axis¹⁶⁸ (Figure 1.2). For cine images MRI sequence is coordinated with retrospective ECG-gating, producing numerous short time frames evenly spaced in the cardiac cycle. These images are spliced together in a cinematic display so that wall motion of the ventricles, valve motion, and blood flow patterns in the heart and great vessels can be visualised.

1.9.2 CMR most accurately measures ventricular morphology and function

Although echocardiography is a useful and practical tool for assessment of DCM, particularly in the clinical setting, it is heavily operator dependent and reliant on good acoustic windows. Furthermore, to produce volumes M-mode or 2-dimensional echo requires geometrical assumptions based on an idealised cardiac shape that may not be present in remodelled ventricles. In patients with advanced heart failure and LV dilatation, the LV frequently adopts a more globular shape due to sphericalisation¹⁰³. 3D echocardiography overcomes some of the geometrical assumptions but still requires good windows, which may not be possible in a patient with markedly dilated ventricles. CMR offers superior delineation of the soft tissues and the blood-myocardium interface with high temporal resolution¹⁶⁹. The multiplanar nature of CMR allows imaging of the entirety of both ventricles with an almost unrestricted field of view and does not require geometrical assumptions, thus excellent accuracy can be maintained¹⁷⁰.

1.9.3 CMR is highly reproducible and increases power to detect a true difference between patient populations

Fast breath-hold CMR shows excellent results for interstudy reproducibility of biventricular volumes, ejection fraction, and LV mass¹⁶⁹.

The sample sizes required to show an effect can be calculated using the following formula:

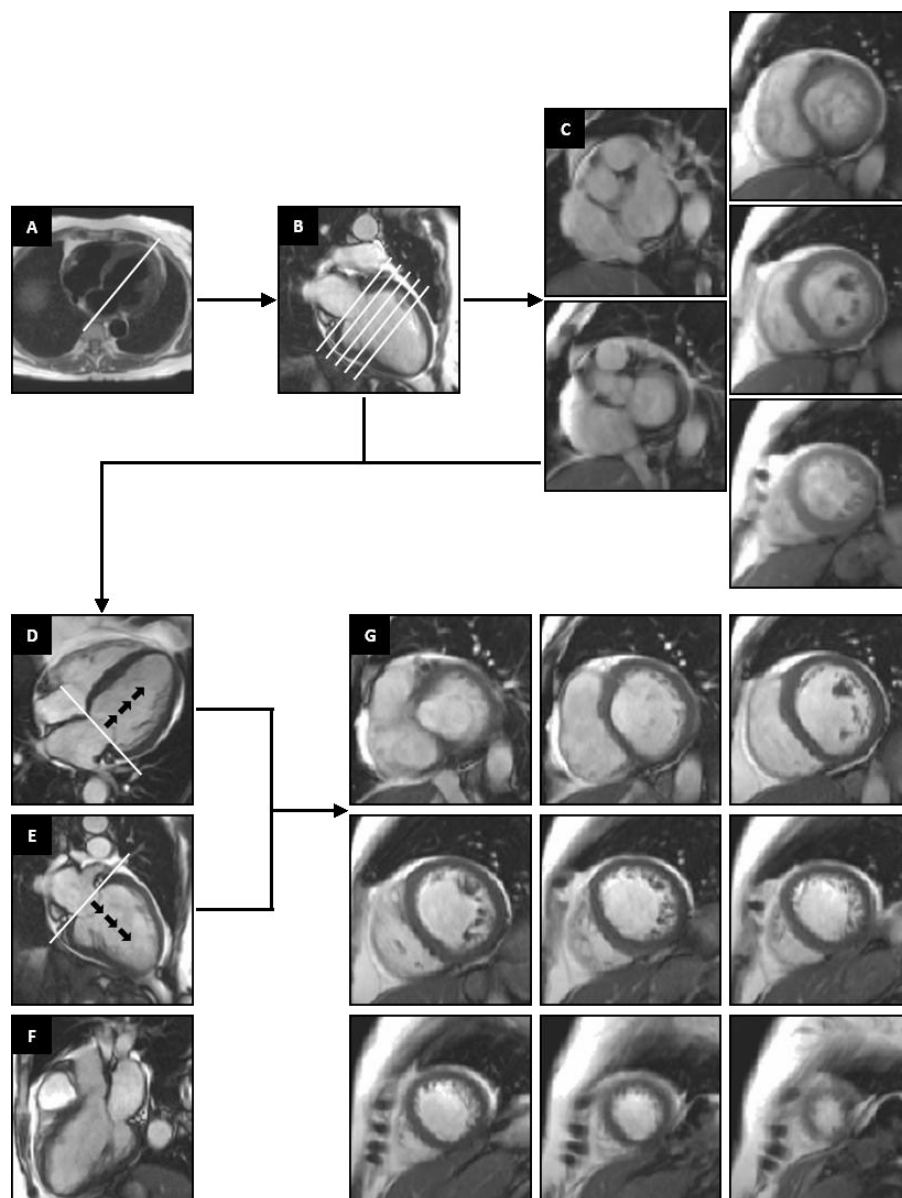
$$n = f(\alpha, P) \sigma^2 / \delta^2$$

where n is the sample size, α the significance level, P the study power required and f the value of the factor for different values of α and P , with σ the interstudy SD and δ the desired difference to be detected.

Thus low SD is important as a reduction in SD leads to a squared reduction in sample size, and superior interstudy reproducibility results in considerably lower calculated sample sizes (reductions of 55% to 93%) required by CMR compared with echocardiography to show clinically relevant changes in LV dimensions and function. For example, to show a 10ml change in LV EDV in patients with heart failure would require 13 patients with CMR versus 66 patients with echocardiography (power of 90%, α error of 0.05) ^{171,172}

Figure 1.2 Acquisition of the cine images and short axis stack.

By using transaxial (Panel A), vertical long-axis (Panel B), and short-axis pilot images (Panel C), 3 points are identified: the left ventricular apex, mid-point of the mitral valve plane, and left ventricular outflow tract. From these, 4-, 2-, and 3-chamber cines are generated (Panels D, E, and F respectively). The 4- and 2-chamber views provide 4 points on the atrio-ventricular ring to plan acquisition of the short-axis stack (Panel G).



1.9.4 Gadolinium-chelated contrast agents aid discrimination of DCM from other causes of heart failure

Gadolinium-chelated contrast agents, which accumulate in areas of interstitial expansion, may be used to document perfusion defects, microvascular ischaemia and areas of fibrosis. Three phases can be assessed after gadolinium injection: the first pass (immediate imaging), early enhancement (first 5 minutes), and late enhancement (5-20 minutes after injection). Late gadolinium enhancement (LGE) detects accumulation of contrast in areas of infarction or fibrosis due to slower contrast kinetics and greater volume of contrast in the extracellular space. The differential patterns of enhancement can discriminate DCM from other causes of heart failure and identify otherwise latent endophenotypes¹⁷³. Approximately one third of DCM patients will have characteristic midwall fibrosis, distinct from that seen in infarction which is typically subendocardial or transmural in an epicardial coronary artery perfusion territory¹⁷⁴. Although x-ray coronary angiography was historically the gold standard for discriminating ischaemic from non-ischaemic disease, there is emerging evidence that CMR with gadolinium enhancement may be more sensitive, particularly given that a significant proportion of patients with a diagnosis of DCM and normal coronary arteries have an ischaemic pattern of LGE indicative of previous undiagnosed infarction¹⁷⁵. Additionally myocarditis in its later stages typically produces a pattern of LGE in the sub-epicardium sparing the sub-endocardium of the lateral wall¹⁷³.

Studies suggest that the presence of mid-wall fibrosis as detected by CMR predicts a worse outcome in terms of all-cause mortality or hospitalisation as well as malignant arrhythmia. The presence of fibrosis appears to be independent of and superior to LVEF in predicting outcome¹⁷⁴.

2 Aims of this Thesis

- 1) To establish a DCM cohort with detailed clinical and CMR imaging data
- 2) To optimise NGS sequencing of TTN as part of a panel of ICC genes
- 3) To describe variation in TTN in DCM patients and compare this to control cohorts
- 4) To determine whether molecular properties of TTNtv can be used to distinguish benign from pathogenic variation
- 5) To determine whether TTNtv are associated with distinct cardiovascular phenotypes and can be correlated with MRI data

3 Materials and methods

3.1 Patient and control cohort selection

All studies were carried out according to institutional guidelines, and with appropriate informed consent from participants. Institutional ethics committees reviewed and approved all protocols (ethics #: 09-H0707-69 and 09/H0504/104). Tissue studies complied with U.K. Human Tissue Act guidelines.

3.1.1 Prospectively recruited unselected DCM subjects

Cohort description:

DCM patients were recruited to the Royal Brompton and Harefield Hospitals NHS Foundation Trust (RBHT) biobank. The biobank prospectively collects, processes and stores specimens, linked to clinical data, consented for donation to research by patients undergoing treatment at the Trust. Patients referred to the RBHT cardiovascular magnetic resonance (CMR) unit from July 2001 to August 2012 for evaluation of a possible diagnosis of DCM and who agreed to provide samples for biobanking were prospectively recruited at the National Institute for Health Research Cardiovascular Biomedical Research Unit, RBHT and Imperial College London (Unselected DCM, n=374). Referrals were from centres across Southern England. A diagnosis of DCM was confirmed, and evaluated against published CMR criteria ($EF > 2sd$ below and end diastolic volume (EDV) $> 2sd$ above the mean normalised for age and sex)^{173 176} by two independent Level 3 accredited CMR cardiologists (Figure 3.1). Where serial scans were available the first diagnostic scan was used. No patients had clinical symptoms or signs of active myocarditis or CMR evidence of infiltrative disease. 40 patients had a history of alcohol excess by UK department of health criteria (14 units/week for women and 21 units/week for men). Significant coronary artery disease (CAD) ($> 50\%$ diameter luminal stenosis in any coronary artery) was excluded in 249 patients by coronary angiography. In 103 patients coronary angiogram was not performed given a low likelihood of myocardial ischemia based on non-invasive testing or the clinical profile of the patient (e.g. a young family member of a patient with known DCM). Bystander CAD, defined as significant CAD evaluated as insufficient to produce DCM phenotype (< 2 affected segments with $< 25\%$ late gadolinium enhancement on CMR) was present in 24 cases.

Clinical data:

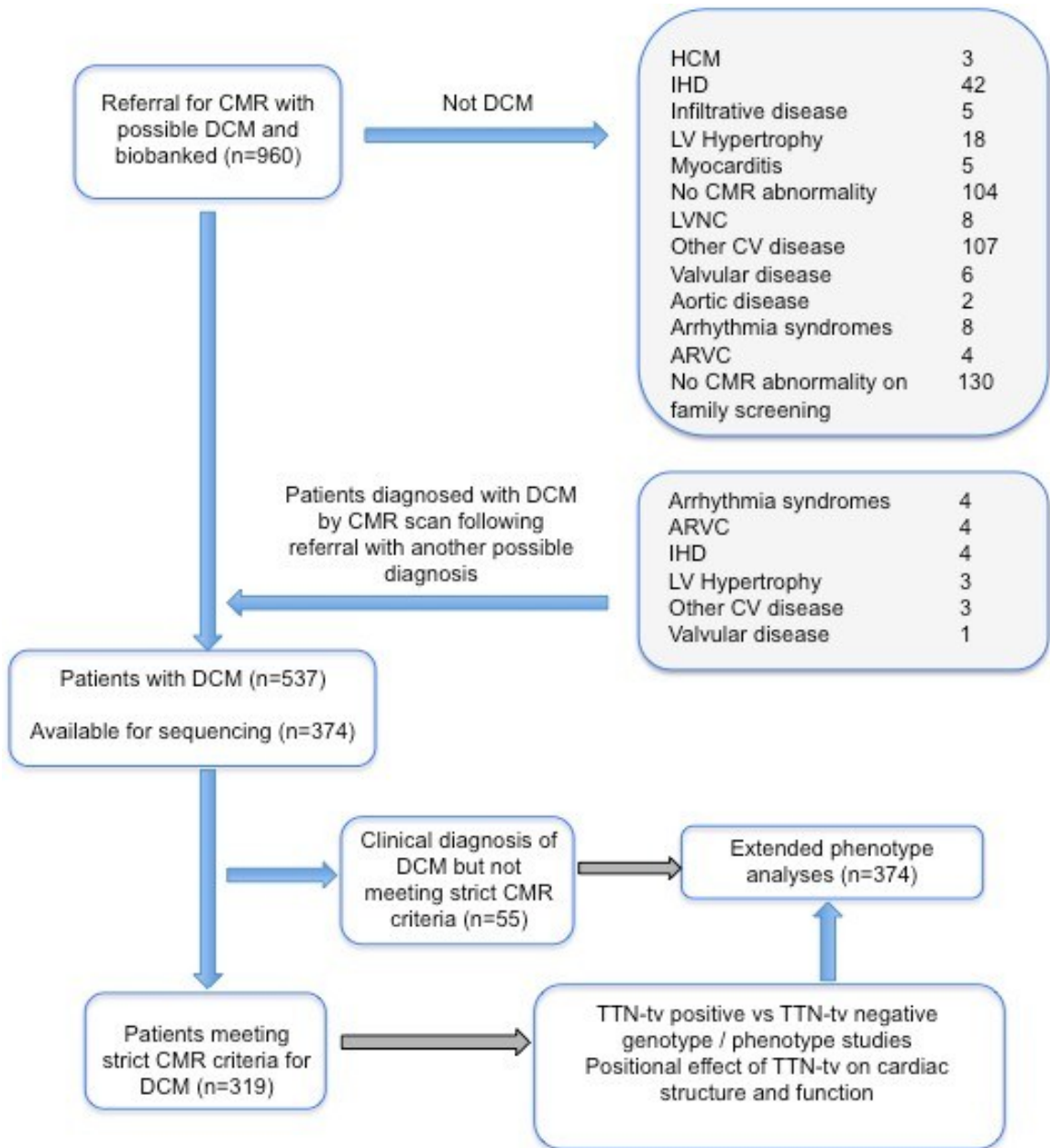
Clinical data were collected for risk factors for non-ischemic DCM including history of alcohol excess, myocarditis, chemotherapy, iron overload and a peripartum presentation. A family history of minimum 3 generations was taken with direct query for DCM, other cardiomyopathy, skeletal myopathy/dystrophy and sudden cardiac death. Significant CAD was excluded as described above. Electrocardiographic data collected included, conduction disease (1st degree heart block: fixed PR interval of >120ms; 2nd degree heart block: Mobitz type 1 or type 2; left bundle branch block (LBBB): QRS duration >120ms with LBBB pattern), atrial fibrillation, sustained and non-sustained ventricular tachycardia (VT) and ventricular fibrillation. Non-sustained VT was defined as 3 or more consecutive ventricular extra-systoles at a rate of >120bpm, sustained VT was defined as >30 seconds consecutive beats of VT. Patient clinical data were obtained by questionnaire at time of enrolment and from clinical records.

3.1.2 End-stage DCM subjects**Cohort description:**

All end-stage DCM patients who were referred for cardiac transplantation and/or left ventricular device implantation between 1993 and 2011 at the Royal Brompton and Harefield Hospital NHS Trust were prospectively enrolled in a tissue bank. 155 subjects for whom peripheral blood was available for DNA analysis were included in this study. This cohort included 84 cases from which frozen LV tissue was available for tissue studies. For the end-stage cohort, diagnosis of non-ischaemic DCM was confirmed from patient's medical records, however detailed clinical data was not collected. Phenotype analysis was not performed on the end-stage cohort.

Figure 3.1 Schematic representation of study design

Patients with a possible diagnosis of DCM (n=960) and additional patients (n=19) originally referred with a possible alternate diagnosis were recruited to the biobank. We excluded 442 patients referred with a possible diagnosis of DCM as not DCM based on CMR characteristics. Of the 537 patients with DCM 374 had been sequenced at the time of preparing this thesis. Of the 374 included here 319 were confirmed with DCM by strict CMR characteristics^{173,176}. A further 55 were diagnosed as DCM based on clinical history and supportive CMR characteristics: borderline LVEDVi and LVEF (EDVi>1.5SD above & EF>1.5SD below the mean for age and sex) were considered supportive CMR characteristics in the presence of corroborating clinical and/or echocardiographic evidence. Abbreviations: HCM, hypertrophic cardiomyopathy; IHD, ischemic heart disease; LVNC, LV non compaction; ARVC, arrhythmogenic right ventricular cardiomyopathy; CV, cardiovascular disease; TTNtv, TTN truncating variant.



3.1.3 Healthy volunteers

Cohort description:

Adult volunteers (age range 18-72 years, mean 40.3 years) were prospectively recruited via advertisement at the MRC Clinical Sciences Centre, Imperial College London. Participants with previously documented cardiovascular disease, hypertension, diabetes or hypercholesterolemia were excluded. Female subjects were excluded if they were pregnant or breastfeeding but were eligible if they took oral contraceptives. Standard published safety contraindications to magnetic resonance imaging were applied with a scanner weight limit of 100kg. Normal cardiac morphology and function was confirmed in all healthy volunteers by CMR. Scans were reviewed by a Level 3 accredited CMR cardiologist and phenotype data was used as a baseline against which to compare the morphology of the DCM cohort.

3.1.4 Population Cohorts

The Framingham Heart Study (FHS) is a multi-generation, prospective, population-based study aimed at identifying the causes of cardiovascular disease¹⁷⁷. In 1948, 5,209 men and women aged between 30 to 62, living in the town of Framingham, Massachusetts were enrolled. Between 1971 and 1975, the study enrolled a second generation, the Offspring Cohort, which consisted of 5,124 children of the original cohort and their spouses. The Offspring cohort has since been examined every three to eight years, with the last exam (exam 8), occurring between 2005 and 2008. All FHS phenotypic data was retrieved from NCBI dbGaP (Accession: phs000007.v18.p7). A random subset of 1,623 unrelated individuals in the FHS Offspring Cohort was sequenced as part of this study. Sequence data is available from NCBI dbGaP (Accession: phs000307.v3.p7).

The Jackson Heart Study (JHS)¹⁷⁸ is a population-based, prospective study of cardiovascular disease. Between 2000 and 2003, the study enrolled 5,301 African-Americans aged 35 to 84 living in the Jackson, Mississippi metropolitan area. All JHS phenotypic data was downloaded from NCBI dbGaP (Accession: phs000286.v3.p1). A random subset of 1,980 unrelated subjects were sequenced as part of this study. The JHS sequence data is available from NCBI dbGaP (Accession: phs000498.v1.p1).

Details of TTNtv were provided for the population cohorts and these together with TTNtv in the healthy volunteer cohort were compared against TTNtv in DCM for frequency, expression and distribution.

3.2 CMR phenotyping

3.2.1 CMR of prospectively enrolled DCM subjects:

Cardiovascular magnetic resonance was performed on Siemens Sonata 1.5-T or Avanto 1.5-T (Erlangen, Germany). Cine images were acquired using a steady-state free-precession sequence in standard 2-, 3- and 4-chamber long-axis views (TE [echo time]/TR [repetition time] 1.6/3.2 ms, flip angle 60°), with subsequent sequential 8-mm short-axis slices (2-mm gap) from the atrioventricular ring to the apex. Ventricular volumes and function were measured for both ventricles using standard techniques¹⁷². Left atrial (LA) area and length were recorded manually from 2-chamber and 4-chamber images. The biplane area-length method was used to calculate LA volume (LAV)¹⁷⁹. In addition to maximum wall thickness measured in the short axis, mean wall thickness was measured at mid-ventricular (papillary muscle) level in the septum and lateral wall. LAV, LV and right ventricular (RV) volumes, wall thicknesses and LV mass were indexed to body surface area (BSA). Image analysis was performed using semi-automated software (CMRtools, Cardiovascular Imaging Solutions, London, United Kingdom). Late gadolinium enhancement (LGE) images were acquired 10 min after intravenous gadolinium-DTPA (Schering, Berlin, Germany; 0.1 mmol/kg) in identical short-axis planes using an inversion-recovery gradient echo sequence, as previously described¹⁸⁰. Midwall LGE was only deemed to be present when the area of signal enhancement could be seen in both phase-swapped images and in a cross-cut long-axis image, as previously described¹⁸⁰.

3.2.2 CMR of prospectively enrolled healthy volunteers:

Healthy volunteer CMR studies were performed on a 1.5T Philips Achieva system (Best, Netherlands). Cine images were acquired using a balanced-steady-state free-precession sequence in standard 2-, 3- and 4-chamber long-axis views (TE/TR 1.5/3.0 ms, flip angle 60°), with subsequent sequential 8-mm short-axis slices (2-mm gap) from the atrioventricular ring to the apex. Gadolinium enhancement was not

performed. Ventricular volumes were measured and indexed as above for the DCM cohort. Anteroseptal and inferoseptal wall thickness were measured and thus was not used for comparison against the DCM cohort. LAV was not measured.

3.3 DNA preparation and quality control

DCM samples from prospectively recruited cases were prepared at the BRU genetics laboratory that is equipped with an automated DNA extraction platform (EZ1 Advanced XL, Qiagen). DNA samples were extracted from whole blood (2 x 350µl) as per the manufacturer's instructions (EZ1 DSP DNA Blood Kit).

Healthy volunteer samples were prepared by the Hammersmith Hospital Haematology Laboratory that is equipped with an automated DNA extraction platform (QIAcube, Qiagen). DNA samples were extracted from whole blood as per the manufacturer's instructions.

DNA samples from the end-stage cohort were obtained from the Royal Brompton and Harefield Transplant Unit Tissue Typing laboratory archive.

DNA quantity and quality was assessed by nanodrop (Thermo Scientific) and samples aliquoted into stock solutions and diluted preparations in 96 well plates and stored at -80C. A 260/280 ratio of >1.7 was accepted for these samples.

3.4 Target enrichment of TTN

3.4.1 Assay design

SureSelect (Agilent technologies) custom hybridisation capture probes were used to capture 204 genes implicated in cardiovascular disease, including *TTN*. RNA baits were designed using Agilent's eArray platform (<https://earray.chem.agilent.com/earray/>). Other than increased bait tiling across the target (five-fold), standard eArray parameters were used. Baits targeted all exons of all Ensembl *TTN* transcripts downloaded from Ensembl version 54, including UTRs, a 100-bp extension into adjacent introns to ensure efficient target capture and downstream sequencing¹⁸¹, and 1.25 kb of upstream sequence to capture putative variation in key regulatory, non-coding motifs. A total of 6340 unique 120-mer RNA

baits were generated for TTN, covering a target region of 168369 bp, including 112916 protein-coding bases.

Methods for the optimisation of the final assay design are presented in chapter 5 (Design and optimisation of the Titin sequencing assay), as the optimisation strategy was heavily influenced by the data presented there.

3.4.2 DNA library preparation

DNA was extracted using standard automated approaches as above and quality and quantity was re-assessed by agarose gel electrophoresis and/or fluorometry (Qubit, Life Technologies). 3 µg of gDNA was diluted in 120ul of low TE and sheared using the Covaris E System. Libraries were constructed with the SureSelect SOLiD XT (Agilent) kits using the Bravo Automated Liquid Handling Platform (Agilent) as per the manufacturer's protocol. Libraries were visualised before and after hybridisation on the Bioanalyser (Agilent) using the HS DNA Assay as per manufacturer's protocol. Libraries were hybridised with custom designed SureSelect capture libraries at 60°C for 24 hours, captured with streptavidin coated magnetic beads (Invitrogen Dynabeads) and amplified with 12 cycles of PCR with indexing primers. Libraries were quantified using qPCR and pooled in equimolar amounts before template bead preparation using 500pg of input library, in an automated ePCR EZBead station. Template beads were enriched using standard protocols and loaded onto a sequencing slide. The SOLiD Experiment Tracking System software was used for run analysis reports.

3.4.3 Next generation sequencing platforms

SOLiD v3, v3+ and SOLiDv4 platforms were used for pilot studies; the SOLiD 5500XL platform (Life Technologies) was used for NGS experiments. The SOLiDv3 generates up to 25Gb of sequence data from a single run using single reads. The SOLiD v3+ generates up to 60Gb and the SOLiD v4 up to 100Gb of sequence data from a single run using paired end (50bp forward, 35bp reverse) technology. The SOLiD 5500 generates up to 150Gb of sequence data (75bp forwards, 35bp reverse).

3.5 Whole genome sequencing

TTN was sequenced using whole genome sequencing in 54 subjects from the end-stage DCM cohort. Sequencing was performed by Complete Genomics, and variants were called using Complete Genomic Analysis Tools

(<http://www.completegenomics.com/analysis-tools/cgatools/> version: 2.2.0.26).

Variants were filtered out if they met any of the following conditions; i) low confidence or incomplete calls flagged by the caller, ii) read-depth less than 10, iii) allele frequency less than 0.15 for heterozygous calls.

3.6 LV tissue studies

Tissue studies were performed on frozen LV tissue samples from 84 of our end-stage DCM cohort. RNA sequencing was performed on all 84 samples, and protein studies in a subset of these. Control tissue used for protein studies were from 6 unused donor hearts with no known cardiac disease.

3.7 RNA sequencing

Total RNA was extracted from frozen LV samples from 84 end-stage DCM cases using Trizol (Life Technologies) following the manufacturer's protocol and quantified by UV spectrophotometry. RNA quality was measured on the Agilent 2100 Bioanalyser using Agilent's RNA 6000 reagents and 250 µg of sample. RINs ranged between 6.3 and 8.7 with a mean of 7.6. 4µg of total RNA was used for library preparation with the TruSeq RNA Sample Preparation Kit (Illumina). Barcoded cDNA fragments of poly(A)+ RNA were then sequenced on a HiSeq 2000 instrument from Illumina with 2 x 100 bp PE chemistry. Reads were mapped against the reference genome GRCh37 using TopHat 1.4.1¹⁸² supplied with the gene annotation of the Ensembl database¹⁸³. Point mutations in the RNA-seq data were detected using the SAMtools suite¹⁸⁴ considering only uniquely mapping reads with no more than 2 mismatches in 100 bp. All detected truncating mutations were checked by Sanger sequencing.

These data were used for TTN variant analyses in conjunction with whole genome and targeted sequencing data. For each individual, variants observed with the same zygosity in RNA-seq and whole genome data were considered validated and used to assess allelic balance across the gene and transcript specific degradation.

Thirteen TTN truncations were identified in genomic DNA from the 84 samples used in these studies. The allele balance of 10 truncating variants could be interrogated in both DNA & RNA. There were no discordant zygosity calls. Two variants affected canonical splice sites, and so were not directly represented in the RNA sequence, and one further variant was not called as indels >2bp are not interrogated by the variant calling software, but all 3 were present in the whole genome or targeted sequencing data and confirmed by Sanger sequencing.

3.7.1 Exon expression estimation

To estimate exon usage, reads covering exons (inclusion reads) and reads spanning exons (exclusion reads) were counted, normalised for exon length and the proportion of transcripts incorporating each exon was calculated (proportion spliced in (PSI) score¹⁸³). Estimated exon usage was derived from the RNA-sequencing of 84 LV tissue samples; these values were then applied across all cohorts to give the estimated expression level of each exon containing a TTNtv.

3.8 Sanger sequencing

All truncating and predicted splice variants identified in NGS and RNA-seq data were Sanger sequenced. Approximately half of these variants were validated. Design of the variant calling strategy and Sanger validation results are described in detail in chapter 5 (Design and optimisation of the Titin sequencing assay).

3.9 Bioinformatics

3.9.1 Bioinformatic analysis of targeted resequencing

SOLiD 5500xl paired-end reads were demultiplexed and aligned to human reference genome (hg19) in colour space using LifeScope™ v2.5.1 “targeted.reseq.pe” pipeline. SOLiD Accuracy Enhancement Tool (SAET) was used to improve colour call accuracy prior to mapping. SAET implements a spectral alignment error correction algorithm that uses quality values and properties of colour space. All other LifeScope parameters were used as default. Duplicate reads and those mapping with a quality score <8 were removed. Variant calling was performed using diBayes (SNPs) and small.indel modules, as well as GATK v1.5-20¹⁸⁵ and SAMtools v0.1.18. Variants called by any of these methods were taken forward for Sanger

validation. Alignment metrics and target capture analysis were carried out by Picard v1.40, BedTools v2.12 and in-house Perl scripts. GATK CallableLociWalker module was used to classify the target genomic regions callable by minimum depth >4 with base quality >20 and mapping quality >10.

3.9.2 Variant annotation

To facilitate standardised variant annotation in accordance with international guidelines we developed a Locus Reference Genomic sequence (LRG)¹⁸⁶ for *TTN* (<http://www.lrg-sequence.org>, LRG_391). Variants are described relative to an inferred complete meta-transcript (LRG_391_t1) manually curated by the HAVANA group that incorporates all *TTN* exons, with the exception of a single alternative terminal exon unique to the shorter Novex-3 isoform (Figure 3.2 and Tables 3.1 & 3.2). Variants in the Novex-3 terminal exon are reported relative to LRG_391_t2.

Variants were reported using Human Genome Variation Society nomenclature. The functional consequences of variants were predicted using the Ensembl Perl API¹⁸⁷ Variant Effect Predictor¹⁸⁸. Variants were classified as truncating if their consequence included one of following sequence ontology terms: “transcript_ablation”, “splice_donor_variant”, “splice_acceptor_variant”, “stop_gained”, “stop_lost” or “frameshift_variant”.

Figure 3.2 A unified representation of TTN transcripts.

A cartoon shows exon usage in the principal TTN cardiac transcripts (exon boundaries denoted by black lines). The meta-transcript (LRG_t1 / ENST00000589042) is a manually curated inferred complete transcript, incorporating exons of all known TTN isoforms (including fetal and non-cardiac isoforms) with the exception of the alternative terminal exon found only in the Novex-3 transcript (LRG_t2 / ENST00000360870, shown in green). The principal adult cardiac long & short isoforms are known as N2BA and N2B: exon usage for reference versions of these isoforms (ENST00000591111, ENST00000460472) is shown, though exon usage *in vivo* is more variable. Novex-1 and Novex-2 are minor cardiac isoforms that differ from N2B only in the inclusion of Novex-1 and Novex-2 (ENST00000359218, ENST00000342175) unique exons (shown in the red and blue respectively within the N2B track) positioned within the constitutively expressed position of the early I-band and notably also absent from the N2BA isoform. The lowest plot shows exon usage in human DCM (LV) as the "proportion spliced in" for each exon - that is, an estimate of the proportion of transcripts that incorporate each exon, derived from RNA sequencing of LV tissue from 84 patients with DCM. PSI for the alternate terminal exon of Novex-3 can not be accurately estimated: estimates are spuriously overinflated due to poly-A enrichment of RNA during library preparation (all other transcripts share the same terminal exon)

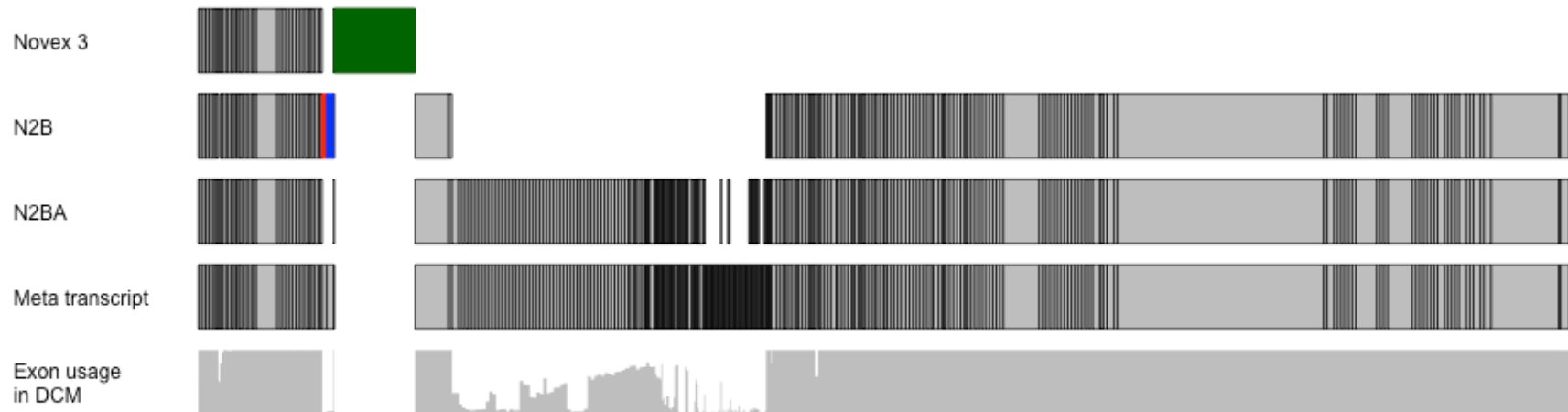


Table 3.1 Reference transcript and protein identifiers for TTN

Isoform	Description	LRG	Ensembl transcript	Ensembl protein	RefSeq transcript	RefSeq protein	UniProt	Length (aa)
meta-transcript	Inferred complete meta-transcript	391_t1	ENST00000589042	ENSP00000467141	NM_001267550.1	NP_001254479		35991
N2-BA	Principle cardiac long isoform		ENST00000591111	ENSP00000465570	NM_001256850.1	NP_001243779	Q8WZ42-1	34350
N2-A	Soleus / skeletal long isoform		ENST00000342992	ENSP00000343764	NM_133378.4	NP_596869.4	Q8WZ42-4	33423
N2-B	Principle cardiac short isoform		ENST00000460472	ENSP00000434586	NM_003319.4	NP_003310.4	Q8WZ42-3	26926
Novex-1	Minor cardiac short isoform		ENST00000359218	ENSP00000352154	NM_133432.3	NP_597676.3	Q8WZ42-8	27051
Novex-2	Minor cardiac short isoform		ENST00000342175	ENSP00000340554	NM_133437.3	NP_597681.3	Q8WZ42-7	27118
Novex-3	Minor small cardiac isoform	391_t2	ENST00000360870	ENSP00000354117	NM_133379.3	NP_596870.2	Q8WZ42-6	5604

Table 3.2 Overview of TTN transcripts and exon usage

For each of 364 TTN exons, Hg19 and LRG genomic co-ordinates are shown, along with the coordinates of the amino-acid residues encoded by that exon (LRG meta-transcript protein coordinates). The recommended exon numbering is shown alongside the historical numbering system in use in much of the literature.

The exon rank in published cardiac transcripts is shown next for seven important transcripts, followed by a description of the exon and its relationship to important protein features.

Finally, an estimate of exon usage in the heart is shown. PSI indicates “proportion spliced in” – the proportion of expressed transcripts that contain the exon listed

Recommended exon number (HAVANA)	Historical exon number (GenBank)	Hg19 genomic start position	Hg19 Genomic end position	LRG genomic start position	LRG genomic end position	Meta-transcript Protein start position	Meta-transcript Protein end position	Meta-transcript exon number	N2BA exon number	N2A exon number	N2B exon number	Novex-1 exon number	Novex-2 exon number	Novex-3 exon number	Exon status	Protein Region	Protein Domain	Average exon PSI
1	1	179672150	179671939	28380	28591	-	-	1	1	1	1	1	1	1	5' utr			100.00%
2	2	179669382	179669279	31148	31251	1	31	2	2	2	2	2	2	2		Z-disk	Ig-like 1	100.00%
3	3	179667068	179666865	33462	33665	31	99	3	3	3	3	3	3	3		Z-disk	Ig-like 1	100.00%
4	4	179665409	179665122	35121	35408	99	195	4	4	4	4	4	4	4		Z-disk	Ig-like 2	100.00%
5	5	179664637	179664552	35893	35978	195	223	5	5	5	5	5	5	5		Z-disk		100.00%
6	6	179664458	179664214	36072	36316	224	305	6	6	6	6	6	6	6		Z-disk		100.00%
7	7	179659979	179659649	40551	40881	305	415	7	7	7	7	7	7	7		Z-disk		100.00%
8	8	179659278	179659126	41252	41404	416	466	8	8	8	8	8	8	8		Z-disk	Z-repeat 1	100.00%
9	9	179658268	179658131	42262	42399	467	512	9	9	9	9	9	9	9		Z-disk	Z-repeat 2	100.00%
10	10	179656924	179656799	43606	43731	513	554	10	10	10	10	10	10	10		Z-disk	Z-repeat 3	100.00%
11	11	179655572	179655435	44958	45095	555	600	11	11	11				11		Z-disk	Z-repeat 4	57.32%
12	12	179654842	179654705	45688	45825	601	646	12	12	12	11	11	11	12		Z-disk	Z-repeat 5	79.99%
13	13	179654224	179654087	46306	46443	647	692	13	13	13	12	12	12	13		Z-disk	Z-repeat 6	98.26%
14	14	179650868	179650575	49662	49955	693	790	14	14	14	13	13	13	14		Z-disk	Z-repeat 7	100.00%
15	15	179650469	179650347	50061	50183	791	831	15	15	15	14	14	14	15		Z-disk		100.00%
16	16	179649078	179648797	51452	51733	832	925	16	16	16	15	15	15	16		near Z-disk		100.00%
17	17	179648512	179648447	52018	52083	926	947	17	17	17	16	16	16	17		near Z-disk		100.00%
18	18	179647791	179647533	52739	52997	948	1034	18	18	18	17	17	17	18		near Z-disk	Ig-like 3	100.00%
19	19	179647329	179647266	53201	53264	1034	1055	19	19	19	18	18	18	19		near Z-disk		100.00%
20	20	179647154	179646939	53376	53591	1055	1127	20	20	20	19	19	19	20		near Z-disk	Ig-like 4	100.00%
21	21	179645990	179645848	54540	54682	1127	1175	21	21	21	20	20	20	21		near Z-disk	Ig-like 4	100.00%

22	22	179644932	179644727	55598	55803	1175	1243	22	22	22	21	21	21	22		near Z-disk		100.00%
23	23	179644189	179643956	56341	56574	1244	1321	23	23	23	22	22	22	23		near Z-disk	Ig-like 5	100.00%
24	24	179643845	179643601	56685	56929	1322	1403	24	24	24	23	23	23	24		near Z-disk	Ig-like 5	100.00%
25	25	179642702	179642431	57828	58099	1403	1494	25	25	25	24	24	24	25		near Z-disk	Ig-like 6	100.00%
26	26	179642311	179642147	58219	58383	1494	1549	26	26	26	25	25	25	26		near Z-disk	Ig-like 6	100.00%
27	27	179642044	179641876	58486	58654	1549	1605	27	27	27	26	26	26	27		near Z-disk	Ig-like 7	100.00%
28	28	179641776	179640083	58754	60447	1605	2170	28	28	28	27	27	27	28		near Z-disk / I-band	Ig-like 7/8/9/10	100.00%
29	29	179639929	179639648	60601	60882	2170	2264	29	29	29	28	28	28	29		I-band	Ig-like 11	100.00%
30	30	179639200	179638934	61330	61596	2264	2353	30	30	30	29	29	29	30		I-band	Ig-like 12	100.00%
31	31	179638837	179638565	61693	61965	2353	2444	31	31	31	30	30	30	31		I-band	Ig-like 13	100.00%
32	32	179638452	179638189	62078	62341	2444	2532	32	32	32	31	31	31	32		I-band	Ig-like 14	100.00%
33	33	179638096	179637836	62434	62694	2532	2619	33	33	33	32	32	32	33		I-band		100.00%
34	34	179636198	179635938	64332	64592	2619	2706	34	34	34	33	33	33	34		I-band	Ig-like 15	100.00%
35	35	179635402	179635139	65128	65391	2706	2794	35	35	35	34	34	34	35		I-band		100.00%
36	36	179635047	179634787	65483	65743	2794	2881	36	36	36	35	35	35	36		I-band		100.00%
37	37	179634666	179634406	65864	66124	2881	2968	37	37	37	36	36	36	37		I-band	Ig-like 16	100.00%
38	38	179633660	179633400	66870	67130	2968	3055	38	38	38	37	37	37	38		I-band	Ig-like 17	100.00%
39	39	179632882	179632741	67648	67789	3055	3102	39	39	39	38	38	38	39		I-band	Ig-like 18	100.00%
40	40	179632651	179632486	67879	68044	3102	3157	40	40	40	39	39	39	40		I-band	Ig-like 18	100.00%
41	41	179631339	179631108	69191	69422	3158	3235	41	41	41	40	40	40	41		I-band		100.00%
42	42	179629538	179629254	70992	71276	3235	3330	42	42	42	41	41	41	42		I-band	Ig-like 19	100.00%
43	43	179629029	179628904	71501	71626	3330	3372	43	43	43	42	42	42	43		I-band	Ig-like 20	100.00%
44	44	179623899	179623711	76631	76819	3372	3435	44	44	44	43	43	43	44		I-band	Ig-like 20	100.00%
45	45	179622643	179622269	77887	78261	3435	3560	45				44			Novex-1			1.55%
46	46	179621524	179620949	79006	79581	3560	3752	46					44		Novex-2			6.91%
47	47	179617907	179617851	82623	82679	3752	3771	47	45	45	44	45	45	45		I-band		100.00%
48	48	179616766	179610047											46	Novex-3			83.24%
49	49	179606648	179603868	93882	96662	3771	4698	48	46		45	46	46		N2B unique	I-band	Ig-like 21/22/23	98.72%

															sequence			
50	50	179603087	179602809	97443	97721	4698	4791	49	47	46	46	47	47			I-band	Ig-like 24	100.00%
51	51	179600801	179600238	99729	100292	4791	4979	50	48	47						I-band	Ig-like 25/26	39.13%
52	52	179599715	179599434	100815	101096	4979	5073	51	49	48						I-band	Ig-like 27	21.50%
53	53	179599333	179599055	101197	101475	5073	5166	52	50	49						I-band	Ig-like 28	14.70%
54	54	179598619	179598341	101911	102189	5166	5259	53	51	50						I-band	Ig-like 29	11.82%
55	55	179598244	179597966	102286	102564	5259	5352	54	52	51						I-band	Ig-like 30	10.11%
56	56	179597848	179597561	102682	102969	5352	5448	55	53	52						I-band	Ig-like 31	11.57%
57	57	179597445	179597167	103085	103363	5448	5541	56	54	53						I-band	Ig-like 32	11.44%
58	58	179597074	179596793	103456	103737	5541	5635	57	55	54						I-band	Ig-like 33	14.53%
59	59	179596698	179596420	103832	104110	5635	5728	58	56	55						I-band	Ig-like 34	14.99%
60	60	179596310	179596032	104220	104498	5728	5821	59	57	56						I-band	Ig-like 35	32.69%
61	61	179595930	179595652	104600	104878	5821	5914	60	58	57						I-band	Ig-like 36	46.73%
62	62	179595519	179595232	105011	105298	5914	6010	61	59	58						I-band	Ig-like 37	38.86%
63	63	179595098	179594820	105432	105710	6010	6103	62	60	59						I-band	Ig-like 38	12.33%
64	64	179594672	179594391	105858	106139	6103	6197	63	61	60						I-band	Ig-like 39	14.22%
65	65	179594293	179594015	106237	106515	6197	6290	64	62	61						I-band	Ig-like 40	13.38%
66	66	179593896	179593618	106634	106912	6290	6383	65	63	62						I-band	Ig-like 41	8.06%
67	67	179593505	179593227	107025	107303	6383	6476	66	64	63						I-band	Ig-like 42	11.29%
68	68	179593124	179592837	107406	107693	6476	6572	67	65	64						I-band	Ig-like 43	9.98%
69	69	179592590	179592312	107940	108218	6572	6665	68	66	65						I-band	Ig-like 44	12.10%
70	70	179592098	179591817	108432	108713	6665	6759	69	67	66						I-band	Ig-like 45	65.99%
71	71	179590773	179590495	109757	110035	6759	6852	70	68	67						I-band	Ig-like 46	53.72%
72	72	179590376	179590095	110154	110435	6852	6946	71	69	68						I-band	Ig-like 47	52.78%
73	73	179589265	179588987	111265	111543	6946	7039	72	70	69						I-band	Ig-like 48	35.48%
74	74	179588870	179588583	111660	111947	7039	7135	73	71	70						I-band	Ig-like 49	36.44%
75	75	179588423	179588145	112107	112385	7135	7228	74	72	71						I-band	Ig-like 50	40.79%
76	76	179588051	179587773	112479	112757	7228	7321	75	73	72						I-band	Ig-like 51	38.93%
77	77	179587664	179587386	112866	113144	7321	7414	76	74	73						I-band	Ig-like 52	56.34%
78	78	179587273	179586986	113257	113544	7414	7510	77	75	74						I-band	Ig-like 53	40.31%
79	79	179586861	179586574	113669	113956	7510	7606	78	76	75						I-band	Ig-like 54	46.83%
80	80	179585929	179585648	114601	114882	7606	7700	79	77	76						I-band	Ig-like 55	49.39%
81	81	179585390	179585112	115140	115418	7700	7793	80	78	77						I-band	Ig-like 56	49.99%
82	82	179584991	179584710	115539	115820	7793	7887	81	79	78						I-band	Ig-like 57	53.78%
83	83	179584559	179584281	115971	116249	7887	7980	82	80	79						I-band	Ig-like 58	54.55%
84	84	179584178	179583891	116352	116639	7980	8076	83	81	80						I-band	Ig-like 59	10.19%

85	85	179583700	179583422	116830	117108	8076	8169	84	82	81						I-band	Ig-like 60	7.78%
86	86	179583327	179583049	117203	117481	8169	8262	85	83	82						I-band	Ig-like 61	6.45%
87	87	179582948	179582670	117582	117860	8262	8355	86	84	83						I-band	Ig-like 62	6.13%
88	88	179582537	179582250	117993	118280	8355	8451	87	85	84						I-band	Ig-like 63	6.05%
89	89	179582109	179581822	118421	118708	8451	8547	88	86	85						I-band	Ig-like 64	10.65%
90	90	179580501	179580220	120029	120310	8547	8641	89	87	86						I-band	Ig-like 65	58.46%
91	91	179579991	179579713	120539	120817	8641	8734	90	88	87						I-band	Ig-like 66	60.89%
92	92	179579300	179579019	121230	121511	8734	8828	91	89	88						I-band	Ig-like 67	64.10%
93	93	179578902	179578624	121628	121906	8828	8921	92	90	89						I-band	Ig-like 68	67.76%
94	94	179578099	179577812	122431	122718	8921	9017	93	91	90						I-band	Ig-like 69	68.15%
95	95	179577702	179577424	122828	123106	9017	9110	94	92	91						I-band	Ig-like 70	68.04%
96	96	179577320	179577042	123210	123488	9110	9203	95	93	92						I-band	Ig-like 71	69.89%
97	97	179576949	179576671	123581	123859	9203	9296	96	94	93						I-band	Ig-like 72	69.51%
98	98	179576076	179575789	124454	124741	9296	9392	97	95	94						I-band	Ig-like 73	69.66%
99	99	179575649	179575362	124881	125168	9392	9488	98	96	95						I-band	Ig-like 74	73.85%
100	100	179574583	179574293	125947	126237	9488	9585	99	97	96						I-band	Ig-like 75	76.61%
101	101	179572540	179572253	127990	128277	9585	9681	100	98	97						I-band	Ig-like 76	80.67%
102	102	179571681	179571589	128849	128941	9681	9712	101	99	98					N2A unique sequence	I-band	Ig-like 77	79.94%
103	103	179571466	179571181	129064	129349	9712	9807	102	100	99					N2A unique sequence	I-band	Ig-like 77	82.34%
104	104	179570084	179569901	130446	130629	9807	9868	103	101	100					N2A unique sequence	I-band		81.95%
105	105	179569693	179569604	130837	130926	9869	9898	104	102	101					N2A unique sequence	I-band		81.86%
106	106	179569504	179569237	131026	131293	9899	9988	105	103	102					N2A unique sequence	I-band		78.62%
107	107	179569134	179568874	131396	131656	9988	10075	106	104	103					N2A unique sequence	I-band	Ig-like 78	80.73%
108	108	179567390	179567181	133140	133349	10075	10145	107	105	104					N2A unique sequence	I-band	Ig-like 79	82.10%
109	109	179566972	179566895	133558	133635	10145	10171	108	106	105					N2A unique sequence	I-band	Ig-like 79	83.83%
110	110	179566793	179566767	133737	133763	10171	10180	109	107	106						I-band		84.85%
111	111	179566314	179566255	134216	134275	10180	10200	110	108	107						I-band		87.28%
112	112	179565930	179565847	134600	134683	10200	10228	111	109	108						I-band		87.24%
113	113	179563641	179563570	136889	136960	10228	10252	112	110	109						I-band		84.96%

114	114	179561895	179561848	138635	138682	10252	10268	113	111	110						I-band		83.03%
115	115	179560996	179560592	139534	139938	10268	10403	114	112	111						I-band		84.40%
116	116	179560137	179560075	140393	140455	10403	10424	115	113	112						I-band		82.86%
117	117	179559633	179559556	140897	140974	10424	10450	116	114	113						I-band		75.14%
118	118	179559403	179559326	141127	141204	10450	10476	117	115	114						I-band		73.95%
119	119	179558735	179558649	141795	141881	10476	10505	118	116	115						I-band		65.93%
120	120	179558416	179558336	142114	142194	10505	10532	119	117	116						I-band		62.58%
121	121	179557307	179557224	143223	143306	10532	10560	120	118	117						I-band	PEVK 1	64.28%
	122	179556826	179556743	143704	143787	10560	10588	121	119	118						I-band	PEVK 2	58.50%
122	123	179555113	179555072												alternative exon - removed from Havana model			
123	124	179554623	179554540	145907	145990	10588	10616	122	120	119						I-band	PEVK 3	82.93%
124	125	179554322	179554242	146208	146288	10616	10643	123	121	120						I-band	PEVK 4	28.92%
125	126	179554100	179554017	146430	146513	10643	10671	124	122	121						I-band	PEVK 5	32.66%
126	127	179553863	179553780	146667	146750	10671	10699	125	123	122						I-band	PEVK 6	22.35%
127	128	179553505	179553404	147025	147126	10699	10733	126	124	123						I-band		24.28%
128	129	179552951	179552838	147579	147692	10733	10771	127	125	124						I-band		38.24%
129	130	179550325	179550245	150205	150285	10771	10798	128	126	125						I-band		45.85%
130	131	179550057	179549980	150473	150550	10798	10824	129	127	126						I-band		27.25%
131	132	179549716	179549633	150814	150897	10824	10852	130	128	127					Low cardiac expression? ¹⁷	I-band	PEVK7	17.33%
132	133	179549476	179549393	151054	151137	10852	10880	131	129	128					Low cardiac expression? ¹⁷	I-band	PEVK8	18.14%
133	134	179549140	179549057	151390	151473	10880	10908	132	130	129					Low cardiac expression? ¹⁷	I-band		14.66%
134	135	179548809	179548726	151721	151804	10908	10936	133	131	130						I-band	PEVK9	18.60%
135	136	179548018	179547938	152512	152592	10936	10963	134	132	131						I-band		36.96%
136	137	179547630	179547424	152900	153106	10963	11032	135	133	132						I-band		83.51%
137	138	179546465	179546388	154065	154142	11032	11058	136	134	133						I-band		83.20%
138	139	179546177	179546103	154353	154427	11058	11083	137	135	134						I-band		3.79%
139	140	179545898	179545806	154632	154724	11083	11114	138	136	135					Low cardiac expression? ¹⁷	I-band		1.77%
140	141	179545058	179544981	155472	155549	11114	11140	139	137	136					Low cardiac expression? ¹⁷	I-band		1.45%
141	142	179544782	179544621	155748	155909	11140	11194	140	138	137					Low cardiac expression? ¹⁷	I-band		1.41%

142	143	179544409	179544326	156121	156204	11194	11222	141	139	138					Low cardiac expression? ¹⁷	I-band	PEVK10	1.13%
143	144	179544143	179544066	156387	156464	11222	11248	142	140	139					Low cardiac expression? ¹⁷	I-band	PEVK11	1.93%
144	145	179543557	179543474	156973	157056	11248	11276	143	141	140						I-band	PEVK12	81.62%
145	146	179543224	179543141	157306	157389	11276	11304	144	142	141						I-band	PEVK13	79.60%
146	147	179542935	179542852	157595	157678	11304	11332	145	143	142						I-band	PEVK14	80.45%
147	148	179542644	179542348	157886	158182	11332	11431	146	144	143					Low cardiac expression? ¹⁷	I-band		2.37%
148		179542014	179541928	158516	158602	11431	11460	147							New in meta transcript			1.24%
149	149	179540722	179540648	159808	159882	11460	11485	148	145	144					Low cardiac expression? ¹⁷	I-band		3.06%
150	150	179540481	179540398	160049	160132	11485	11513	149										0.94%
151	151	179539840	179539766	160690	160764	11513	11538	150	146	145						I-band		19.51%
152	152	179539136	179539041	161394	161489	11538	11570	151	147	146						I-band		77.13%
153	153	179538437	179538360	162093	162170	11570	11596	152	148	147						I-band		68.18%
154	154	179537430	179537362	163100	163168	11596	11619	153	149	148						I-band		16.13%
155	155	179537208	179537134	163322	163396	11619	11644	154	150	149					Low cardiac expression? ¹⁷	I-band		3.15%
156	156	179536994	179536698	163536	163832	11644	11743	155	151	150					Low cardiac expression? ¹⁷	I-band		3.50%
157	157	179535897	179535817	164633	164713	11743	11770	156	152	151					Low cardiac expression? ¹⁷	I-band	PEVK15	6.09%
158	158	179535022	179534945	165508	165585	11770	11796	157	153	152						I-band	PEVK16	49.51%
159	159	179534402	179534319	166128	166211	11796	11824	158										1.77%
160	160	179534174	179534100	166356	166430	11824	11849	159										1.11%
161	161	179532448	179532365	168082	168165	11849	11877	160										2.35%
162	162	179532252	179532169	168278	168361	11877	11905	161										2.32%
163	163	179532046	179531963	168484	168567	11905	11933	162										1.15%
164	164	179531628	179531551	168902	168979	11933	11959	163										1.24%
165	165	179530518	179530435	170012	170095	11959	11987	164										2.90%
166	166	179530187	179530104	170343	170426	11987	12015	165										2.40%
167	167	179529653	179529579	170877	170951	12015	12040	166										2.62%
168	168	179529464	179529381	171066	171149	12040	12068	167										2.20%
169	169	179529264	179529187	171266	171343	12068	12094	168										2.22%
170	170	179528870	179528742	171705	171788	12094	12122	169							Boundaries re-			2.71%

															annotated in meta-transcript			
171	171	179528629	179528546	171901	171984	12122	12150	170										2.42%
172	172	179528437	179528354	172093	172176	12150	12178	171										2.49%
173	173	179528243	179528160	172287	172370	12178	12206	172										5.97%
174	174	179528076	179527993	172454	172537	12206	12234	173										4.30%
175	175	179527782	179527693	172748	172837	12234	12264	174	154	153					Genomic position remapped in meta-transcript	I-band		73.00%
176	176	179527539	179527456	172991	173074	12264	12292	175										3.32%
177	177	179527343	179527260	173187	173270	12292	12320	176										5.27%
178	178	179527145	179527065	173385	173465	12320	12347	177										2.86%
179	179	179526949	179526869	173581	173661	12347	12374	178										5.22%
180	180	179526755	179526675	173775	173855	12374	12401	179										4.76%
181	181	179526569	179526486	173961	174044	12401	12429	180										1.83%
182	182	179523982	179523899	176548	176631	12429	12457	181	155	154					Genomic position changed in meta- remapped	I-band	PEVK17	15.91%
183	183	179523815	179523732	176715	176798	12457	12485	182	156	155					Genomic position remapped in meta-transcript	I-band	PEVK18	8.96%
184	184	179523521	179523432	177009	177098	12485	12515	183										42.35%
185	185	179523278	179523195	177252	177335	12515	12543	184										7.84%
186	186	179523082	179522999	177448	177531	12543	12571	185										10.83%
187	187	179522884	179522804	177646	177726	12571	12598	186										7.52%
188	188	179522688	179522608	177842	177922	12598	12625	187										9.65%
189	189	179522494	179522414	178036	178116	12625	12652	188										8.35%
190	190	179522308	179522225	178222	178305	12652	12680	189										3.69%
191	191	179519722	179519639	180808	180891	12680	12708	190										8.54%
192	192	179519555	179519472	180975	181058	12708	12736	191										6.00%
193	193	179519261	179519172	181269	181358	12736	12766	192										27.73%
194	194	179519018	179518935	181512	181595	12766	12794	193										4.29%
195	195	179518822	179518739	181708	181791	12794	12822	194										8.57%
196	196	179518624	179518544	181906	181986	12822	12849	195										9.99%

197	197	179518428	179518348	182102	182182	12849	12876	196										12.17%
198	198	179518234	179518154	182296	182376	12876	12903	197										9.52%
199	199	179518042	179517965	182482	182565	12903	12931	198										4.52%
200	200	179517851	179517768	182679	182762	12931	12959	199										3.36%
201		179517658	179517575	182872	182955	12959	12987	200										3.57%
202	201	179517463	179517380	183067	183150	12987	13015	201										2.99%
203	202	179517268	179517185	183262	183345	13015	13043	202	157	156						I-band	PEVK19	6.25%
204	203	179517074	179516991	183456	183539	13043	13071	203	158	157						I-band	PEVK20	14.71%
205	204	179516906	179516823	183624	183707	13071	13099	204	159	158						I-band	PEVK21	22.43%
206	205	179516694	179516611	183836	183919	13099	13127	205	160	159						I-band		24.60%
207	206	179516476	179516393	184054	184137	13127	13155	206	161	160						I-band		21.56%
208	207	179516263	179516180	184267	184350	13155	13183	207	162	161						I-band	PEVK 22	22.41%
209	208	179516047	179515970	184483	184560	13183	13209	208	163	162						I-band		25.04%
210	209	179515561	179515478	184969	185052	13209	13237	209	164	163						I-band	PEVK 23	23.19%
211	210	179514998	179514891	185532	185639	13237	13273	210	165	164						I-band	PEVK 24	11.71%
212	211	179514621	179514544	185909	185986	13273	13299	211	166	165						I-band	PEVK 25	12.63%
213	212	179514358	179514281	186172	186249	13299	13325	212										5.33%
214	213	179514058	179513975	186472	186555	13325	13353	213										3.40%
215	214	179512191	179512108	188339	188422	13353	13381	214										3.94%
216	215	179511871	179511791	188659	188739	13381	13408	215										2.35%
217	216	179511286	179511212	189244	189318	13408	13433	216										3.45%
218	217	179510757	179510647	189773	189883	13433	13470	217	167	166						I-band		8.38%
219	218	179509343	179509275	191187	191255	13470	13493	218	168	167						I-band	PEVK26	8.17%
220	219	179507044	179506964	193486	193566	13493	13520	219	169	168	47	48	48			I-band	PEVK27	99.55%
221	220	179506042	179505968	194488	194562	13520	13545	220	170	169	48	49	49			I-band	PEVK28	98.41%
222	221	179505357	179505268	195173	195262	13545	13575	221	171	170	49	50	50			I-band		100.00%
223	222	179504837	179504775	195693	195755	13575	13596	222	172	171	50	51	51			I-band	PEVK29	100.00%
224	223	179504515	179504426	196015	196104	13596	13626	223	173	172	51	52	52			I-band	PEVK30	100.00%
225	224	179502146	179502096	198384	198434	13626	13643	224	174	173	52	53	53			I-band	PEVK31	73.18%
226	225	179501526	179501125	199004	199405	13643	13777	225	175	174	53	54	54			I-band	Ig-like 80	100.00%
227	226	179500968	179500690	199562	199840	13777	13870	226	176	175	54	55	55			I-band	Ig-like 81	100.00%
228	227	179500442	179500167	200088	200363	13870	13962	227	177	176	55	56	56			I-band	Ig-like 82	100.00%
229	228	179500031	179499892	200499	200638	13962	14008	228	178	177	56	57	57			I-band		100.00%
230	229	179499576	179499450	200954	201080	14009	14051	229	179	178	57	58	58			I-band		100.00%
231	230	179499356	179499093	201174	201437	14051	14139	230	180	179	58	59	59			I-band		100.00%
232	231	179498810	179498544	201720	201986	14139	14228	231	181	180	59	60	60			I-band	Ig-like 83	100.00%

233	232	179498403	179498140	202127	202390	14228	14316	232	182	181	60	61	61			I-band	Ig-like 84	100.00%
234	233	179498053	179497914	202477	202616	14316	14362	233	183	182	61	62	62			I-band		100.00%
235	234	179497771	179497645	202759	202885	14363	14405	234	184	183	62	63	63			I-band		100.00%
236	235	179497519	179497253	203011	203277	14405	14494	235	185	184	63	64	64			I-band	Ig-like 85	100.00%
237	236	179497140	179496874	203390	203656	14494	14583	236	186	185	64	65	65			I-band		100.00%
238	237	179496027	179495761	204503	204769	14583	14672	237	187	186	65	66	66			I-band	Ig-like 86	100.00%
239	238	179495670	179495531	204860	204999	14672	14718	238	188	187	66	67	67			I-band		100.00%
240	239	179495094	179494968	205436	205562	14719	14761	239	189	188	67	68	68			I-band		100.00%
241	240	179494170	179494028	206360	206502	14761	14808	240	190	189	68	69	69			I-band	Ig-like 87	100.00%
242	241	179490123	179490000	210407	210530	14809	14850	241	191	190	69	70	70			I-band	Ig-like 87	100.00%
243	242	179489458	179489192	211072	211338	14850	14939	242	192	191	70	71	71			I-band	Ig-like 88	47.15%
244	243	179487494	179487397	213036	213133	14939	14971	243	193	192	71	72	72			I-band	Ig-like 89	100.00%
245	244	179486735	179486567	213795	213963	14972	15028	244	194	193	72	73	73			I-band	Ig-like 89	100.00%
246	245	179486468	179486202	214062	214328	15028	15117	245	195	194	73	74	74			I-band	Ig-like 90	100.00%
247	246	179486095	179485829	214435	214701	15117	15206	246	196	195	74	75	75			I-band	Ig-like 91	100.00%
248	247	179485720	179485442	214810	215088	15206	15299	247	197	196	75	76	76			I-band	Ig-like 92	100.00%
249	248	179485352	179484944	215178	215586	15299	15435	248	198	197	76	77	77			I-band	Ig-like 93/94	100.00%
250	249	179484839	179484715	215691	215815	15435	15477	249	199	198	77	78	78			I-band	Ig-like 94	100.00%
251	250	179484614	179484348	215916	216182	15477	15566	250	200	199	78	79	79			I-band		100.00%
252	251	179483580	179483311	216950	217219	15566	15656	251	201	200	79	80	80			I-band	Ig-like 95	100.00%
253	252	179483218	179482916	217312	217614	15656	15757	252	202	201	80	81	81			A-band	Fibronectin type-III 1	100.00%
254	253	179482808	179482506	217722	218024	15757	15858	253	203	202	81	82	82			A-band	Fibronectin type-III 2	100.00%
255	254	179482239	179482052	218291	218478	15858	15920	254	204	203	82	83	83			A-band	Fibronectin type-III 3	100.00%
256	255	179481961	179481847	218569	218683	15921	15959	255	205	204	83	84	84			A-band	Fibronectin type-III 3	100.00%
257	256	179481740	179481456	218790	219074	15959	16054	256	206	205	84	85	85			A-band		100.00%
258	257	179481357	179481206	219173	219324	16054	16104	257	207	206	85	86	86			A-band	Fibronectin type-III 4	100.00%
259	258	179480515	179480368	220015	220162	16105	16154	258	208	207	86	87	87			A-band	Fibronectin type-III 4	100.00%
260	259	179480211	179480034	220319	220496	16154	16213	259	209	208	87	88	88			A-band	Fibronectin type-III 5	100.00%
261	260	179479695	179479574	220835	220956	16213	16254	260	210	209	88	89	89			A-band	Fibronectin type-III 5	100.00%
262	261	179479480	179479193	221050	221337	16254	16350	261	211	210	89	90	90			A-band	Ig-like 96	100.00%
263	262	179479075	179478779	221455	221751	16350	16449	262	212	211	90	91	91			A-band	Fibronectin type-III 6	100.00%
264	263	179478664	179478478	221866	222052	16449	16511	263	213	212	91	92	92			A-band	Fibronectin type-III 7	100.00%
265	264	179478003	179477888	222527	222642	16511	16550	264	214	213	92	93	93			A-band	Fibronectin type-III 7	100.00%
266	265	179477799	179477500	222731	223030	16550	16650	265	215	214	93	94	94			A-band	Fibronectin type-III 8	100.00%
267	266	179477303	179477004	223227	223526	16650	16750	266	216	215	94	95	95			A-band	Fibronectin type-III 9	100.00%
268	267	179476889	179476784	223641	223746	16750	16785	267	217	216	95	96	96			A-band	Fibronectin type-III 10	100.00%

269	268	179476681	179476485	223849	224045	16785	16851	268	218	217	96	97	97			A-band	Fibronectin type-III 10	100.00%
270	269	179476404	179476099	224126	224431	16851	16953	269	219	218	97	98	98			A-band	Fibronectin type-III 11	100.00%
271	270	179475998	179475720	224532	224810	16953	17046	270	220	219	98	99	99			A-band	Ig-like 97	100.00%
272	271	179475116	179474817	225414	225713	17046	17146	271	221	220	99	100	100			A-band	Fibronectin type-III 12	100.00%
273	272	179474713	179474411	225817	226119	17146	17247	272	222	221	100	101	101			A-band	Fibronectin type-III 13	100.00%
274	273	179474297	179473935	226233	226595	17247	17368	273	223	222	101	102	102			A-band	Ig-like 98	100.00%
275	274	179473635	179473333	226895	227197	17368	17469	274	224	223	102	103	103			A-band	Fibronectin type-III 14	100.00%
276	275	179473204	179472905	227326	227625	17469	17569	275	225	224	103	104	104			A-band	Fibronectin type-III 15	100.00%
277	276	179472808	179472512	227722	228018	17569	17668	276	226	225	104	105	105			A-band	Fibronectin type-III 16	100.00%
278	277	179472412	179472128	228118	228402	17668	17763	277	227	226	105	106	106			A-band	Ig-like 99	100.00%
279	278	179472041	179471748	228489	228782	17763	17861	278	228	227	106	107	107			A-band	Fibronectin type-III 17	100.00%
280	279	179470440	179470141	230090	230389	17861	17961	279	229	228	107	108	108			A-band	Fibronectin type-III 18	100.00%
281	280	179470022	179469714	230508	230816	17961	18064	280	230	229	108	109	109			A-band	Ig-like 100	100.00%
282	281	179469625	179469435	230905	231095	18064	18127	281	231	230	109	110	110			A-band	Fibronectin type-III 19	100.00%
283	282	179469032	179468603	231498	231927	18128	18271	282	232	231	110	111	111			A-band	Fibronectin type-III 20	100.00%
284	283	179467317	179467009	233213	233521	18271	18374	283	233	232	111	112	112			A-band	Fibronectin type-III 21	100.00%
285	284	179466877	179466729	233653	233801	18374	18423	284	234	233	112	113	113			A-band	Ig-like 101	100.00%
286	285	179466641	179466609	233889	233921	18424	18434	285	235	234	113	114	114			A-band	Ig-like 101	100.00%
287	286	179466514	179466385	234016	234145	18435	18478	286	236	235	114	115	115			A-band	Ig-like 101	100.00%
288	287	179466291	179465992	234239	234538	18478	18578	287	237	236	115	116	116			A-band	Fibronectin type-III 22	100.00%
289	288	179465898	179465581	234632	234949	18578	18684	288	238	237	116	117	117			A-band	Fibronectin type-III 23	100.00%
290	289	179464577	179464281	235953	236249	18684	18783	289	239	238	117	118	118			A-band	Ig-like 102	100.00%
291	290	179464172	179463873	236358	236657	18783	18883	290	240	239	118	119	119			A-band	Fibronectin type-III 24	100.00%
292	291	179463789	179463475	236741	237055	18883	18988	291	241	240	119	120	120			A-band	Fibronectin type-III 25	100.00%
293	292	179463381	179463233	237149	237297	18988	19037	292	242	241	120	121	121			A-band	Fibronectin type-III 26	100.00%
294	293	179462785	179462635	237745	237895	19038	19088	293	243	242	121	122	122			A-band	Fibronectin type-III 26	100.00%
295	294	179462546	179462265	237984	238265	19088	19182	294	244	243	122	123	123			A-band	Ig-like 103	100.00%
296	295	179460536	179460234	239994	240296	19182	19283	295	245	244	123	124	124			A-band	Fibronectin type-III 27	100.00%
297	296	179459373	179459071	241157	241459	19283	19384	296	246	245	124	125	125			A-band	Fibronectin type-III 28	100.00%
298	297	179458969	179458688	241561	241842	19384	19478	297	247	246	125	126	126			A-band	Ig-like 104	100.00%
299	298	179458594	179458295	241936	242235	19478	19578	298	248	247	126	127	127			A-band	Fibronectin type-III 29	100.00%
300	299	179458202	179457900	242328	242630	19578	19679	299	249	248	127	128	128			A-band	Fibronectin type-III 30	100.00%
301	300	179457810	179457502	242720	243028	19679	19782	300	250	249	128	129	129			A-band	Fibronectin type-III 31	100.00%
302	301	179457387	179457106	243143	243424	19782	19876	301	251	250	129	130	130			A-band	Ig-like 105	100.00%
303	302	179457004	179456705	243526	243825	19876	19976	302	252	251	130	131	131			A-band	Fibronectin type-III 32	100.00%
304	303	179456619	179456326	243911	244204	19976	20074	303	253	252	131	132	132			A-band	Fibronectin type-III 33	100.00%

305	304	179456231	179453265	244299	247265	20074	21063	304	254	253	132	133	133			A-band	Ig-like 106, Fibronectin type-III 34/35/36, Ig-like 107, Fibronectin type-III 37/38, Ig-like 108, Fibronectin type-III 39/40	100.00%
306	305	179452946	179452626	247584	247904	21063	21170	305	255	254	133	134	134			A-band	Fibronectin type-III 41	100.00%
307	306	179452527	179452243	248003	248287	21170	21265	306	256	255	134	135	135			A-band	Ig-like 109	100.00%
308	307	179452144	179451845	248386	248685	21265	21365	307	257	256	135	136	136			A-band	Fibronectin type-III 42	100.00%
309	308	179451534	179451232	248996	249298	21365	21466	308	258	257	136	137	137			A-band	Fibronectin type-III 43	100.00%
310	309	179450074	179449799	250456	250731	21466	21558	309	259	258	137	138	138			A-band	Ig-like 110	100.00%
311	310	179449695	179449396	250835	251134	21558	21658	310	260	259	138	139	139			A-band	Fibronectin type-III 44	100.00%
312	311	179449305	179449003	251225	251527	21658	21759	311	261	260	139	140	140			A-band	Fibronectin type-III 45	100.00%
313	312	179448633	179448334	251897	252196	21759	21859	312	262	261	140	141	141			A-band	Fibronectin type-III 46	100.00%
314	313	179447954	179447667	252576	252863	21859	21955	313	263	262	141	142	142			A-band	Ig-like 111	100.00%
315	314	179447319	179447023	253211	253507	21955	22054	314	264	263	142	143	143			A-band	Fibronectin type-III 47	100.00%
316	315	179446935	179446633	253595	253897	22054	22155	315	265	264	143	144	144			A-band	Fibronectin type-III 48	100.00%
317	316	179446531	179446226	253999	254304	22155	22257	316	266	265	144	145	145			A-band	Fibronectin type-III 49	100.00%
318	317	179445336	179445049	255194	255481	22257	22353	317	267	266	145	146	146			A-band		100.00%
319	318	179444956	179444666	255574	255864	22353	22450	318	268	267	146	147	147			A-band	Fibronectin type-III 50	100.00%
320	319	179444575	179444288	255955	256242	22450	22546	319	269	268	147	148	148			A-band	Fibronectin type-III 51	100.00%
321	320	179444120	179443533	256410	256997	22546	22742	320	270	269	148	149	149			A-band	Ig-like 112	100.00%
322	321	179443442	179443338	257088	257192	22742	22777	321	271	270	149	150	150			A-band	Fibronectin type-III 52	100.00%
323	322	179442912	179442715	257618	257815	22777	22843	322	272	271	150	151	151			A-band	Fibronectin type-III 53	100.00%
324	323	179442625	179442329	257905	258201	22843	22942	323	273	272	151	152	152			A-band	Fibronectin type-III 54	100.00%
325	324	179442237	179441650	258293	258880	22942	23138	324	274	273	152	153	153			A-band	Ig-like 113, Fibronectin type-III 55	100.00%
326	325	179441558	179441256	258972	259274	23138	23239	325	275	274	153	154	154			A-band	Fibronectin type-III 56	100.00%
327	326	179441143	179424038	259387	276492	23239	28941	326	276	275	154	155	155			A-band	Fibronectin type-III 57, Ig-like 114, Fibronectin type-III 58/59, Ig-like 115, Fibronectin type-III 60/61/62, Ig-like 116, Fibronectin type-III 63/64/65, Ig-like 117, Fibronectin type-III 66/67, Ig-like 118, Fibronectin type-III 68/69/70, Ig-like	100.00%

																	119, Fibronectin type-III 71/72/73, Ig-like 120, Fibronectin type-III 74/75, Ig-like 121, Fibronectin type-III 76/77/78, Ig-like 122, Fibronectin type-III 79/80/81, Ig-like 123, Fibronectin type-III 82/83, Ig-like 124, Fibronectin type-III 84/85/86, Ig-like 125, Fibronectin type-III 87/88/89, Ig-like 126, Fibronectin type-III 90/91, Ig-like 127, Fibronectin type-III 92/93/94, Ig-like 128, Fibronectin type-III 96/96/97, Ig-like 129, Fibronectin type-III 98	
328	327	179423364	179423068	277166	277462	28941	29040	327	277	276	155	156	156			A-band	Fibronectin type-III 99	100.00%
329	328	179422962	179422375	277568	278155	29040	29236	328	278	277	156	157	157			A-band	Fibronectin type-III 100	100.00%
330	329	179422282	179421980	278248	278550	29236	29337	329	279	278	157	158	158			A-band	Fibronectin type-III 101	100.00%
331	330	179421871	179421575	278659	278955	29337	29436	330	280	279	158	159	159			A-band	Fibronectin type-III 102	100.00%
332	331	179419879	179419592	280651	280938	29436	29532	331	281	280	159	160	160			A-band	Ig-like 130	100.00%
333	332	179419479	179419180	281051	281350	29532	29632	332	282	281	160	161	161			A-band	Fibronectin type-III 103	100.00%
334	333	179418943	179418641	281587	281889	29632	29733	333	283	282	161	162	162			A-band	Fibronectin type-III 104	100.00%
335	334	179418534	179418229	281996	282301	29733	29835	334	284	283	162	163	163			A-band	Fibronectin type-III 105	100.00%
336	335	179418123	179416357	282407	284173	29835	30424	335	285	284	163	164	164			A-band	Ig-like 131, Fibronectin type-III 106/107, Ig-like 132, Fibronectin type-III 108/109	100.00%
337	336	179415987	179415694	284543	284836	30424	30522	336	286	285	164	165	165			A-band	Fibronectin type-III 110	100.00%
338	337	179415000	179414713	285530	285817	30522	30618	337	287	286	165	166	166			A-band	Ig-like 133	100.00%
339	338	179414596	179414297	285934	286233	30618	30718	338	288	287	166	167	167			A-band	Fibronectin type-III 111	100.00%
340	339	179414200	179412134	286330	288396	30718	31407	339	289	288	167	168	168			A-band	Fibronectin type-III 112/113, Ig-like 134, Fibronectin type-III 114/115, Ig-like 135,	100.00%

																	Fibronectin type-III 116		
341	340	179412032	179411730	288498	288800	31407	31508	340	290	289	168	169	169				A-band	Fibronectin type-III 117	99.80%
342	341	179411632	179411327	288898	289203	31508	31610	341	291	290	169	170	170				A-band	Fibronectin type-III 118	99.54%
343	342	179411229	179410939	289301	289591	31610	31707	342	292	291	170	171	171				A-band	Ig-like 136	99.59%
344	343	179410843	179410547	289687	289983	31707	31806	343	293	292	171	172	172				A-band	Fibronectin type-III 119	99.47%
345	344	179410420	179410115	290110	290415	31806	31908	344	294	293	172	173	173				A-band	Fibronectin type-III 120	99.02%
346	345	179409233	179408928	291297	291602	31908	32010	345	295	294	173	174	174				A-band	Fibronectin type-III 121	99.14%
347	346	179408842	179408561	291688	291969	32010	32104	346	296	295	174	175	175				A-band	Ig-like 137	99.28%
348	347	179408389	179407796	292141	292734	32104	32302	347	297	296	175	176	176				A-band	Fibronectin type-III 122/123	99.43%
349	348	179407676	179407389	292854	293141	32302	32398	348	298	297	176	177	177				A-band	Ig-like 138	99.48%
350	349	179407290	179406991	293240	293539	32398	32498	349	299	298	177	178	178				A-band	Fibronectin type-III 124	99.43%
351	350	179406311	179406009	294219	294521	32498	32599	350	300	299	178	179	179				A-band	Fibronectin type-III 125	99.59%
352	351	179405097	179404795	295433	295735	32599	32700	351	301	300	179	180	180				A-band	Fibronectin type-III 126	99.68%
353	352	179404693	179404109	295837	296421	32700	32895	352	302	301	180	181	181				A-band	Ig-like 139, Fibronectin type-III 127	99.78%
354	353	179403978	179403673	296552	296857	32895	32997	353	303	302	181	182	182				A-band	Fibronectin type-III 128	99.84%
355	354	179403566	179403267	296964	297263	32997	33097	354	304	303	182	183	183				A-band	Fibronectin type-III 129	99.83%
356	355	179402644	179402069	297886	298461	33097	33289	355	305	304	183	184	184				A-band	Ig-like 140	99.85%
357	356	179401970	179401665	298560	298865	33289	33391	356	306	305	184	185	185				A-band	Fibronectin type-III 130	99.88%
358	357	179401302	179400709	299228	299821	33391	33589	357	307	306	185	186	186				A-band	Fibronectin type-III 131, Ig-like 141	99.11%
359	358	179400576	179394968	299954	305562	33589	35458	358	308	307	186	187	187				M-band	Ig-like 142, Fibronectin type-III 132, Protein kinase, Ig-like 143/144/145/146/147/148	99.89%
360	359	179394843	179394687	305687	305843	35459	35511	359	309	308	187	188	188				M-band	Ig-like 149	99.62%
361	360	179393946	179393255	306584	307275	35511	35741	360	310	309	188	189	189				M-band	Ig-like 150	99.80%
362	361	179393154	179393001	307376	307529	35742	35793	361	311	310	189	190	190				M-band	Ig-like 151	99.95%
363	362	179392475	179392173	308055	308357	35793	35894	362	312	311	190	191	191				M-band		99.42%
364	363	179392034	179390716	308496	309814	35894	35991	363	313	312	191	192	192				M-band	Ig-like 152	100.00%

3.9.3 Splice variant prediction

To identify splice variants outside of the essential splice site comprised of two conserved intronic bases, AlamutHT¹⁹⁰ was used to calculate Maximum Entropy¹⁹¹, Neural Network (NNSplice), Splice Site Finder (SSF), Human Splice Finder (HSF), and Gene Splicer (GS) scores for reference and alternate alleles. The FHS cohort was used to generate a threshold value for calling splice variant predictions: for each variant in the splicing region (donor: -3 to +6, acceptor: -20 to +3), the wild type splicing score was subtracted from the variant score to give a score change value and then converted to percentiles. Variants that scored above the 90th percentile by at least 3 algorithms *and* above the 70th percentile by all applied algorithms were considered conservative splicing variant predictions. In the instance where one or more algorithm did not produce a score for a variant, 4/4 or 3/3 algorithms scoring above the 70th percentile was considered sufficient. For variants meeting conservative splicing variant criteria in the FHS cohort the minimum absolute score change for each prediction algorithm was applied as a threshold across all cohorts (Table 3.3). These thresholds were selected to be more conservative than the previously applied Maximum Entropy Score difference ≤ -2 threshold¹⁷ and to exclude variants found more frequently than in 1 in 1,000 individuals.

Table 3.3 Minimum absolute score change for each prediction algorithm applied as threshold across all cohorts

	Absolute score change threshold	
	70 th percentile	90 th percentile
Maximum Entropy (MaxEnt)	-0.89	-1.73
Neural Network (NNSplice)	-0.53	-5.9
Splice Site Finder (SSF)	-0.077	-0.590
Human Splice Finder (HSF)	-0.25	-3.2
Gene Splicer (GS)	-0.91	-1.81

3.10 Statistical analyses

3.10.1 Overview

Statistical analyses were performed using R (<http://www.R-project.org>). Comparisons between groups were performed using Mann-Whitney or Fisher's exact tests, as appropriate. Standard linear regressions were used to evaluate the relationship between *TTN* genotype and cardiovascular phenotypes. Multivariate models were generated using known clinical covariates and optimised to minimise Bayesian

information criterion. The relationships between morphologic parameters and *TTN* genotype were assessed by ANOVA between nested linear models.

3.10.2 Determining the likelihood that a *TTN*tv found in an individual with DCM is pathogenic

Excluding Novex, low expression & predicted splice site variants:

Total *TTN*tv burden in controls = $31/3911 = 0.8\%$

Total *TTN*tv burden in cases = $72/529 = 14\%$

In order to estimate the proportion of variants in cases that are truly pathogenic, the burden of *TTN*tv in controls without overt disease is taken as an estimate of the burden of benign variation in both cohorts (0.8%). The burden of pathogenic *TTN*tv in cases is therefore estimated at 13.2%.

In an individual with DCM, the likelihood that a *TTN*tv is pathogenic = $13.2/0.8 = 16.5$, and the probability of pathogenicity = $13.2/14 = 94\%$

3.10.3 Stratification of DCM and linear modelling

For prospectively recruited DCM subjects, linear modelling was used to robustly assess the relationship between *TTN* genotype and cardiac phenotype. In these analyses, we sought to determine which of four inter-related positional factors were most informative: distance of the *TTN*tv from the N-terminus; cardiac expression of the variant exon; the transcripts affected by the *TTN*tv; and location inside/outside the A-band. Taken together, these genotype-based descriptors significantly improved model performance for six parameters, namely LV EF, RV EF, LV SVi, RV SVi, LVMi, and lateral WTi. For LV EF, presence inside/outside the A-band was most informative, whereas distance from the N-terminus was the strongest predictor for the other parameters. Definitive models were then constructed, with *TTN* genotype defined by presence/absence of *TTN*tv and the distance of the *TTN*tv from the N-terminus, while taking age and sex into account. With these refinements, *TTN* genotype remained a significant predictor of five phenotypic indices (LV EF, RV EF, LV SVi, RV SVi and lateral WTi)

4 Deep phenotyping of a DCM cohort

4.1 Introduction

A principle goal of genetic research is to identify specific genotypes that are associated with human phenotypes¹⁹². Increasing the accuracy and completeness of phenotype information ensures that genetic investigations have maximal power to identify a variant influencing the presence of this manifestation, and determine genotype-phenotype correlation¹⁹³. Deep phenotyping, the precise and comprehensive analysis of phenotypic abnormalities in which the individual components of the phenotype are observed and described,¹⁹⁴ is vital in a complex and heterogeneous condition such as DCM. With advances in high throughput genotyping technologies the rate limiting step has become the collection of sensitive and specific phenotype information in large samples of study participants¹⁹⁵. Geneticists are now accustomed to highly precise technologies: genotype miscall rates are now typically less than 0.5%, whereas measurement error in phenotypes has, until recently, received less attention. Confounding factors, including study heterogeneity originating from clinical sources such as phenocopy or a common final disease picture can be minimized as careful measurement of environmental risk factors and disease sub-phenotypes can provide valuable clues to pathophysiology¹⁹⁵. The importance and utility of high quality phenotype data was demonstrated by the identification of MLL2 mutations as a cause of Kabuki syndrome, a discovery made possible only once a phenotypic stratification strategy was employed⁸⁶.

CMR has been applied for the measurement of LV volumes, systolic function and LV mass for several years in the clinical arena¹⁷⁶. The excellent accuracy and reproducibility that make CMR the gold standard technique¹⁶⁹ also make it the ideal tool for deep phenotyping in this study. The greatest restriction to the use of CMR probably remains the limited distribution and availability of scanners, and in the research setting the cost of such detailed phenotyping for a large cohort. As such until recent studies published by the RBHT CMR unit^{179,180}, studies using CMR phenotyping were limited to much smaller cohorts¹⁹⁶⁻²⁰⁰.

4.2 Aims

- To collect and curate data on a large cohort of idiopathic DCM patients, with detailed and accurate phenotype data gained through CMR analysis to maximise the power of genotype phenotype analyses
- To explore the data collected for this cohort both independently and in comparison to a healthy volunteer cohort in order to identify any unexpected or unexplained measurements or distributions as a QC step
- To describe the cohort and ensure selection of appropriate statistical tests for subsequent analyses.

4.3 Results

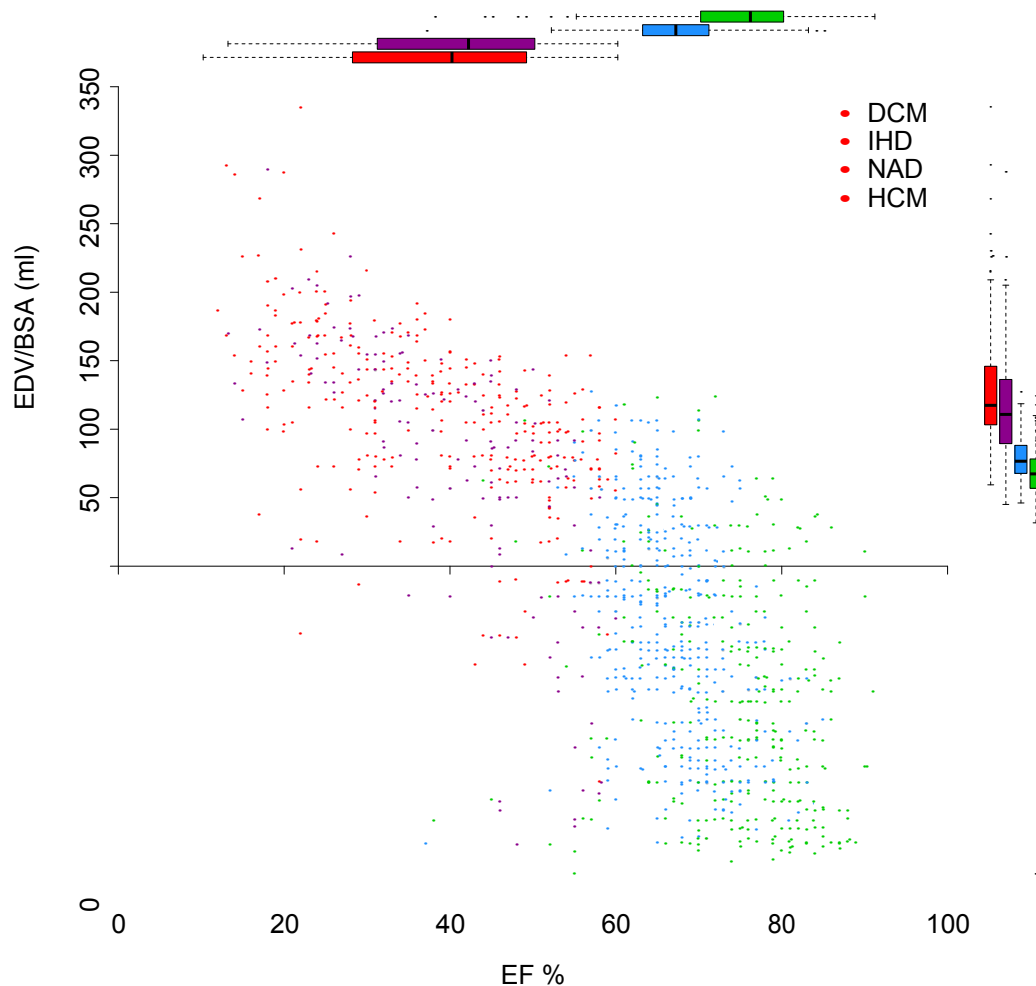
Consecutive patients referred for cardiac evaluation by CMR were prospectively recruited and biobanked. Figure 4.1 shows the relation of LVEF and LVEDVi for the four most frequent diagnoses following initial assessment and highlights complete overlap in these parameters for DCM and ischaemic cardiomyopathy (IHD) patients as well as a lesser overlap for HCM and ostensibly normal scans.

4.3.1 Establishing a DCM cohort using CMR phenotyping and comparison to the normalised mean for LVEF and LVEDVi

The difficulty in making an accurate diagnosis of DCM (discussed in introduction section 1.2.3 Aetiology) and the particular utility of CMR in this study is demonstrated by the proportion of patients referred with a possible diagnosis of DCM in whom the diagnosis was excluded following CMR: Of the 960 patients referred with a preliminary diagnosis of DCM only 518 were confirmed to have DCM, and 442 were assigned a different diagnosis. A full list of the alternate diagnoses is shown in Figure 3.1. A large proportion of those in whom DCM was excluded were relatives of a DCM proband undergoing family screening (n=130). An additional 19 patients referred with other preliminary diagnoses were found to have DCM on CMR, see Figure 3.1 for full details. In total 537 patients with DCM were recruited and biobanked at the time of preparation of this thesis. Full phenotype data was collected and recorded for the first 374 cases (for whom TTN sequence data was also available).

Figure 4.1 Relation of LVEF and LVEDVi for the four most frequent diagnoses on initial CMR analysis

IHD= ischaemic cardiomyopathy, NAD= no abnormality detected on CMR, HCM= hypertrophic cardiomyopathy



Of the 374 included here 319 were confirmed to have DCM using strict CMR characteristics^{173,176} (LVEDVi>2SD above & LVEF>2SD below the mean for age and sex) by two experienced, Level 3 CMR accredited cardiologists blinded to information on genetic status. Normalised mean LVEDVi and LVEF for age and sex were derived from a study by Maciera et al¹⁷⁶ undertaken at the RBHT CMR unit using the same scanners as this study, thus reducing bias in measurement between cohorts. The authors found that most clinical parameters of LV volume and systolic/diastolic function are significantly and independently influenced by gender, age and BSA. Their findings suggest that the interpretation of LV parameters in borderline clinical cases, such as early cardiomyopathy, should be referred to age, gender and body surface area (BSA) normalised values in order to determine normality, or severity of abnormality. An algorithm was developed to perform comparison to normalised

values as recommended. Following this all cases given a diagnosis of DCM but not meeting strict criteria were reevaluated and passed to a level 3 CMR accredited cardiologist to confirm exclusion from this study or inclusion based on clinical history and supportive CMR characteristics: borderline EDVi and EF (EDVi > 1.5SD above & EF > 1.5SD below the mean for age and sex) were considered supportive CMR

characteristics in the presence of corroborating clinical and/or echocardiographic evidence. A further 55 cases were classified as DCM based on clinical history and supportive CMR characteristics. Most patients in this group met criteria for reduced LVEF but not increased LVEDVi (LVEDVi > 2SD below and LVEF 1.5-2SD above the mean (n=8); LVEDVi 1.5-2SD above and LVEF > 2SD below the mean (n=40); LVEDVi 1.5-2SD above and LVEF 1.5-2SD below the mean (n=8)). One mild familial case that would otherwise have been excluded was included based on meeting criteria once adjusted for age and sex.

Figure 4.2 shows a formal comparison of the effect of age and sex on continuous CMR parameters for the DCM cohort. Interestingly four important markers of DCM severity follow opposing trends compared to the healthy asymptomatic subjects studied by Maciera et al ¹⁷⁶: In healthy subjects LVEF increases with age whereas in DCM LVEF decreases with increasing age for both males and females. In healthy subjects LVEDVi decreases with increasing age whereas in DCM LVEDVi remains unchanged in males and increases slightly with age in females. In healthy subjects LV SVi decreases with increasing age whereas LV SVi increases with age in both males and females. In healthy subjects LV mass indexed to BSA (LVMI) decreases with increasing age in males and remains unchanged in females whereas in DCM LVMI increases with age in both males and females.

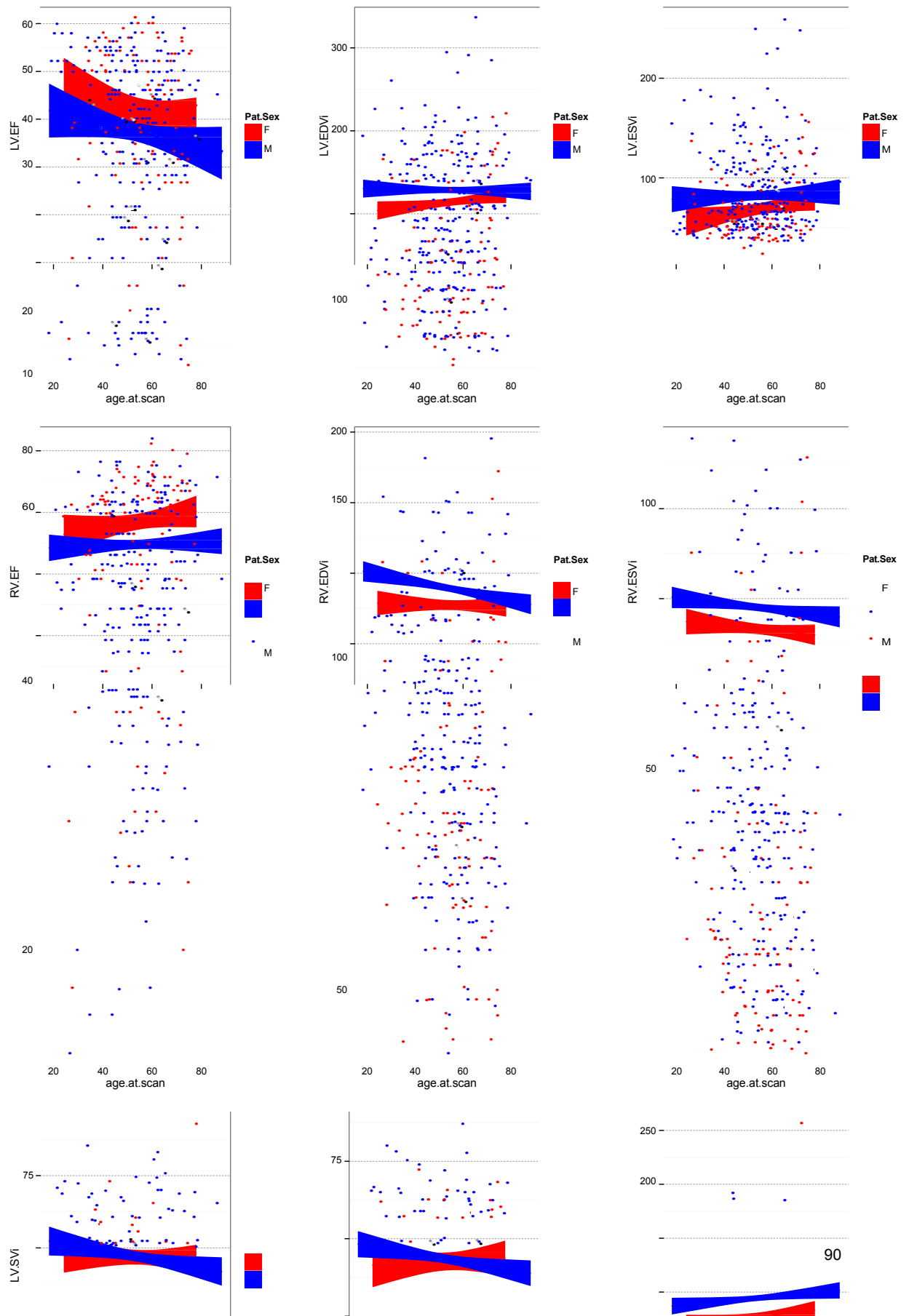
Table 4.1 Full demographic data and results for categorical variables.

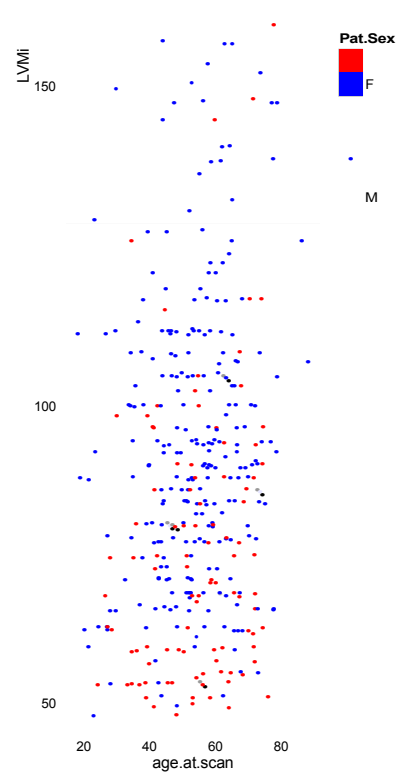
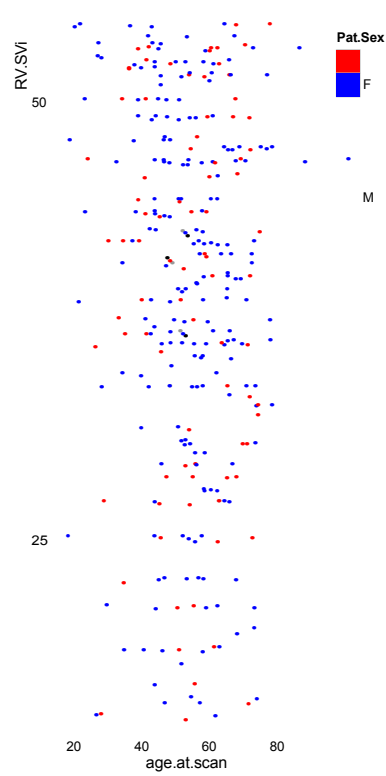
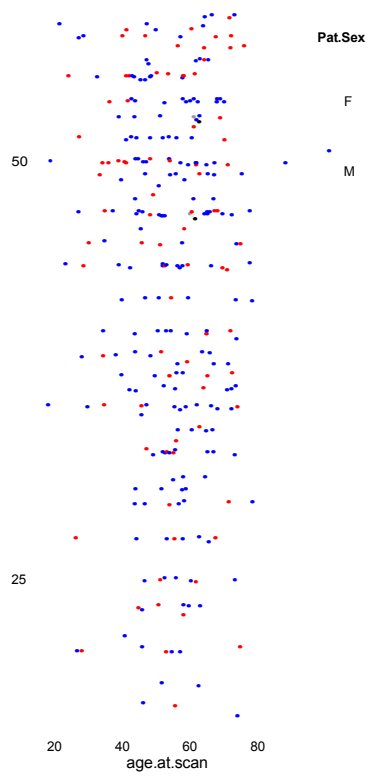
See general methods for details of variables recorded

Categorical variable		N	Number with data available	Percentage
Ethnicity			374	
	Caucasian	329		88.0%
	Asian subcontinent	16		4.3%
	Afrocaribbean	12		3.2%
	African	4		1.1%
	Chinese	3		0.8%
	Mixed	1		0.3%
	Other	9		2.4%
Female		104	374	27.7%
Preliminary diagnosis			374	
	DCM	354		94.7%
	Arrhythmia	4		1.1%
	ARVC	4		1.1%
	IHD	4		1.1%
	HCM	2		0.5%
	Other	6		1.6%
Family history			299	
	DCM	51		17.1%
	SCD	15		5.0%
	ARVC	1		0.3%
	HCM	0		0%
	Skeletal myopathy	0		0%
NYHA class			345	
	1	159		46.1%
	2	138		40.0%
	3	46		13.3%
	4	2		0.6%
History of myocarditis		14	301	4.7%
Normal coronary arteries		250	274	91.2%
Bystander myocardial infarction		24	368	6.5%
Alcohol excess		45	320	14.1%
Chemotherapy		15	298	5.0%
Iron overload		5	318	1.6%
Peripartum		8	334	2.4%
Conduction disease		97	306	31.7%
Sustained VT		30	131	22.9%
Atrial fibrillation		91	315	28.9%
Mitral regurgitation		62	373	16.6%
LVNC		7	366	1.9%
GAD suggestive of myocarditis		26	361	7.2%
Mid wall GAD			367	
	None	246		65.6%
	Limited	89		23.7%
	Extensive	32		8.5%
Most affected midwall segment			122	
	Anterior	4		3.3%
	Inferior	16		13.1%
	Lateral	24		19.7%
	Septal	78		63.9%

Figure 4.2 The effect of age and sex on continuous CMR parameters for the DCM cohort

Each parameter is plotted against age at time of scan. Plots show individual cases as points (red=female, blue=male) with regression lines and 95% confidence intervals (red=female, blue=male) .





4.3.2 Exploration of CMR parameters in a DCM cohort and comparison to healthy volunteers

Frequency density and distribution of cardiac morphology and function as measured by CMR was explored in DCM cases and healthy volunteers. As some measurements are expected to differ for males and females, both the DCM and healthy volunteer cohorts were stratified accordingly ¹⁷⁶. Figure 4.3 shows frequency density plots of CMR parameters for males and females in the DCM and healthy volunteer cohorts alongside probability density histograms with raw and smoothed density curves for the DCM cohort only (probability density histograms are not stratified by gender). Volumetric and contractility parameters show the expected difference between males and females for both DCM patients and healthy volunteers (namely lower EF and larger volumes in males compared to females) with the greatest difference between sexes observed for LVMi and right ventricular EF, EDVi and ESVi (Figure 4.3a-c). Negligible gender difference was observed for biventricular SVi and for all wall thickness measurements.

For all cardiac parameters the spread of the DCM cohorts was greater than that of the healthy volunteers, as may be expected, likely reflecting the broad spectrum of disease in DCM cohorts. The mean of four important markers of DCM severity (LV EF, EDVi, ESVi and LVMi) was substantially distinct from the healthy volunteer cohort ($P < 9.8 \times 10^{-58}$), and three additional parameters were significantly different ($P < 0.0007$) calculated using a Wilcoxon rank sum test with continuity correction (Table 4.2). On visual inspection, the distributions of CMR phenotypes for the DCM cohort are grossly normal. For LVEF and RVEF the frequency density has a shoulder suggestive of a bimodal distribution, which could indicate population stratification or the presence of two subpopulations. Exploring this using the probability density histograms the shoulder in the frequency density curve appears to be a stochastic effect suggesting this is a function of the degree of smoothing applied in the generation of the plot rather than a true bimodal distribution (Figure 4.3a and d). LVEF has a platykurtic distribution, limited on the right by the diagnostic criteria ($EF > 2SD$ below the normalised mean) and on the left by compatibility with life. ESVi and to a lesser extent EDVi show a positive skew, most likely due to exclusion of low EDVi by diagnostic criteria limiting the distribution on the left. Left ventricular EF, EDVi and ESVi are most discriminatory for DCM, while the distribution for right ventricular EF, EDVi and ESVi show greater overlap with the healthy volunteer cohort and the biventricular SVi is not discriminatory. Wall thickness for the healthy

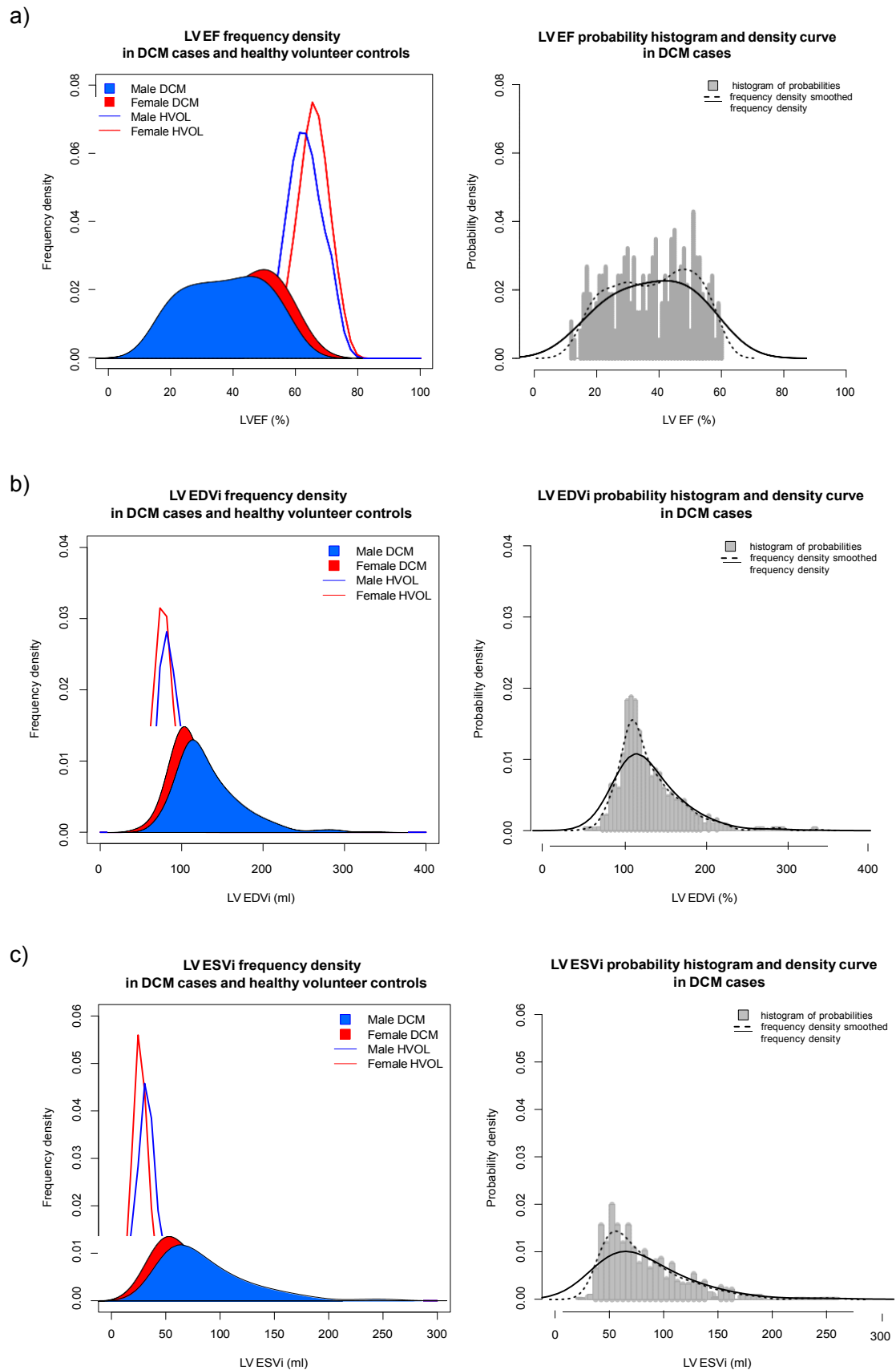
volunteer cohort was not measured at the same position as in the DCM cohort (the healthy volunteer cohort comprises a separate study) so direct comparison cannot be made (Figure 4.4).

Table 4.2 Comparison of means assesses the difference between DCM and healthy volunteer cohorts.

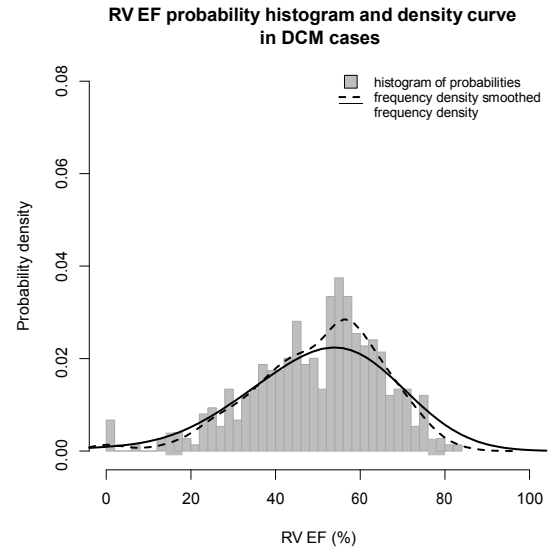
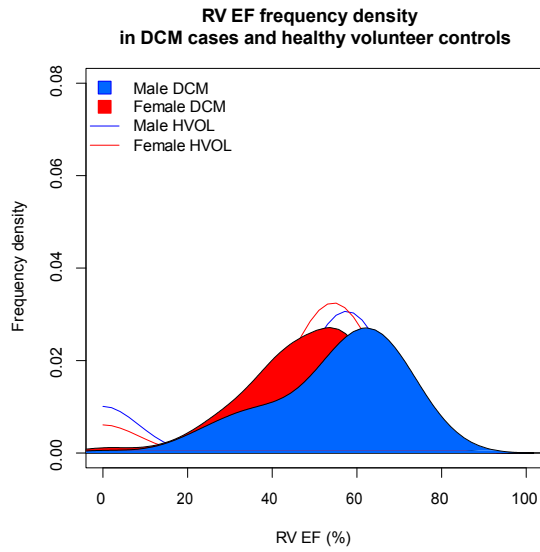
	DCM		Healthy volunteer		Confidence intervals	Difference in location	P value
	Mean	sd	Mean	sd			
LV EF	39.1	12.9	64.3	5.4	23; 27	25	6.0×10^{-102}
LV EDVi	129.2	39.4	81.9	14.1	-43; -35.9	-39.3	4.1×10^{-78}
LV ESVi	82.0	39.5	29.6	8.4	-46.6; -38.7	-42.5	4.3×10^{-101}
LV SVi	47.1	13.5	52.3	7.9	3.24; 6.82	5.0	4.1×10^{-08}
RV EF	51.7	14.3	46.2	22.1	-2; 2	5.6×10^{-05}	0.912
RV EDVi	88.0	24.8	77.0	39.1	-4.66; 3.33	-0.5	0.777
RV ESVi	44.4	22.1	34.2	18.9	-8.83; -2.69	-5.6	0.0001
RV SVi	43.7	13.3	42.6	21.8	1.68; 6.17	3.9	0.0007
LVMi	92.1	27.6	61.6	14.0	-30.4; -24.5	-27.4	9.8×10^{-58}
Max WTi	10.4	2.2					
Septal WTi	8.3	1.9					
Lateral WTi	6.1	1.6					
LAVi	62.2	25.8					
Age at diagnosis	52.1	13.9					
Age at scan	53.7	13.6					

Figure 4.3 Frequency density plots of CMR parameters for males and females in the DCM and healthy volunteer cohorts alongside probability density histograms with raw and smoothed density curves for the DCM cohort

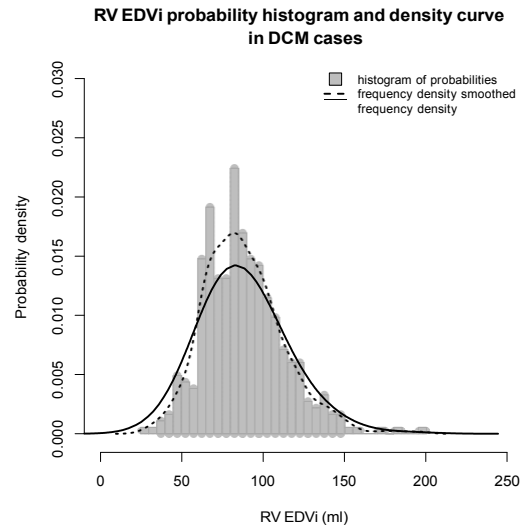
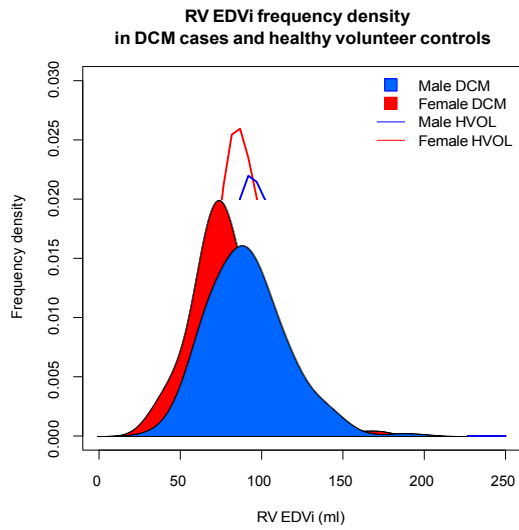
(probability density histograms are not stratified by gender)



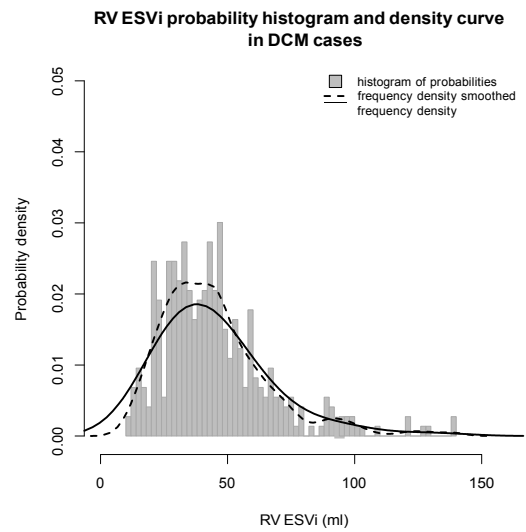
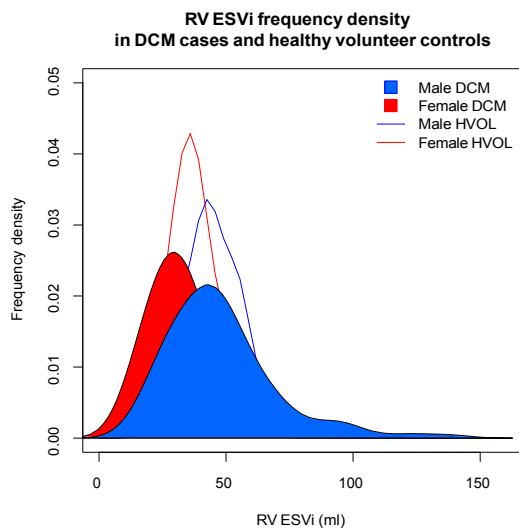
d)



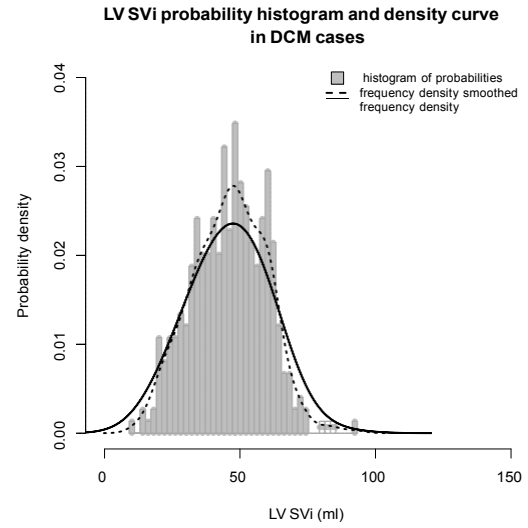
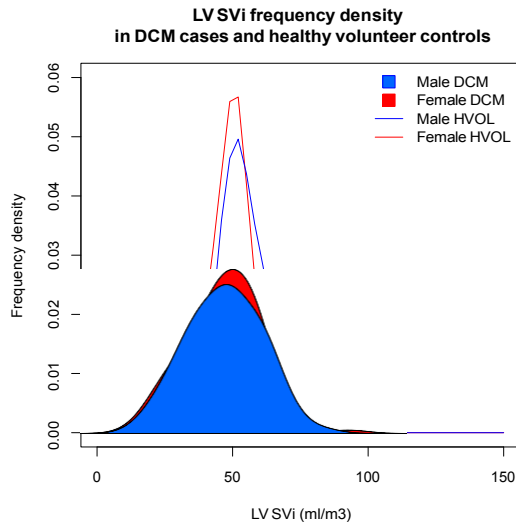
e)



f)



g)



h)

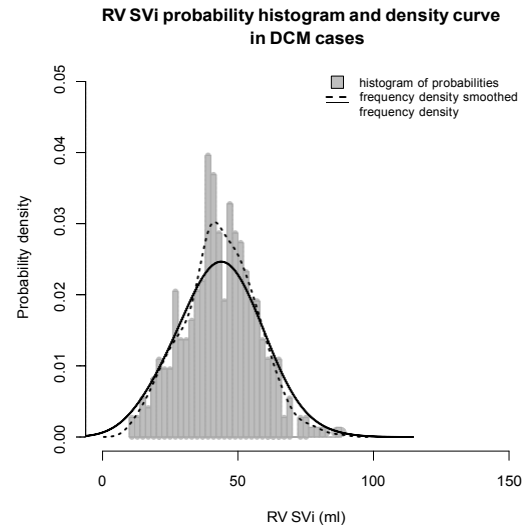
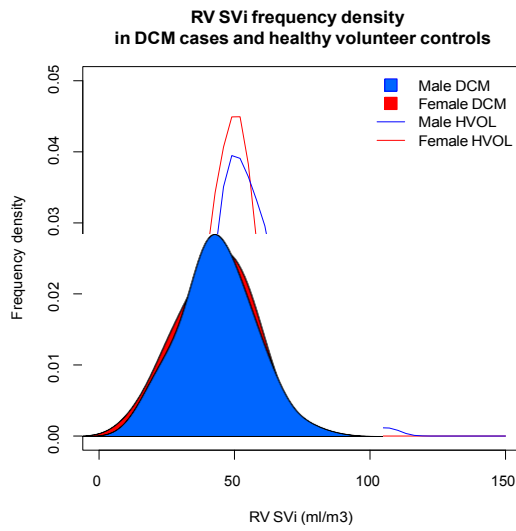


Figure 4.4 Frequency density plots of CMR parameters for males and females in the DCM cohort alongside probability density histograms with raw and smoothed density curves

(probability density histograms are not stratified by gender)

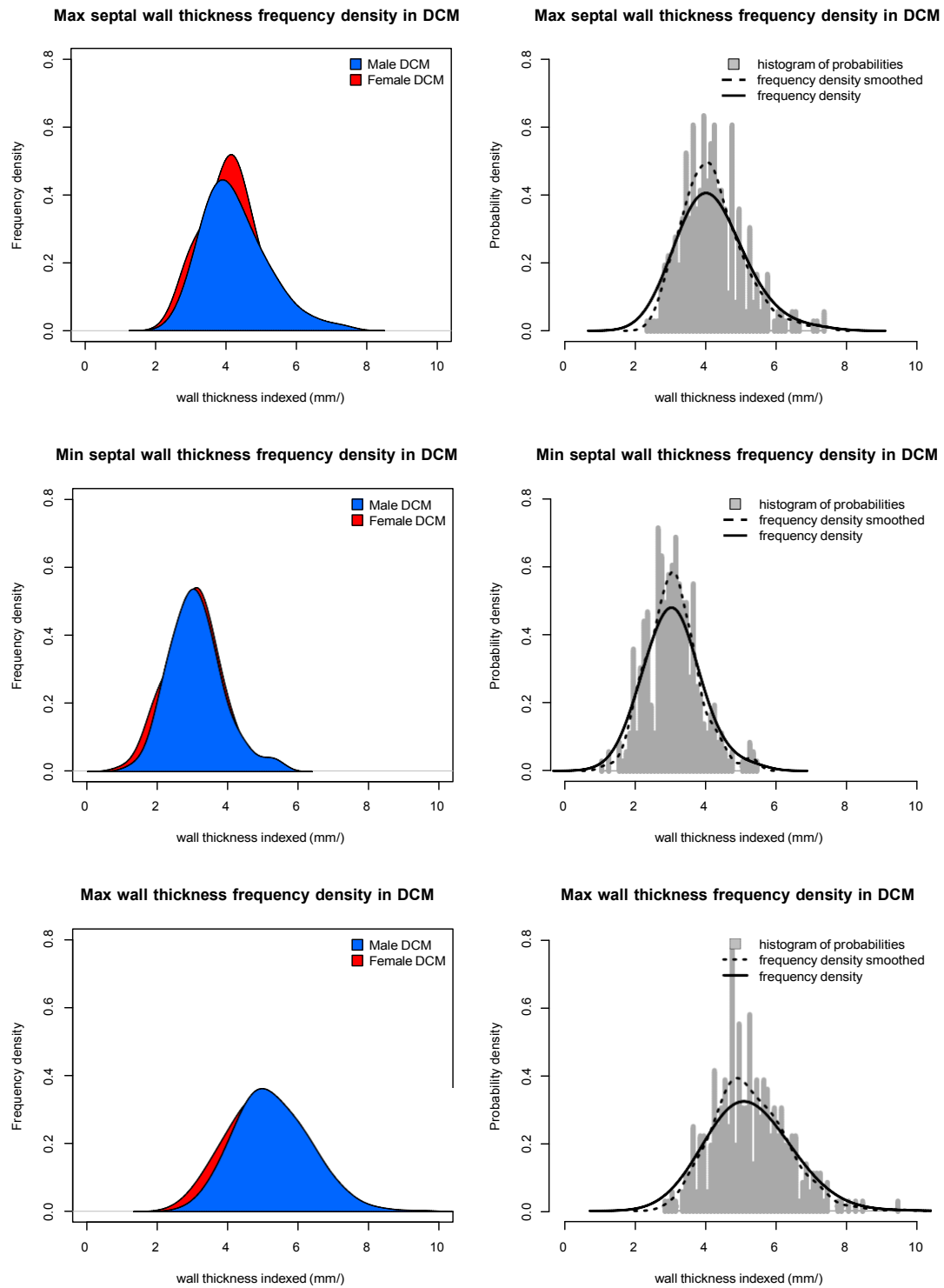
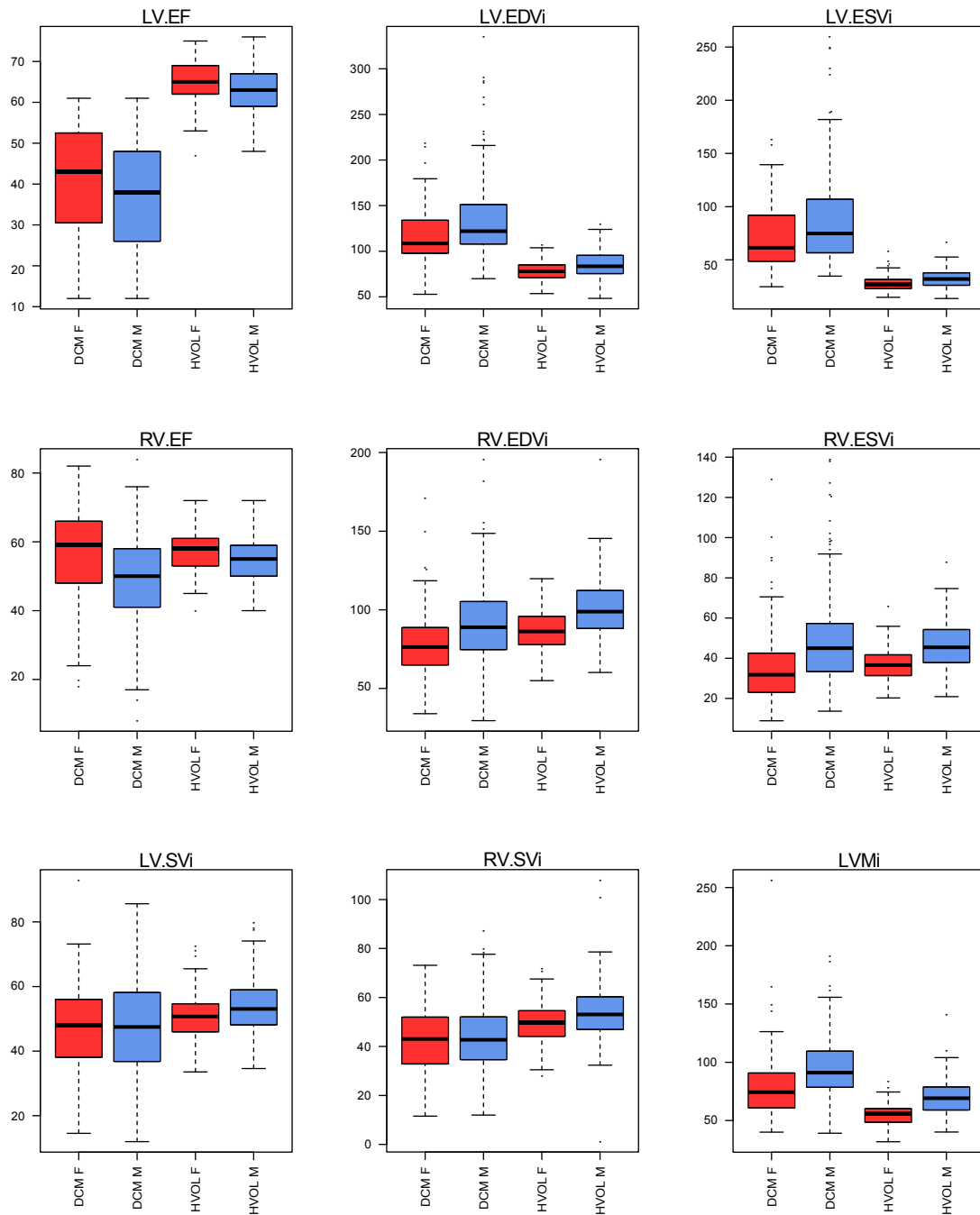


Figure 4.5 Overview comparison of continuous CMR parameters between DCM and healthy volunteer cohorts stratified by gender.

DCM= dilated cardiomyopathy, HVOL= healthy volunteers, M=male, F=female



4.4 Discussion

4.4.1 Demographics

This DCM cohort was prospectively recruited from consecutive referrals to the RBHT CMR unit and is representative of the typical through-the-door patients seen in a CMR unit in South East England. As such, the majority of patients are Caucasian (88%) and male (72%). The frequency of midwall late gadolinium enhancement (32%) is consistent with previous studies in DCM¹⁷⁴. The quantification of valve regurgitation is a strength of CMR because of its capability of measuring accurate ventricular stroke volumes and comparing this with the measurement of great vessel flow from velocity mapping. In isolated valve regurgitation, the regurgitant volume and fraction are simply calculated from the difference in LV and RV stroke volumes, thus isolated mitral regurgitation results in an excess LV stroke volume. Despite a general consensus and the wealth of publications stating that mitral valve disease and ventricular tachycardias are well recognised complications of DCM, there is a paucity of publications quoting their prevalence in DCM^{174,201-203}. In this cohort 19% of patients had moderate to severe MR as determined by CMR, slightly fewer than previously published estimates using ECHO (24%)²⁰⁴ although this might be expected due to the higher accuracy of CMR. In this cohort there was sufficient data to make a determination regarding sustained VT at, or prior to, the time of diagnosis in 131 (35%) patients. Thirty patients (22%) had sustained VT documented electrographically (standard or ambulatory ECG). In 101 patients there was sufficient evidence to exclude sustained VT (absence of sustained VT on ambulatory ECG or clear documentation of absence of suggestive symptoms i.e. palpitations, dizziness, syncope/pre-syncope in conjunction with absence of VT on standard ECG), but in 244 cases (65%) ambulatory ECG monitoring was not performed and/or documentation of symptoms prior to presentation was not sufficient to exclude sustained VT. The prevalence of sustained VT in this cohort (23%) is higher than previously published (3% sustained VT requiring hospitalisation and 4% SCD)¹⁷⁴ and this may be due to a negative bias in ascertainment of cases without sustained VT, i.e. symptomatic or high risk patients are more likely to undergo ambulatory ECG monitoring, and there is likely a bias toward failure to document negative findings e.g. absence of suggestive symptoms. However, since I was blind to TTN genotype while collecting this data any inflation in the prevalence of sustained VT in the cohort as a whole should not influence the validity of genotype-phenotype correlation.

Due to the difficulty in determining whether a particular risk factor for DCM is the primary cause in any given case, or merely a coincident or contributory factor, patients with a history of alcohol excess (14.1%), chemotherapy (5%), iron overload (1.6%), myocarditis (4.7%), small (bystander) myocardial infarction (6.5%) or peripartum presentation (2.4%) were not excluded. If these risk factors were indeed the cause of DCM in these patients it would contaminate the homogeneity of the idiopathic DCM phenotype in this cohort and could reduce the power of this study to detect genotype-phenotype correlation. As an example of the problems associated with determining causality in an individual, excessive alcohol intake is reported in a wide range (3–40%) of DCM patients, but data regarding the incidence, prevalence, mortality and morbidity rates related to alcoholic cardiomyopathy are lacking. Environmental factors (e.g. cobalt, arsenic) and genetic predisposition (e.g. HLA-B8, alcohol dehydrogenase alleles) have been proposed as modifiers in the development of alcoholic cardiomyopathy. Further, it is difficult to establish a definite causal relationship between heavy alcohol consumption and heart failure, given the beneficial effects seen with low to moderate to levels of consumption and the fact that many heavy alcohol users never develop overt heart failure. Ultimately there exist no unique identifying features, clinical or laboratory (including histological changes), that set alcoholic cardiomyopathy apart from other causes of heart failure and the diagnosis remains one of exclusion^{205,206}. In view of the probabilistic nature of assigning a cause to DCM, and since the presence of a relative risk factor does not preclude a genetic factor as the primary cause, cases were included where there is lack of certainty regarding aetiology. In this way the cohort also remains representative of DCM patients typically seen in clinic. In contrast, patients with evidence of active myocarditis (endomyocardial biopsy or CMR) or ischaemic heart disease deemed sufficient to account for severity of DCM (large myocardial infarction on CMR) were excluded.

4.4.2 CMR parameters of morphology and function

A small number of DCM patients and healthy volunteers had outlying measurements and these scans were reviewed. Outlying indexed measurements in DCM patients were found to be due to particularly high or low BSA, and extremes of age due to the difference in normalised mean LVEDVi and LVEF. Outlying indexed measurements in the healthy volunteers were all in young adult males with a low BSA. It is

recognized that indexing by BSA has reduced utility at extremes of BSA, and that using lean body mass, at least in males, is more precise²⁰⁷. However, lean body mass was only available for the healthy volunteer cohort and is not routinely measured or recorded for patients. After a review of the literature I decided to use to BSA as this is superior to height, height squared, and weight and is the most commonly used and clinically relevant index²⁰⁸.

The distribution of the measurements for the DCM cohort was robustly assessed in R, using a Wilcoxon rank sum test with continuity correction, and visualized by the probability density histogram and frequency curves. One issue that was considered was the question of outliers. On one hand, it is best to be inclusive of as much data as possible to increase power. On the other it is important to identify outliers that may be in some way unrepresentative of the cohort overall. In this case each outlier in turn was examined and the majority were retained. A second issue is that of data transformation. This is typically used in order to normalise distribution and thus improve statistical analysis. However, using non-transformed data keeps analysis focussed on the originally measured parameter. The data presented here, without transformation, was judged compatible with downstream genotype-phenotype analysis using robust methods in spite of the presence of outlier cases.

4.4.3 Limitations of relying on CMR

Amongst limitations of CMR, metallic implants and intracardiac devices (pacemaker, ICD) represent the only absolute contra-indication²⁰⁹. Although gadolinium contrast agents are relatively safe, their use is restricted in chronic renal failure (GFR <30 ml/min) due to the small but real risk of nephrogenic systemic fibrosis, which is increased by concurrent heart failure²¹⁰. Severe claustrophobia, clinical instability and the first trimester of pregnancy are among relative contra-indications and a CMR scanner weight limit of 100kg is applied. CMR remains somewhat limited by arrhythmias interfering with ECG triggering and by a patients poor ability to perform breath-holds, both of which are not infrequent in a patient with advanced heart failure. Although the study was designed to prospectively recruit DCM cases without selection bias, this effort may have been undermined by the inherent limitations of CMR. The limitations for intracardiac devices and arrhythmias may have reduced the proportion of patients with arrhythmogenic DCM, including LMNA associated DCM

and thus may have introduced occult genetic stratification of the cohort ²¹¹. The limitation for clinically unstable patients, breatholding and arrhythmias may have reduced the proportion of patients with advanced or decompensated disease. Not all patients with DCM are referred for CMR assessment. Recruiting patients from the CMR unit may have introduced bias for severe yet stable DCM, or for unusual cases.

Although detailed CMR phenotyping was performed for all DCM patients on a research basis and in a standardised manor, all other clinical data was collected from patient records. As a result there is a frustrating amount of missing clinical data, either because tests were not performed (in particular ambulatory ECG recording ¹⁹⁵) or due to difficulty obtaining clinical records from GPs and specialists in other hospitals.

4.5 Summary

This DCM cohort represents prospective recruitment over several years in a large, specialist CMR unit. The depth and breadth of phenotype data coupled with high throughput sequencing of biobanked DNA samples in such a large cohort presents an unrivalled resource for genotype-phenotype associations. Although not entirely homogenous in terms of ethnicity and potential contributing risk factors, the DCM cohort is representative of the type of patient seen in a typical CMR unit in South East England and as such, the results of this study can be applied directly to such a clinic. The variance for continuous parameters is higher in DCM than in healthy volunteers. This broad distribution of CMR measurements reflects the fact that the cohort represents the full spectrum of disease from mild to severe. As expected, lower contractility and larger volumes are observed in males compared to females for both the DCM and healthy volunteer cohorts. Deviation from normal distribution in the DCM cohort, likely due to the diagnostic criteria, is seen for EF, EDVi and ESVi.

5 Design and optimisation of the Titin sequencing assay

5.1 Introduction

Several factors could be expected to influence overall sequencing performance. At the time of commencing this thesis we had early access to SureSelect (Agilent) target enrichment compatible with the SOLiD sequencing platform. Throughout we have continued to work collaboratively with Agilent using early access products including their assay re-design service. The scope of research into TTN genetics had been restricted by size of the TTN gene and its extensive alternative splicing events until the advent of NGS and target enrichment technologies^{17,212,213}. Current data about the genomic co-ordinates of all TTN exons and which of these are expressed in the heart and were thus relevant to cardiac biology were not available. As such development of a comprehensive and consistently high performance assay involved concurrent troubleshooting and investigation on several fronts including target definition, factors affecting hybridisation, effect of sequencing parameters, iterative redesign of the assay, optimisation of library preparation, and efficiency of overall performance.

Performance of a given target enrichment assay is assessed relative to the target genomic region(s) it is designed to capture. The initial assay design (Assay 1) targeted all published exons in the two major cardiac transcripts of TTN (N2BA, ENST00000591111 & N2B, ENST00000460472)¹⁸⁹. During the iterative redesign process we took the opportunity to re-define the assay target following review of the evidence for each TTN exon and generation of a meta-transcript (ENST00000589042) by HAVANA¹⁸³. Evidence of exon usage in the LV, derived de-novo from RNA-sequencing data (see section 3.7 RNA Sequencing) was also incorporated.

Study design, including variant calling parameters, was planned to maximise sensitivity (potentially at the expense of specificity), and variants were then validated by Sanger sequencing. The technical sensitivity of capture-based NGS approaches is very high at good sequencing depth (>99% at 20x read depth²¹⁴) and in this study the major determinant of sensitivity relates to target enrichment efficacy and uniformity. Even with good overall depth of sequencing there are particular regions that are either not enriched successfully or where reads cannot be mapped uniquely to the genome due to repetitive sequence (TTN contains several short exons with a

high degree of homology). Callability describes the percentage (%) of target sequence which is sequenced with sufficient depth and quality such that a heterozygous variant can be identified²¹⁵, and defines the proportion of target sequence that can be studied. This is likely the most informative single metric for addressing assay performance and sensitivity of variant calling, and is used here as the primary metric of assay performance. Callability can also be used to confirm equivalent overall sequencing performance between cohorts enabling fair comparison of results.

Coverage is often used interchangeably to describe depth of sequencing (vertical coverage) and the proportion of target sequenced (horizontal coverage)^{216,217}. To avoid any potential confusion the term coverage is not used here. Similarly sequence capture (and many other terms) are used interchangeably with target enrichment in the published literature²¹⁷⁻²¹⁹. Here target enrichment refers to the enrichment of target DNA relative to genomic DNA during NGS library preparation. Sequence capture is used to describe the usable sequence data obtained following NGS and is a function of both target enrichment, NGS performance and mapping and QC of sequencing reads.

5.2 Aims

- Definition of the genomic regions encoding the TTN gene
- Identification and optimisation of factors affecting target enrichment and overall sequencing performance
- Development of a final comprehensive and consistently high performance assay.
- To gather high quality sequence data for TTN in our DCM cohort

5.3 Results

5.3.1 Definition of the target region and development of a TTN LRG

As discussed in section 1.7 of the introduction (Titin structure and function in the cardiomyocyte), TTN undergoes extensive alternative splicing, including 15 splice variants recorded in Ensembl^{220,221}. Until recently, variant reporting in coding coordinates was difficult, as it required the use of several transcripts to obtain

complete exon annotation¹⁷. It was also unclear which exons were truly included in each transcript and were expressed in the heart^{17 212}.

To facilitate standardised variant annotation in accordance with international guidelines we developed a Locus Reference Genomic sequence (LRG)¹⁸⁶ for *TTN* (<http://www.lrg-sequence.org>, LRG_391). LRG records contain internationally recognized stable reference genomic DNA sequence for a genomic region, essential transcripts and proteins for variant reporting and an exon numbering system.

Together with the European Bioinformatic Institute (EBI, www.ebi.ac.uk) we developed an inferred complete meta-transcript (LRG_391_t1) manually curated by the HAVANA group that incorporates all *TTN* exons, with the exception of a single alternative terminal exon unique to the shorter novex-3 isoform¹⁸⁹ (Figure 3.2 A unified representation of TTN transcripts and tables 3.1 Reference transcript and protein identifiers for Titin and 3.2 Overview of TTN transcripts and exon usage). A second LRG transcript, LRG_391_t2, novex3 isoform (NM_133379.4), was also included to allow reporting of variants within the alternate terminal exon²¹³.

5.3.2 GC ratio affects hybridization efficiency

Variability in assay performance is a function of efficiency of hybridisation which is known to be sensitive to factors such as GC content¹⁸¹. A previous study has shown that a distinctive feature of the SureSelect system is that it outperforms other hybridisation-based enrichment systems in regions of low GC content²²². We have found that the SureSelect system performs relatively poorly in regions of high GC content, particularly in the first exon of many genes²²³. The poorly captured region in *TTN* has a higher GC ratio than the rest of the gene (Figure 5.1) suggesting that this may be the underlying cause of poor performance in this region. However when callability is plotted against GC ratio it can be seen that there is no correlation in this instance (Figure 5.2). Indeed both the *TTN* exons with the highest and lowest GC ratio are captured effectively.

Figure 5.1 Callability and GC content of TTN exons

Points in the upper plot show the percentage of each exon sequenced with sufficient depth and quality to call variants, calculated as a mean across samples from the DCM cohort. Points in the lower plot show the GC ratio for each exon. Exon numbering is according to LRG_391_t1. The first, GC rich, exon is captured well but a short section between exon 165 and 195 is captured poorly.

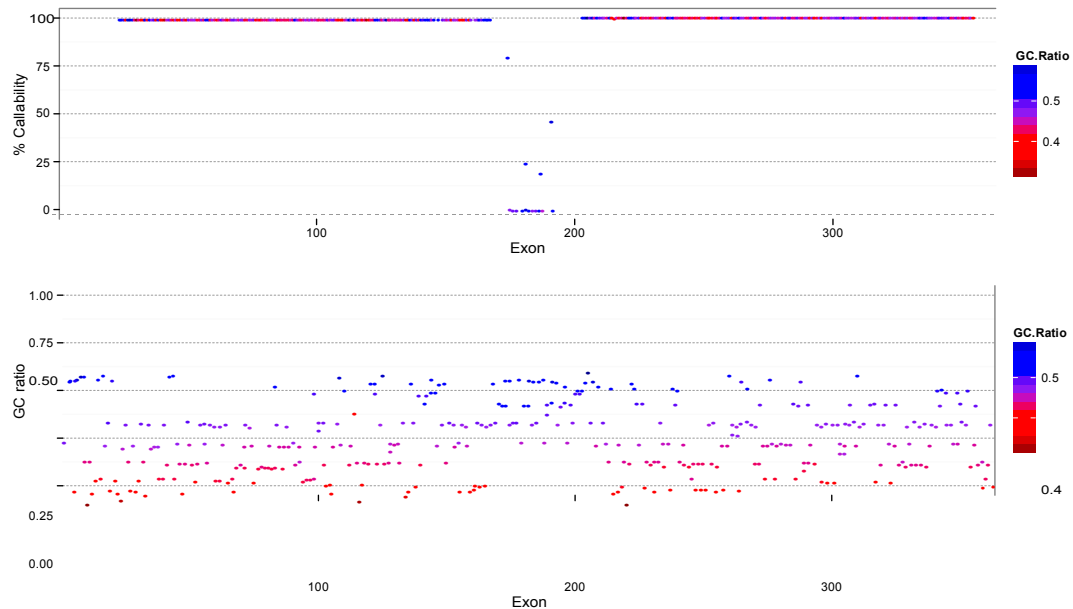
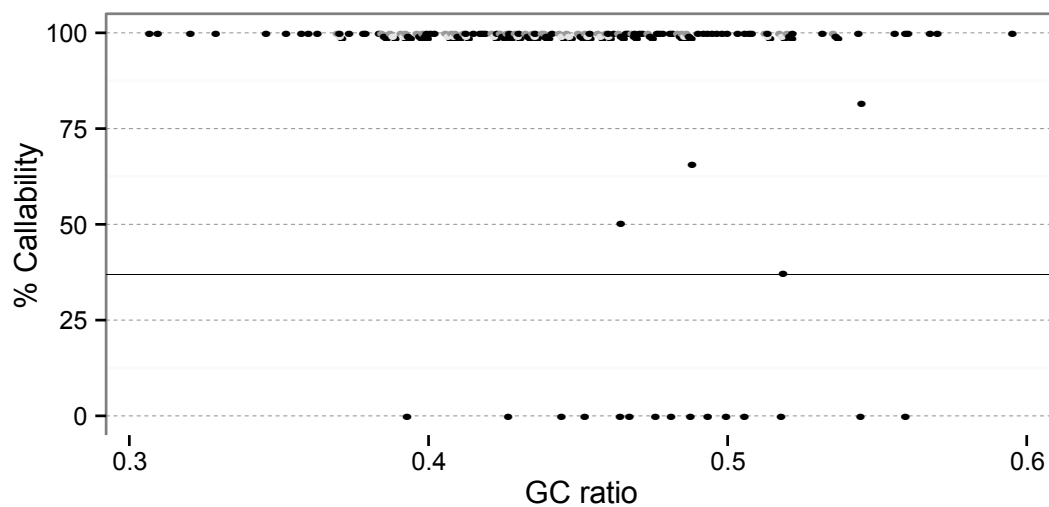


Figure 5.2 Relationship between GC content and callability for TTN exons

The percentage callable for each exon calculated as a mean across samples is plotted against the GC ratio for that exon. Each point represents an exon in LRG_391_t1. Most exons with $GC > 0.5$ are captured well.

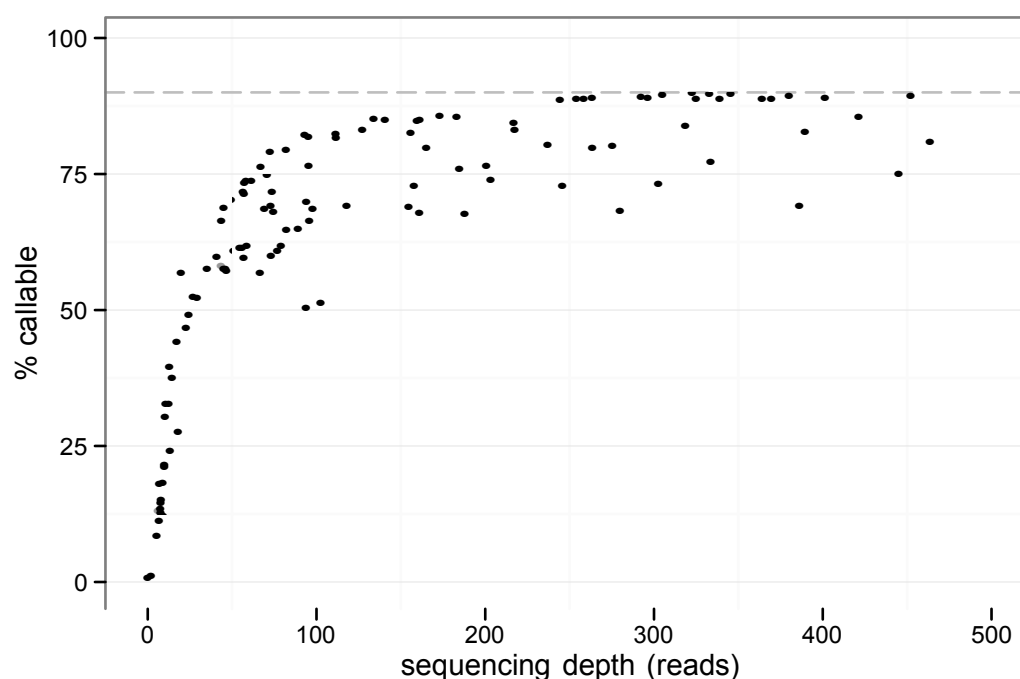


5.3.3 Sequencing depth affects overall assay performance

Irrespective of target enrichment design, the percentage of bases callable increases with increasing sequencing depth (Figure 5.3) illustrating the relationship between read depth and capability at a global level. The percentage callable in early runs which had a lower mean depth of sequencing was strongly dependent on the sequencing depth of each sample. At a mean read depth < x100, increase in read depth greatly increases callability, after which the effect begins to plateau. At an average read depth of approximately x300 the callability reaches saturation (approximately 95% callable) after which increasing read depth does not improve callability. The fact that the maximum callability is less than 100% may be explained by factors such as high GC content leading to poor hybridisation efficiency or inadequate tiling leading to ineffective capture of particular genomic sequences, and increasing mean read depth will not resolve these problems²²³ (see below).

Figure 5.3 Relationship between sequencing depth and callability for target enrichment assays

Samples plotted here are from a variety of pilot studies, using several different SureSelect assay designs including both off the shelf whole exome and smaller custom assays. For each sample the percentage of bases that are callable across the complete target is plotted against the mean sequencing depth achieved for that sample (reads on target are normalised for target size). The dashed line represents the maximum % callable achieved for any sample, as the callability seems to have approached saturation.



5.3.4 Assay design optimisation

Using Assay 1 the median callability achieved for TTN at optimal sequencing depth was 97%. Callability for other genes in the assay was variable. It has been shown previously that SureSelect target enrichment has uneven performance: enrichment, and thus read depth, is variable across a given design or target²²³. One solution to improve assay performance is to increase the mean depth of sequencing per sample, in an effort to push the depth of sequencing of borderline regions above variant calling threshold. However this is inefficient as increasing overall sequencing depth will also increase sequencing of regions that are well captured, consequently it is preferable to redistribute read depth across the target through iterative re-design.

Some regions of the assay target were captured consistently poorly despite optimal average sequencing depth (Figure 5.4). A minority of regions were excluded from the initial design due to strict eArray bait design parameters such as repeat masking (<https://earray.chem.agilent.com/earray>). Although not quantifiable, failure in bait design during production may also have occurred. In collaboration with Agilent we undertook iterative redesign of the whole assay to redistribute sequence depth together with specific optimisation of TTN target enrichment by removal of repeat masking and manual optimisation of bait design for poorly performing regions. Agilent have been working on algorithms for this, increasing the number of baits in poorly performing regions, and reducing the number in regions that capture well. This can be achieved by either altering the number of copies of each bait, or by increasing the number of distinct baits. The exact details of the algorithm have not been shared with us.

During the assay redesign we took the opportunity to include additional TTN exons based on the newly developed LRG (see above)²¹³, and thus the target for Assay 1 differs slightly to all other designs. We provided Agilent with the same target criteria as for Assay 1 (100bp overhang into exons, plus upstream sequence), albeit with additional TTN exons, along with aligned sequencing reads from our first sequencing runs, and received a “rebalanced” library design.

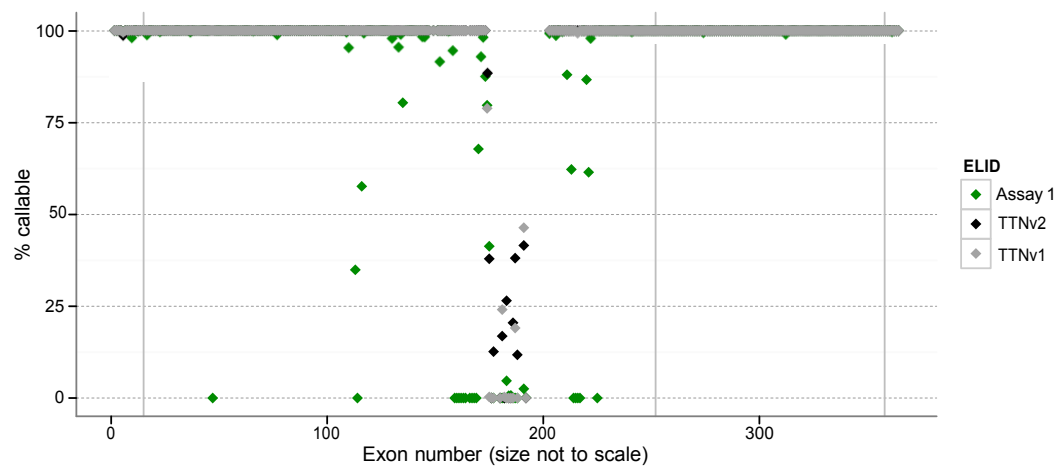
Two single gene iterations were trialed for TTN, assay designs TTNv1 and TTNv2, which differed from each other in that repeat masking was used for TTNv1 but not TTNv2, and as such TTNv2 contained more baits (target size 271,902bp vs. 281,431bp). Callability of some TTN exons was improved by assay redesign, but a

small region of 30 short exons with a high degree of sequence homology (exon 165-195) remains poorly captured (Figure 5.4). Both the small size and high homology of these exons may affect both the capture and eventual mapping of sequencing reads.

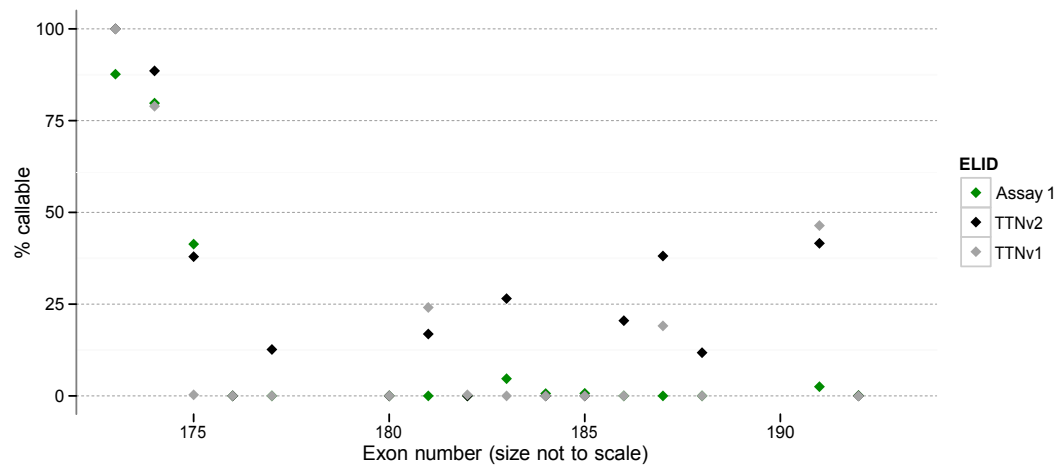
Figure 5.4 Callability per exon for iterative assay designs

Points for each assay show the percentage of each exon sequenced with sufficient depth and quality to call variants, calculated as a mean across samples. Exon numbering is according to LRG_391. Vertical grey lines mark the domain boundaries between Z-disc, I-band, A-band, and M-band (from left to right). Points denote exon number and are not representative of exon size. Figure 5.4 a) shows all coding exons (2-364). Figure 5.4 b) shows just the region from within the I-band (exons 173-193) where TTNv1 and TTNv2 perform differently.

a)



b)



Although both TTNv1 and v2 outperformed Assay 1, TTNv2 slightly outperformed TTNv1: Both assays failed to capture the same four whole exons, and v1 failed to capture one further exon. Assay TTNv1 partially captured 12 exons and assay v2 partially captured 10 exons (Table 5.1). That callability of exons 175-198 remains poor despite assay redesign and optimal sequencing depth suggests factors other than assay design influence overall sequencing performance of this region.

Table 5.1 Exons not fully captured by TTN single gene assays TTNv1 and TTNv2

The number of bases per exon which are callable, sequenced at low read depth insufficient to call variants (low capture), and with no sequence capture mapped (not covered) are shown for each of the assays.

exon	TTN v1			TTN v2		
	Callable	Low capture	Not covered	Callable	Low capture	Not covered
175	0	64	45	89	0	0
176	0	35	68	0	80	23
177	0	0	103	0	30	73
180	0	53	47	0	41	59
181	0	32	71	0	68	35
182	0	65	38	0	0	103
183	0	33	70	15	51	37
184	0	13	96	0	3	106
185	0	0	103	0	0	103
186	0	0	103	0	5	98
187	0	0	100	22	53	25
188	0	56	44	50	26	24
191	40	21	42	41	18	44
192	0	24	29	0	0	53
197	0	0	54	0	0	54
198	25	42	33	80	0	0
216	87	7	6	93	5	2

The final optimized design, Assay 2, comprised optimisation of Assay 1 plus TTNv2 and performs consistently well. Assay 1 was designed before the TTN meta-transcript was developed and the LRG was created. Thus baits were not designed for some of the exons and performance of Assay 1 assessed by callability with regard to the LRG meta-transcript is consistently inferior to other assays. In order to bring meta-transcript target capture of samples already sequenced using Assay 1 in line with subsequent samples sequenced using Assay 2, a 'Top-up Assay' was produced. The design of the Top-up Assay was simply those additional baits by which Assay 2 differed from Assay 1. New libraries were prepared and sequenced. Sequencing reads for Assay 1 and the Top-up Assay were combined ('Assay 1 & Top-up') prior to mapping to the genome and callability achieved using this approach is comparable to Assay 2 (Figure 5.5) for most samples.

5.3.5 Library preparation and NGS performance

There are several ways in which library preparation and sequencing can affect overall performance. The most simple is differences in design of the assay used for library preparation, as discussed above. However a few of the samples sequenced with Assay 1 & Top-up still showed poor callability (<90%, Figure 5.5).

Libraries were prepared and sequenced in batches. One possible explanation for the small but distinct group of samples performing poorly is a batch effect in library preparation or sequencing. However this does not appear to be the case: there are a few samples in each run which perform poorly (Figure 5.6).

Libraries were barcoded and pooled for sequencing, thus if an individual sample was underrepresented in the pool it would have been sequenced to a shallower depth and this may adversely affect performance as discussed above. The depth of sequencing for individual libraries ranged from approximately 20x to >400x. Assay 1 was sequenced at less than 50x, which together with the assay design may explain poor performance. Some Assay 1 & Top-up libraries performed poorly despite optimal sequencing depth, thus sequencing depth does not appear to be the cause of their poor performance (Figure 5.7).

Figure 5.5 Beeswarm plot of callability for each assay design measured against the meta-transcript LRG_391_t1.

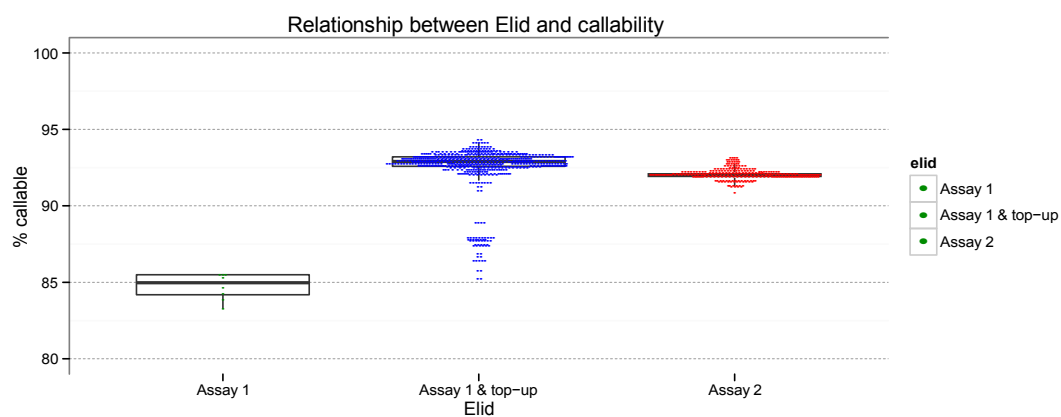


Figure 5.6 Beeswarm plot of callability stratified by sequencing run, measured against the meta-transcript LRG_391_t1.

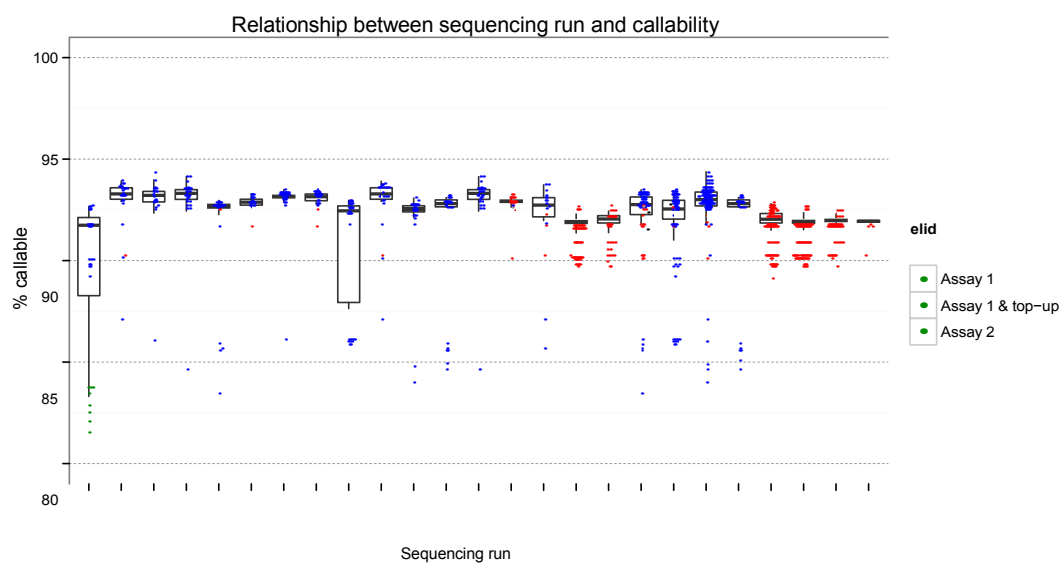
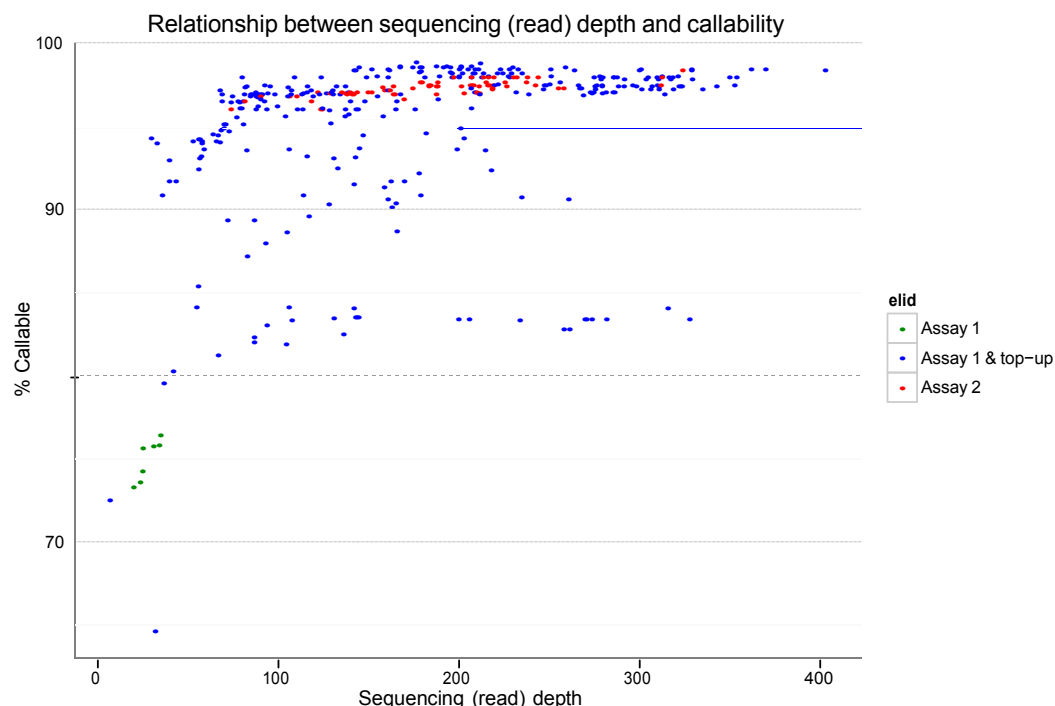


Figure 5.7 Relationship between sequencing depth and callability for TTN



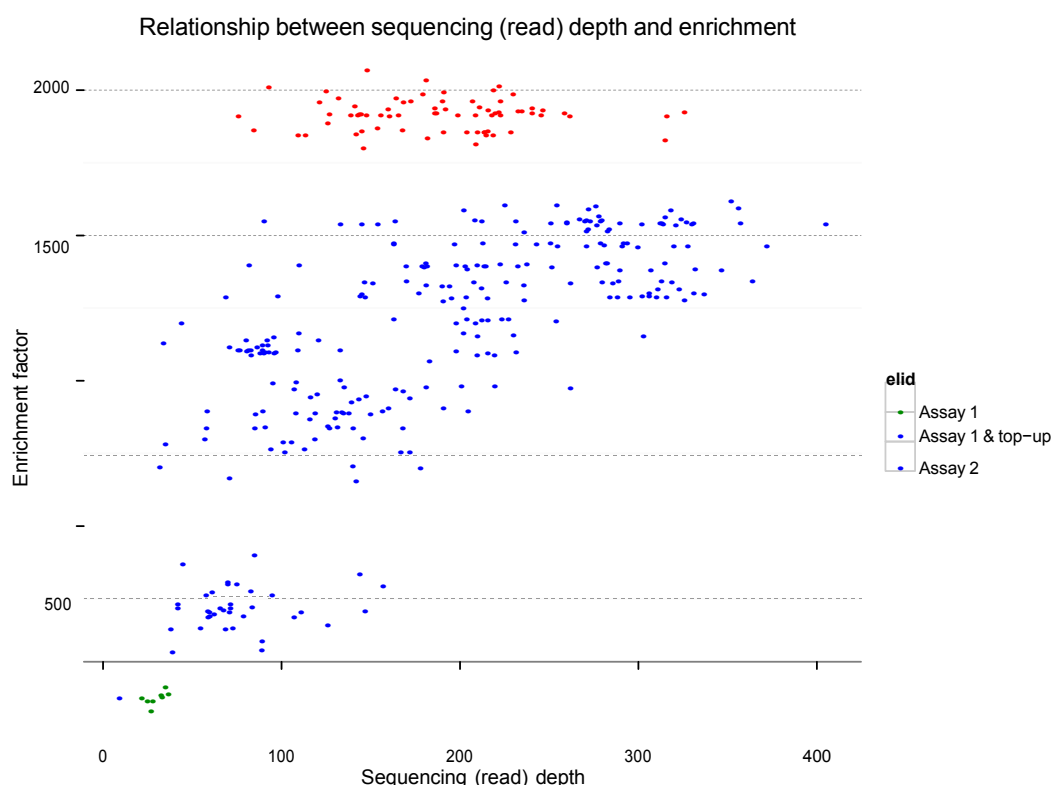
5.3.6 Improving sequencing efficiency

Target enrichment increases the amount of target DNA relative to genomic DNA and thus reduces the sequencing capacity per sample required to obtain usable sequence data for a region of interest. The efficacy of an assay, the target enrichment factor, can be calculated as

$$\text{Enrichment Factor} = \frac{\text{Reads on Target} / \text{Total mapped reads}}{\text{Target size} / \text{Genome size}}$$

Although Assay 1 & Top-up perform comparably to Assay 2 for most samples, the enrichment factor for Assay 2 is far greater (Figure 5.8). Thus libraries prepared with Assay 2 should reach callability saturation at a lower depth of sequencing, however this remains to be determined as all samples enriched using Assay 2 were in fact sequenced to a high depth: min 95.9x, max 98.0x, median 97.2x, mean 97.2x, sd 0.47.

Figure 5.8 Relationship between sequencing depth and enrichment for TTN



5.4 Discussion

Since the first sequencing runs using SureSelect target enrichment sequencing capacity has increased six-fold (from 25Gb of single end reads on the SOLiDv3 to 150Gb of paired end reads (75bp forwards, 35bp reverse) on the SOLiD5500). Increased read length and paired-end reads have increased the proportion of sequence data that maps uniquely to the genome and is thus usable, further increasing practical capacity

Iterative assay redesign has improved capture of some TTN exons and allowed refinement of the target region to reflect the TTN meta-transcript. Manual rebalance and reduction of baits in the regions with highest sequence depth improves evenness of assay performance²²³ and should reduce the saturation threshold for sequencing depth, however this remains to be determined. Given our understanding of the importance of adequate sequencing and our increase in capacity, all samples are sequenced to well above the saturation threshold.

Despite Assay redesign a few TTN exons remain poorly captured. These exons are clustered in a region of high CG content, however there are exons with a much higher GC content that are captured well, suggesting there is some other sequence

factor or tertiary structure at play. The exons in this region are short (81-90bp) and highly homologous, possibly affecting mapping of reads uniquely to the genome. This assay is designed to target exons with a 100bp overhang into introns, and previous studies have shown this improves capture of smaller exons¹⁸¹. Future increases in read length may facilitate unique mapping of these exons and thus improve callability. Indeed, recent sequencing runs of an unrelated cohort enriched using Assay 2 and sequenced on the Illumina HiSeq platform which has longer reads (150bp forward and reverse) had 100% callability. Although the effect of increased read length cannot be dissected from other differences between the Illumina and Life Technologies NGS platforms which use different chemistries, the improved capture on a longer read platform supports unique read mapping as the factor limiting capture in this study.

Some samples sequenced using Assay 1 plus Top-up performed poorly. No batch effects in library preparation, pooling or sequencing were identified, but it may be an effect of the amount or quality of DNA used in the library preparation of those individual samples. All libraries prepared with Assay 2 performed well, as these were prepared and sequenced most recently the consistency in performance may reflect our increased experience in using SureSelect and troubleshooting of the protocol for library preparation on the liquid handling platform (Bravo, Agilent).

5.5 Summary

In summary, optimisation of an effective TTN sequencing assay involved careful definition of the target region through review of TTN exon usage and establishment of a meta-transcript and LRG; optimisation of the target capture assay through iterative design and testing; resolving poorly covered regions both through manual bait design and by removing repeat masking features in the design process; and finally taking into consideration potential batch effects. The net result has been the development of an assay which performs well and provides reliable sequence capture (> 97%) across the TTN gene. There remains a region of 30 short exons (representing approximately 5% of the total coding sequence) which was captured poorly by this assay. Analysis of DNA sequence including GC content suggests this is more likely a problem of mapping than of target enrichment as is supported by results of recent sequencing on a longer read NGS platform that achieve 100% callability. It is possible that TTNtv are under-detected in the region of poor callability, particularly in

the small number of samples sequenced using Assay 1 only. The impact of this is discussed in detail in chapter 6.

6 Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease

6.1 Introduction

Diagnostic sequencing of *TTN* will only be useful if causality can be confidently ascribed to individual variants. *TTN*tv cause severe and familial DCM¹⁷ with age-dependent and incomplete penetrance, features that have hindered definition of the spectrum of clinical manifestations and insights into disease mechanisms.²²⁴ The significance of *TTN*tv found in the general population or in less severe, and often sporadic, DCM remains uncharacterised. Additionally, it is unknown whether molecular properties of *TTN*tv can distinguish benign from pathogenic variation, or indicate the degree of risk for DCM. To address these issues *TTN* was sequenced in five cohorts spanning the full spectrum cardiac physiology: healthy volunteers with full cardiovascular evaluation (n=308), community-based cohorts with longitudinal clinical data (3,603 participants in the Framingham (FHS¹⁷⁷) and Jackson (JHS¹⁷⁸) Heart Studies), prospectively-enrolled unselected DCM patients (n=374) and end-stage DCM patients considered for transplantation (n=155).

6.2 Aims

- Compare sequencing capture between cohorts to assess validity of comparison
- Explore burden and distribution of variants between cohorts to estimate likelihood of pathogenicity and determine factors that distinguish pathogenic variants
- Explore relationship between *TTN* genotype and DCM phenotype to determine the potential for prognostication and personalised management

6.3 Results

6.3.1 Comparison of sequencing capture between cohorts to assess validity of comparison

Prior to comparing the variants found in each cohort it is important to confirm that no error is introduced by the differences in sequencing and variant analysis pipelines for each cohort (see general methods). The sequencing capture for all cohorts was compared and although there were minor differences, these were not significant

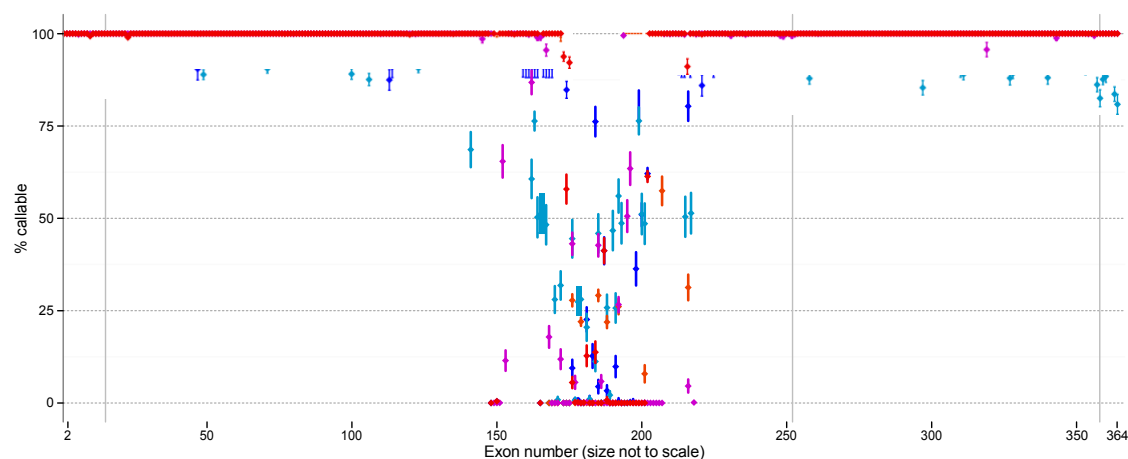
(Figure 6.1). It was not possible to assess capture of TTN in end-stage DCM samples sequenced by Complete Genomics (n=54), however no difference in the frequency of TTNtv was observed in end-stage cohort samples sequenced using targeted capture (20/101, 19.8%) as compared to whole genome sequencing (10/54, 18.5%).

Figure 6.1 TTN sequencing capture for each cohort

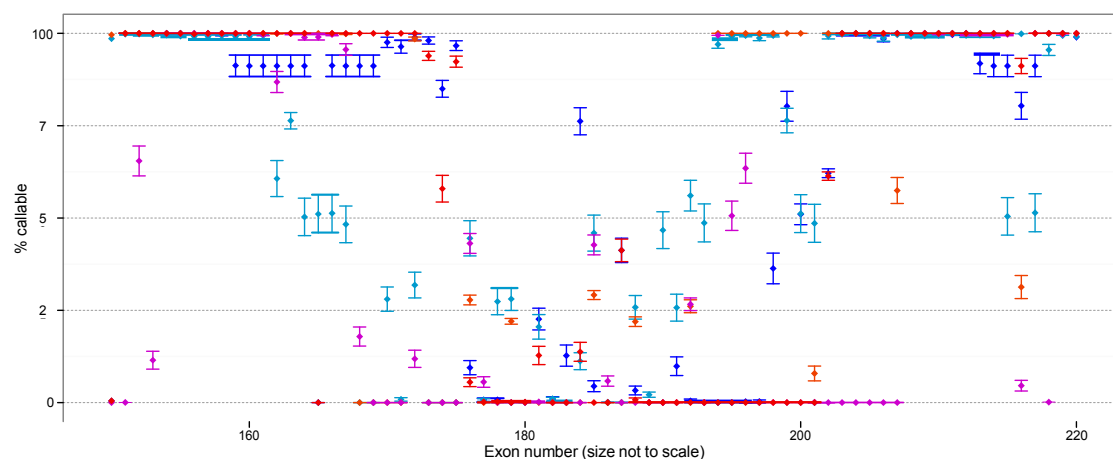
Points for each cohort show the percentage of each exon sequenced with sufficient depth and quality to call variants, calculated as a mean across samples. Error bars show the 95%CI. Exon numbering is according to LRG_391. Vertical grey lines mark the domain boundaries between Z-disc, I-band, A-band, and M-band (from left to right). Points denote exon number and are not representative of exon size. (a) all coding exons (2-364). (b) shows the region within the I-band (exons 150-220; 6315 bp; 2105aa; ~5% of the gene) that exhibited the poorest callability in all cohorts. The exons in this region are small and contain repetitive sequence affecting capture and mapping of reads in all cohorts.

Red = Healthy Volunteers (n=308), Orange = JHS (n=1,980), Magenta = FHS (n=1,623), Dark blue= Unselected DCM (n=374), Pale blue = End-stage DCM (n=155)

a)



b)



6.3.2 Burden of TTNtv in health and disease

Fifty-six TTNtv affecting full-length *TTN* were identified in 3,911 controls (1.4%): of which 9 were found in healthy volunteers (2.9%), 16 in the FHS (1.0%) and 31 in JHS participants (1.6%). Eighty-three DCM patients carried TTNtv (total TTNtv in DCM versus controls, OR=13(9-18), $P = 2.8 \times 10^{-43}$; Table 6.1): 49 in the unselected DCM patients (13%) and 34 in end-stage DCM patients (22%). Comparing variants found in health and disease: nonsense, frameshift, and canonical splice site TTNtv were substantially enriched in DCM patients (78/607 (12.9%), OR = 17(11-25); $P = 1.9 \times 10^{-45}$). Additional variants predicted to alter non-canonical splice signals were also enriched in DCM (OR = 4.2(1.8-9.7); $P = 0.0017$) but often occurred in combination with another TTNtv. A full list of TTNtv and novex-3 truncating variants in each cohort is shown in Tables 6.1-6.5

6.3.3 Distribution of TTNtv in health and disease

The locations of TTNtv were non-uniformly distributed within and between study cohorts (Figure. 6.2). To better understand the effects of individual TTNtv, three interrelated positional factors were explored: location of the variant with respect to the A-band, transcripts affected, and variant exon usage level.

As a previous study of severe and familial DCM showed TTNtv enrichment in the A-band¹⁷, the distribution of TTNtv across the spectrum of health and disease was explored. TTNtv were more commonly located in the A-band in cases than controls (61/87 case variants located in A-band, vs. 21/56 in controls, OR=3.9(1.8-8.3), 1.4×10^{-4}), due both to an enrichment of A-band variants in DCM patients (compared with a uniform distribution: 61/87 variants in A-band, vs. 18235/35991 residues; OR = 2.3(1.4-3.6), $P = 3.4 \times 10^{-4}$), and an opposing trend, towards A-band sparing, in controls (35/56 variants outside A-band, OR=0.58(0.34-1.0), $p = 0.06$). A-band enrichment was most pronounced in end-stage DCM patients (OR = 3.5(1.6-8.5), $P = 7.8 \times 10^{-4}$) and a concordant trend in less severe, sporadic DCM (OR = 1.7(0.97-3.1), $P = 0.07$) (Figure. 6.2 & 6.3). Enrichment of variants in the A-band in DCM was robust to controlling for non-uniform distribution of potential truncation sites along the *TTN* gene (Figure. 6.3).

Table 6.1 *TTN* truncating variants identified in UK prospective DCM cohort

Sample ID	Hg19 genomic start position	LRG genomic start position	Transcript effect	Protein effect	Exon number	Median exon PSI	Present in N2BA	Present in N2B	Present in novex-3	Variant type
Variants affecting meta-transcript (LRG_391_t1)										
10AP01669	179641966	58564	c.4724_4728delTGAAA	p.Met1575SerfsX6	27	1	Y	Y	Y	Frameshift
10RW00786	179635212	65318	c.8307_8308delTG	p.Ala2770HisfsX4	35	1	Y	Y	Y	Frameshift
10DW00512	179605317	95213	c.12643_12644delCA	p.Gln4215ValfsX16	49	0.98	Y	Y		Frameshift
14MW01583	179605203	95327	c.12757C>T	p.Gln4253X	49	0.98	Y	Y		Nonsense
10GG01869	179577042	123488	c.27607G>A	p.Glu9203Lys	96	0.69	Y			Splice variant prediction
10JF01835	179571683	128847	c.29267-2A>C		103	0.82	Y			Canonical splice variant
12SK00375	179571661	128869	c.29062delG	p.Ala9688GlnfsX7	102	0.8	Y			Frameshift
12PB00376	179554624	145906	c.31763-1G>A		123	0.83	Y			Canonical splice variant
10SB00367	179547631	152899	c.33113-1delG		137	0.83	Y			Canonical splice variant
12MD01102	179544980	155550	c.33643+1G>A		142	0.02	Y			Canonical splice variant
12MA01291	179517277	183253	c.39044-9T>A		203	0.06	Y			Splice variant prediction
10RC00179	179500851	199679	c.41447delG	p.Gly13816AlafsX18	227	0.99	Y	Y		Frameshift
10PC00581	179500825	199705	c.41473C>T	p.Arg13825X	227	0.99	Y	Y		Nonsense
10JM01592	179495983	204547	c.43792delG	p.Val14598X	238	0.99	Y	Y		Frameshift
10BC00144	179485581	214950	c.45757_45756dupA	p.Tyr15253IlefsX15	248	0.98	Y	Y		Frameshift

Sample ID	Hg19 genomic start position	LRG genomic start position	Transcript effect	Protein effect	Exon number	Median exon PSI	Present in N2BA	Present in N2B	Present in novex-3	Variant type
10WS00448	179485525	215005	c.45812T>G	p.Leu15271X	248	0.98	Y	Y		Nonsense
10VS00240	179482115	218415	c.47697C>A	p.Cys15899X	255	0.98	Y	Y		Nonsense
10DR00069	179481846	218684	c.48100+1G>A		257	0.98	Y	Y		Canonical splice variant
10CH00929	179480145	220385	c.48527G>A	p.Trp16176X	260	0.98	Y	Y		Nonsense
10JN00724	179477082	223448	c.50170C>T	p.Arg16724X	267	0.98	Y	Y		Nonsense
10LP00987	179477082	223448	c.50170C>T	p.Arg16724X	267	0.98	Y	Y		Nonsense
10RN00513	179477082	223448	c.50170C>T	p.Arg16724X	267	0.98	Y	Y		Nonsense
10LB00067	179474002	226529	c.52036_52035insTT	p.Leu17346PhefsX4	274	0.98	Y	Y		Frameshift
10DH00882	179473511	227020	c.52228_52227dupAGAAA	p.Asp17410ArgfsX25	275	0.98	Y	Y		Frameshift
12AM00400	179470140	230390	c.54106+1G>T		281	0.98	Y	Y		Canonical splice variant
10JS00726	179470136	230394	c.53881+5G>A		280	0.98	Y	Y		Splice variant prediction
10HH00944	179466515	234015	c.55528-1G>A		288	0.98	Y	Y		Canonical splice variant
10JF01881	179466199	234331	c.55525_55531delGACAGGA	p.Asp18509SerfsX29	288	0.98	Y	Y		Frameshift
10SF00976	179466199	234331	c.55525_55531delGACAGGA	p.Asp18509SerfsX29	288	0.98	Y	Y		Frameshift
10RN00513	179462261	238269	c.57544+4A>G		295	100	Y			Splice variant prediction
10AL00811	179458293	242237	c.58957+2T>C		300	0.98	Y	Y		Canonical splice variant
12SR01222	179455521	245009	c.60931C>T	p.Arg20311X	305	0.98	Y	Y		Nonsense
10TM00933	179453427	247103	c.63025C>T	p.Arg21009X	305	0.98	Y	Y		Nonsense
10CS01472	179441341	259189	c.69630C>A	p.Tyr23210X	326	0.97	Y	Y		Nonsense

Sample ID	Hg19 genomic start position	LRG genomic start position	Transcript effect	Protein effect	Exon number	Median exon PSI	Present in N2BA	Present in N2B	Present in novex-3	Variant type
12JL01327	179436553	263977	c.74306dupA	p.Asn24769LysfsX2	327	0.96	Y	Y		Frameshift
10BK01803	179432675	267855	c.78184G>T	p.Glu26062X	327	0.96	Y	Y		Nonsense
10JC00262	179432352	268178	c.78507delT	p.Gly26170ValfsX3	327	0.96	Y	Y		Frameshift
10DM00077	179429597	270933	c.81262_81269delCAGATGCT	p.Gln27088CysfsX5	327	0.96	Y	Y		Frameshift
10PP00413	179429597	270933	c.81262_81269delCAGATGCT	p.Gln27088CysfsX5	327	0.96	Y	Y		Frameshift
12NP00167	179429538	270992	c.81321C>G	p.Tyr27107X	327	0.96	Y	Y		Nonsense
10CP00605	179428346	272184	c.82513delA	p.Ile27505PhefsX20	327	0.96	Y	Y		Frameshift
10PP00465	179425769	274761	c.85090C>T	p.Arg28364X	327	0.96	Y	Y		Nonsense
12MH00114	179424219	276311	c.86640delC	p.His28881ThrfsX2	327	0.96	Y	Y		Frameshift
12JL00046	179423219	277311	c.86967G>A	p.Trp28989X	328	0.95	Y	Y		Nonsense
12ML01299	179422273	278257	c.87716delG	p.Gly29239AspfsX32	330	0.95	Y	Y		Frameshift
10MV00215	179417877	282654	c.89751_89750dupG	p.Val29918SerfsX3	336	0.95	Y	Y		Frameshift
10TD00812	179417541	282989	c.90086_90088delAAGinsA	p.Glu30029AspfsX7	336	0.95	Y	Y		Frameshift
10KF00073	179413670	286860	c.92683C>T	p.Arg30895X	340	0.96	Y	Y		Nonsense
10AG01687	179411209	289321	c.94849_94855delGATGCCC	p.Ala31618TyrfsX37	343	0.97	Y	Y		Frameshift
10KW01906	179410548	289982	c.95415_95416+2delCAGT		344	0.97	Y	Y		Canonical splice variant
10CS01784	179404524	296006	c.98265_98268dupAACA	p.His32757AsnfsX4	353	0.99	Y	Y		Frameshift
14CT01557	179399346	301184	c.101996G>A	p.Trp33999X	359	1	Y	Y		Nonsense
12PF00041	179393849	306681	c.106629delA	p.Ala35544ProfsX2	361	1	Y	Y		Frameshift
Variants affecting Novex 3 transcript only (LRG_391_t2)										
12FB00047	179614105	86425	c.13022C>G	p.Ser4341X	48	na			Y	Nonsense

Table 6.2 *TTN* truncating variants identified in the FHS Offspring cohort

Sample ID	Hg19 genomic start position	LRG genomic start position	Transcript effect	Protein effect	Exon number	Median exon PSI	Present in N2BA	Present in N2B	Present in novex-	Variant type
Variants affecting meta-transcript (LRG_391_t1)										
17431	179659646	40884	c.1245+3A>G		7	1.00	Y	Y	Y	Splice variant prediction
5630	179647533	52997	c.3100G>A		18	1.00	Y	Y	Y	Splice variant prediction
8864	179629515	71015	c.9727C>T	p.Q3243*	42	1.00	Y	Y	Y	Nonsense
15956	179621404	79126	c.10799C>A	p.S3600*	46	0.07				Nonsense
12004	179621351	79179	c.10852C>T	p.Q3618*	46	0.07				Nonsense
11278	179621019	79511	c.11183_11184insG	p.G3728fs	46	0.07				Frameshift
13791	179598245	102285	c.15776-1G>T		55	0.10	Y			Canonical splice variant
14685	179559325	141205	c.31426+1G>C		118	0.73	Y			Canonical splice variant
530	179506963	193567	c.40558+1G>A		220	0.98	Y	Y		Canonical splice variant
1508	179473609	226921	c.52128_52128delT	p.F17376fs	275	0.98	Y	Y		Frameshift
26653	179453946	246584	c.62506C>T	p.R20836*	305	0.98	Y	Y		Nonsense
21151	179442329	258201	c.68824G>A		324	0.97	Y	Y		Splice variant prediction
9484	179430224	270306	c.80635C>T	p.Q26879*	327	0.96	Y	Y		Nonsense
19648	179422272	278258	c.87716_87716delG	p.G29239fs	330	0.95	Y	Y		Frameshift
11733	179404241	296289	c.98551C>T	p.R32851*	353	0.99	Y	Y		Nonsense
21570	179404241	296289	c.98551C>T	p.R32851*	353	0.99	Y	Y		Nonsense
Variants affecting Novex-3 transcript only (LRG_391_t2)										
35	179616683	83847	c.10443_10443delA	p.K3481fs	48	na			Y	Frameshift
21984	179613466	87064	c.13660_13661insA	p.I4554fs	48	na			Y	Frameshift
13496	179613422	87108	c.13705G>T	p.E4569*	48	na			Y	Nonsense
21368	179613187	87343	c.13939_13939delG	p.E4647fs	48	na			Y	Frameshift
14608	179611821	88709	c.15305_15305delC	p.T5102fs	48	na			Y	Frameshift

Table 6.3 *TTN* truncating variants identified in the JHS cohort

Sample ID	Hg19 genomic start position	LRG genomic start position	Transcript effect	Protein effect	Exon number	Median exon PSI	Present in N2BA	Present in N2B	Present in novex-3	Variant type
Variants affecting meta-transcript (LRG_391_t1)										
J127199	179649078	51452	c.2494G>T		16	1.00	Y	Y	Y	Splice variant prediction
J183455	179648447	52083	c.2841G>T		17	1.00	Y	Y	Y	Splice variant prediction
J306717	179640236	60294	c.6355G>T	p.E2119*	28	1.00	Y	Y	Y	Nonsense
J544442	179632884	67646	c.9164-2A>T		39	1.00	Y	Y	Y	Canonical splice variant
J507994	179622355	78175	c.10592C>G	p.S3531*	45	0.02				Nonsense
J193179	179621089	79441	c.11113_11113delA	p.R3705fs	46	0.07				Frameshift
J503785	179621089	79441	c.11113_11113delA	p.R3705fs	46	0.07				Frameshift
J550924	179603088	97442	c.14093-1G>A		50	1.00	Y	Y		Canonical splice variant
J547288	179584983	115547	c.23386C>T	p.R7796*	82	0.54	Y			Nonsense
J120020	179577419	123111	c.27328+5G>A		95	0.68	Y			Splice variant prediction
J226108	179577042	123488	c.27607G>A		96	0.69	Y			Splice variant prediction
J548074	179571683	128847	c.29042-2A>C		102	0.80	Y			Canonical splice variant
J550083	179571683	128847	c.29042-2A>C		102	0.80	Y			Canonical splice variant
J558930	179571683	128847	c.29042-2A>C		102	0.80	Y			Canonical splice variant
J508635	179558336	142194	c.31594G>T		120	0.63	Y			Splice variant prediction
J162495	179514543	185987	c.39895+1G>T		212	0.14	Y			Canonical splice variant
J206585	179480363	220167	c.48460+5G>A		259	0.98	Y	Y		Splice variant prediction
J510863	179480363	220167	c.48460+5G>A		259	0.98	Y	Y		Splice variant prediction
J554679	179480363	220167	c.48460+5G>A		259	0.98	Y	Y		Splice variant prediction
J560750	179480363	220167	c.48460+5G>A		259	0.98	Y	Y		Splice variant prediction

Sample ID	Hg19 genomic start position	LRG genomic start position	Transcript effect	Protein effect	Exon number	Median exon PSI	Present in N2BA	Present in N2B	Present in novex-3	Variant type
J506451	179442329	258201	c.68824G>A		324	0.97	Y	Y		Splice variant prediction
J578580	179435736	264794	c.75123T>A	p.Y25041*	327	0.96	Y	Y		Nonsense
J545511	179425847	274683	c.85008_85011delTAGT	p.V28336fs	327	0.96	Y	Y		Frameshift
J550577	179414849	285681	c.91715_91716insA	p.N30572fs	338	0.96	Y	Y		Frameshift
J545771	179414064	286466	c.92288_92289insAAAAG	p.S30763fs	340	0.96	Y	Y		Frameshift
J567853	179414064	286466	c.92288_92289insAAAAG	p.S30763fs	340	0.96	Y	Y		Frameshift
J154955	179407385	293145	c.97192+4A>G		349	0.98	Y	Y		Splice variant prediction
J541798	179398711	301819	c.102630_102630delA	p.K34210fs	359	1.00	Y	Y		Frameshift
J542193	179397934	302596	c.103408G>T	p.E34470*	359	1.00	Y	Y		Nonsense
J316117	179397250	303280	c.104092C>T	p.R34698*	359	1.00	Y	Y		Nonsense
J551172	179394966	305564	c.106374+1delG		359	1.00	Y	Y		Canonical splice variant
Variants affecting Novex-3 transcript only (LRG_391_t2)										
J133254	179613716	86814	c.13410_13411insA	p.K4470fs	48	na			Y	Frameshift

Table 6.4 *TTN* truncating variants identified in end-stage DCM

Both 20IR01555 and 20JR01203 were independently recruited as probands, but retrospectively identified as very likely related. Both samples have been retained in this cohort as recruitment was not biased by family history. 20JM01785 contains 2 *TTN*tv, a frameshift and a canonical splice variant. 20HS01530 contains 2 *TTN*tv, a frameshift and a splice variant prediction. Without phasing it is not known whether these variants are in cis or trans. For burden testing between cohorts the number of individuals with *TTN*tv rather than the total number of *TTN*tv was used.

Sample ID	Hg19 genomic start position	LRG genomic start position	Transcript effect	Protein effect	Exon number	Median exon PSI	Present in N2BA	Present in N2B	Present in novex-3	Variant type	Previously published
Variants affecting meta-transcript (LRG_391_t1)											
20HS01530	179647533	52997	c.3100G>A	p.Val1034Met	18	100	Y	Y	Y	Splice variant prediction	Y
20SF01658	179647533	52997	c.3100G>A	p.Val1034Met	18	100	Y	Y	Y	Splice variant prediction	Y
20AH01746	179647533	52997	c.3100G>A	p.Val1034Met	18	100	Y	Y	Y	Splice variant prediction	
20VW01643	179506964	193566	c.40558G>C	p.Val13520Leu	220	100	Y	Y		Splice variant prediction	Y
20JT01288	179494968	205562	c.39358C>T	p.Pro13120Ser	240	98	Y	Y		Splice variant prediction	
20MC01968	179494967	205563	c.44281+1G>A		240	98	Y	Y		Canonical splice variant	
20SG01639	179485012	215518	c.46236C>A	p.C15412*	249	100	Y	Y		Nonsense	Y
20JH01781	179483495	217035	c.46782C>A	p.Tyr15594*	252	100	Y	Y		Nonsense	
20JB01501	179478665	221865	c.49346-1G>A		264	98	Y	Y		Canonical splice variant	
20RD01430	179477004	223526	c.50247_50247delT	p.F16749fs	267	100	Y	Y		Frameshift	Y
20JL01448	179471841	228689	c.53488G>T	p.G17830*	279	100	Y	Y		Nonsense	Y
20PD01505	179458948	241582	c.58172delA	p.Asp19391fs	298	100	Y	Y		Frameshift	
20VW01643	179457005	243525	c.59627-1G>A		303	100	Y	Y		Canonical splice variant	Y
20EG01127	179452435	248095	c.63601C>T	p.Arg21201*	307	100	Y	Y		Nonsense	
20SW01468	179450018	250512	c.64453C>T	p.R21485*	310	100	Y	Y		Nonsense	Y
20JD01577	179444429	256101	c.67495C>T	p.R22499*	320	100	Y	Y		Nonsense	
20PS01433	179441649	258881	c.69412+1G>A		325	100	Y	Y		Canonical splice variant	Y

Sample ID	Hg19 genomic start position	LRG genomic start position	Transcript effect	Protein effect	Exon number	Median exon PSI	Present in N2BA	Present in N2B	Present in novex-3	Variant type	Previously published
20CW01447	179441015	259515	c.69843_69843delA	p.K23281fs	327	100	Y	Y		Frameshift	Y
20HS01530	179440067	260463	c.70791_70791delA	p.E23597fs	327	100	Y	Y		Frameshift	Y
20FG01148	179436521	264009	c.74338C>T	p.Arg24780*	327	100	Y	Y		Nonsense	
20SF01123	179434743	265787	c.76116_76117insA	p.Asn25372fs	327	100	Y	Y		Frameshift	
20SB01532	179434473	266057	c.76383_76386delTAAT	p.Asn25462fs	327	100	Y	Y		Frameshift	
20DH01701	179429341	271189	c.81518delC	p.Pro27173fs	327	100	Y	Y		Frameshift	
20LN01500	179424398	276132	c.86459_86460delCT	p.S28820fs	327	100	Y	Y		Frameshift	Y
20BM01499	179422457	278073	c.87624C>A	p.Y29208*	329	100	Y	Y		Nonsense	Y
20DB01655	179417723	282807	c.89900_89903delATTA	p.N29967fs	336	100	Y	Y		Frameshift	Y
20AA01165	179417305	283225	c.90322_90323insT	p.Glu30108*	336	100	Y	Y		Frameshift	
20AK01584	179413187	287343	c.93166C>T	p.R31056*	340	100	Y	Y		Nonsense	Y
20JM01785	179411593	288937	c.94562dupC	p.Thr31522fs	342	96	Y	Y		Frameshift	
20PA01474	179408239	292291	c.96460_96461insA	p.T32154fs	348	99	Y	Y		Frameshift	Y
20DS01225	179404524	296006	c.98265_98268dupAACA	p.His32757fs	353	100	Y	Y		Frameshift	
20LB01523	179404491	296039	c.98299_98300delAG	p.R32767fs	353	100	Y	Y		Frameshift	Y
20PS01566	179404286	296244	c.98506C>T	p.R32836*	353	100	Y	Y		Nonsense	Y
20IR01555	179401029	299501	c.100445C>A	p.S33482*	358	99	Y	Y		Nonsense	Y
20JR01203	179401029	299501	c.100445C>A	p.Ser33482*	358	99	Y	Y		Nonsense	
20RD01346	179400320	300210	c.101021_101022delGA	p.Arg33674fs	359	100	Y	Y		Frameshift	
20JM01785	179393000	307530	c.107377+1G>A		362	100	Y	Y		Canonical splice variant	
Variants affecting Novex-3 transcript only (LRG_391_t2)											
20JL01558	179612656	87874	c.14470_14471insCACA CTCCATA	p.R4824fs	48				Y	Frameshift	Y

Table 6.5 *TTN* truncating variants identified in healthy volunteers

Sample ID	Hg19 genomic start position	LRG genomic start position	Transcript effect	Protein effect	Exon number	Median exon PSI	Present in N2BA	Present in N2B	Present in novex-3	Variant type
Variants affecting meta-transcript (LRG_391_t1)										
14AG01422	179647533	52997	c.3100G>A	p.Val1034Met	18	100	Y	Y	Y	Splice variant prediction
14SS01830	179621351	79179	c.10852C>T	p.Gln3618X	46	7				Nonsense
14JD01896	179588844	111686	c.21142C>T	p.Arg7048X	74	36	Y			Nonsense
14EC01433	179560998	139532	c.30803-2A>G		115	84	Y			Canonical splice variant
14KN01429	179553775	146755	c.32095+5G>A		126	22	Y			Splice variant prediction
14AH01539	179549632	150898	c.32554+1G>C		131	17	Y			Canonical splice variant
14MO01427	179486054	214476	c.45391delA	p.Ile15131TyrfsX46	247	100	Y	Y		Frameshift
14JM01448	179444855	255675	c.67159delA	p.Ile22387X	319	100	Y	Y		Frameshift

Figure 6.2 *TTN* exon usage in the heart and distribution of TTNtv in healthy volunteers, FHS and JHS participants, unselected and end-stage DCM patients.

Upper half shows the location of TTNtv identified in each cohort. TTNtv in the A-band and distal I-band are enriched, as previously described¹⁷, in both end-stage and unselected DCM patients with reciprocal paucity in population and healthy volunteer cohorts.

Lower half: A schematic of the *TTN* meta-transcript is shown together with N2BA, N2B and Novex3 transcripts and corresponding *TTN* sarcomere regions demarcated: Z-disk (red), I-band (blue), A-band (green), and M-band (purple). The meta-transcript (LRG_391_t1 / ENST00000589042) is a manually curated inferred complete transcript, incorporating exons of all known *TTN* isoforms (including fetal and non-cardiac isoforms) with the exception of the alternative terminal exon found only in the novex-3 transcript (LRG_391_t2 / ENST00000360870, shown in dark green). The principal adult cardiac long & short isoforms are known as N2BA and N2B: exon usage for reference versions of these isoforms (ENST00000591111, ENST00000460472) is shown, though exon usage in vivo is variable (see below). Novex-1 and novex-2 are rare cardiac isoforms that differ from N2B only in the inclusion of novex-1 and novex-2 unique exons (ENST00000359218 & ENST00000342175 shown in the red and blue respectively within the N2B track).

PSI : depicts calculated exon usage based on RNASeq data (see section 5.7.1 Exon expression estimation): grey bars represent exon usage (0%-100%) as the "proportion spliced-in" (PSI) for each exon.

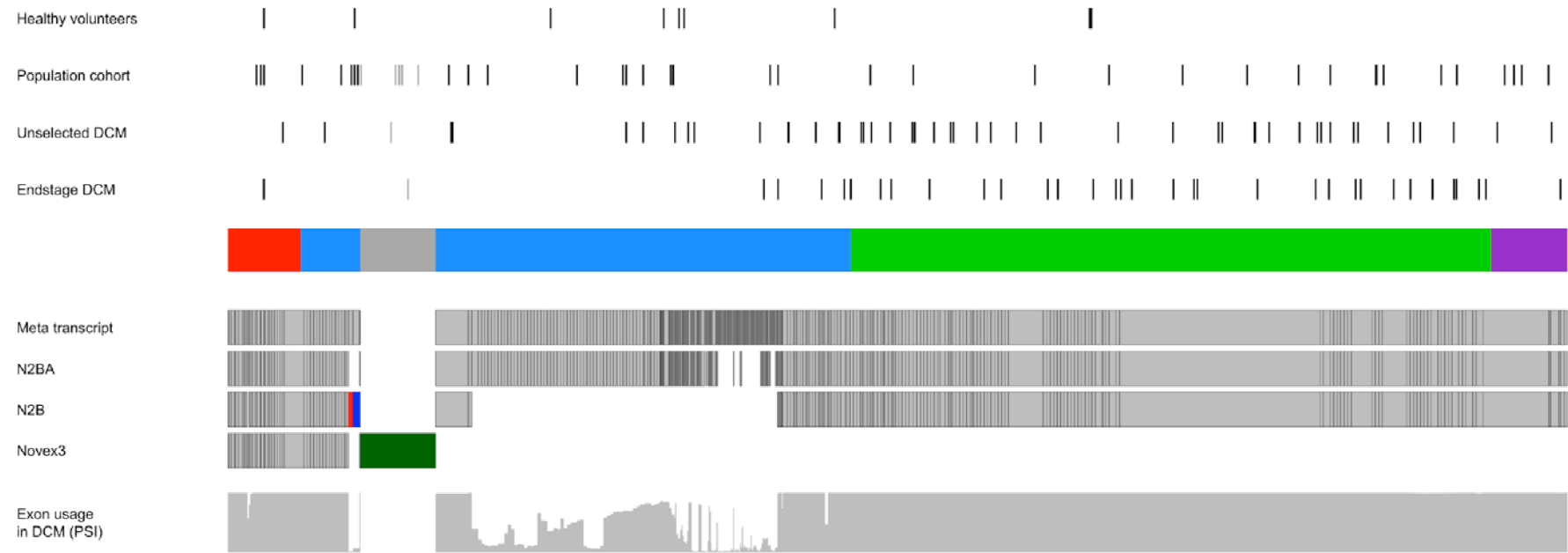
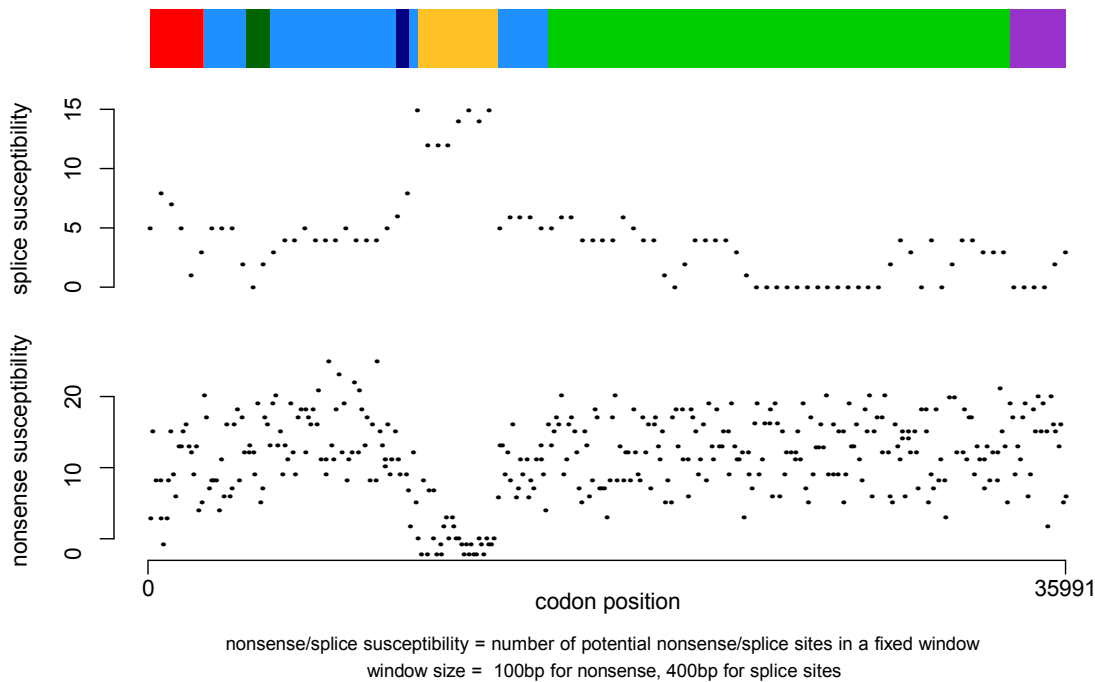


Figure 6.3 Sites susceptible to truncating events are non-uniformly distributed within the TTN gene but do not influence clustering effects in the A band.

In the top track a cartoon of the TTN meta-transcript is shown, with regions of the sarcomere demarcated as follows: Z-disk (red), I-band (blue), A-band (green), and M-band (purple). The “PEVK” domain of Titin is shown in yellow, the N2B sequence in dark green and the N2A sequence in dark blue. As exon lengths are non-uniform the number of exon boundaries (and hence sites at which splicing may be disrupted) are shown using a 400bp sliding window along the TTN transcript. Short exons in the PEVK region lead to an excess of exon boundaries and potential splice site variants, with a relative depletion of splice sites in the A-band ($P= 8.7 \times 10^{-16}$; Fisher test for number of potential splice sites in the A-band vs. rest of the gene by length), where some exons are longer. Despite this, variants affecting splice donor / acceptor sites in DCM cases cluster in the A-band. Nonsense susceptibility is also non-uniform, the PEVK region is relatively invulnerable to nonsense variation ($P= 1.3 \times 10^{-8}$ (Fisher test for number of potential nonsense variants in the A-band vs. rest of the gene by length)), as the codons representing these four amino acids can not be converted to a stop codon through a single nucleotide substitution. While this may explain some of the observed clustering in nonsense variants, an enrichment of variants in the A-band is still observed when controlling for this effect ($P= 9.1 \times 10^{-6}$; Fisher test for number of TTNtv in A-band vs. rest of gene by number of potential variant sites).



Distal exons, including those that encode the A-band portion of TTN, are annotated as being constitutively expressed in cardiac isoforms, whereas many proximal exons, particularly in the I-band, are variably spliced in different isoforms (see introduction section 1.7). Recent studies suggest that variants impacting only a subset of gene transcripts are less likely to cause loss-of-function than variants impacting all isoforms²²⁵. TTNtv that altered both N2BA and N2B were strongly enriched in DCM patients versus controls (OR=19(12-29), $P = 5.5 \times 10^{-46}$) and associated more strongly with DCM than TTNtv that affected only the long isoform (OR=3.8(1.4-9.2), $P=0.008$, Table 6.6). By contrast TTNtv found in controls were more likely to fall outside the two principal cardiac transcripts: in novex, foetal or alternative exons (13 variants in 7406bp encoding non-cardiac exons, vs. 49 variants in 103052bp encoding cardiac exons, OR=3.7(2.0-6.7), $P= 2.0 \times 10^{-4}$). Variants unique to the shorter novex-3 *TTN* isoform were not significantly enriched in DCM (0.37% DCM versus 0.15% controls, OR=2.5(0.36-13), $P =0.24$). Though this study is not powered to definitively exclude a small excess of variants in the novex-3 terminal exon, this isoform does not span the cardiac sarcomere¹⁸⁹ and these variants are unlikely to be pathogenic. Given the lack of evidence of pathogenicity these variants are excluded from subsequent analyses.

Average cardiac TTN exon usage (PSI) was determined de novo from human heart LV samples (n=84, subset of end-stage cohort) and used to annotate each TTN exon (section 3.7, RNA Sequencing, and Table 3.2). PSI gives important information distinct from conventional transcript information. A number of exons annotated as incorporated into the N2BA isoform previously had limited evidence of cardiac expression, and the new RNAseq data supports this: 39/122 exons annotated as incorporated into the N2BA isoform had minimal evidence of cardiac expression (PSI < 0.15). Conversely, three exons annotated as constitutively expressed appeared to have variable inclusion (intermediate PSI 0.15-0.9), as did two exons that are not included in the conventional N2BA/N2B descriptions. This is shown pictorially in Figure 6.4.

Treating all cohorts as an ordered variable with four levels (healthy volunteers, general population, ambulatory unselected DCM, and end-stage DCM), a strong relationship can be observed between cohort and mutant exon usage (Kruskal-Wallis chi-square, $P = 4.9 \times 10^{-3}$), with TTNtv-containing exons in controls having lower usage than TTNtv-containing exons in DCM patients ($P = 2.5 \times 10^{-4}$) (Figure. 6.5). This suggests that some variants observed in controls may be tolerated either because the putative truncating

event is rescued by exclusion of the variant exon, or because the truncated transcript and/or protein are expressed only at relatively low levels.

Figure 6.4 Usage of *TTN* exons across isoforms.

Each data point represents the PSI of an exon annotated as incorporated into the respective isoform

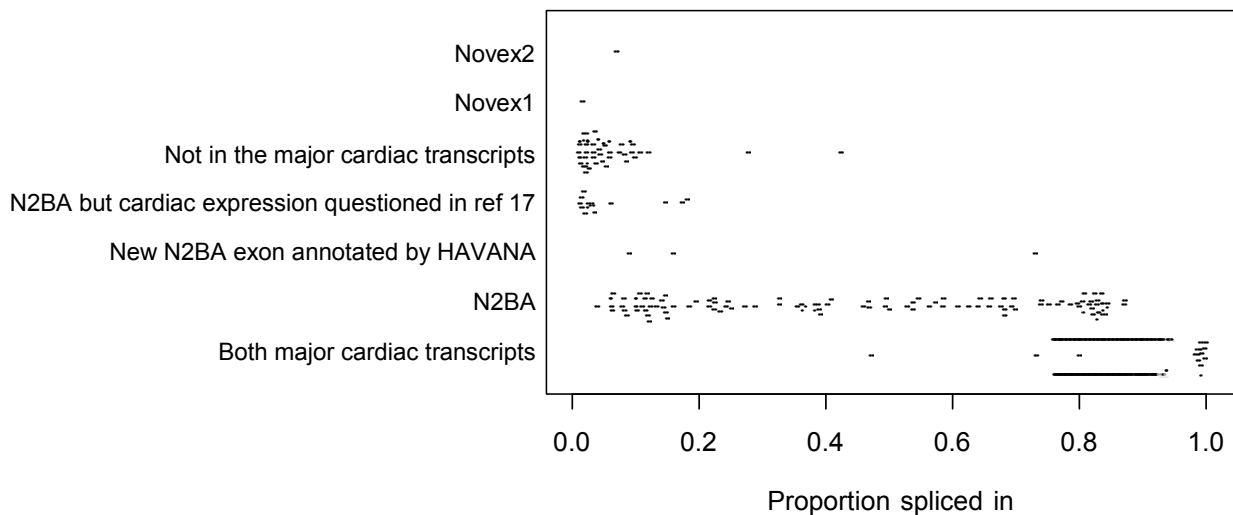


Figure 6.5 Usage of *TTN* exons containing TTNtv across all subject groups

TTN exon usage is represented as PSI "proportion spliced-in" for each exon, which is an estimate of the proportion of transcripts that incorporate each exon. Each data point represents the PSI of an exon containing a TTNtv separated by cohort. For splice variants, PSI is derived as the average PSI for exons either side of the splice junction. The PSI for exons containing TTNtv in control subjects is significantly lower than in unselected DCM cases ($P=0.003$, Wilcox Mann-Whitney) and all DCM cases (unselected and end-stage DCM cohorts) ($P = 2.5 \times 10^{-4}$).

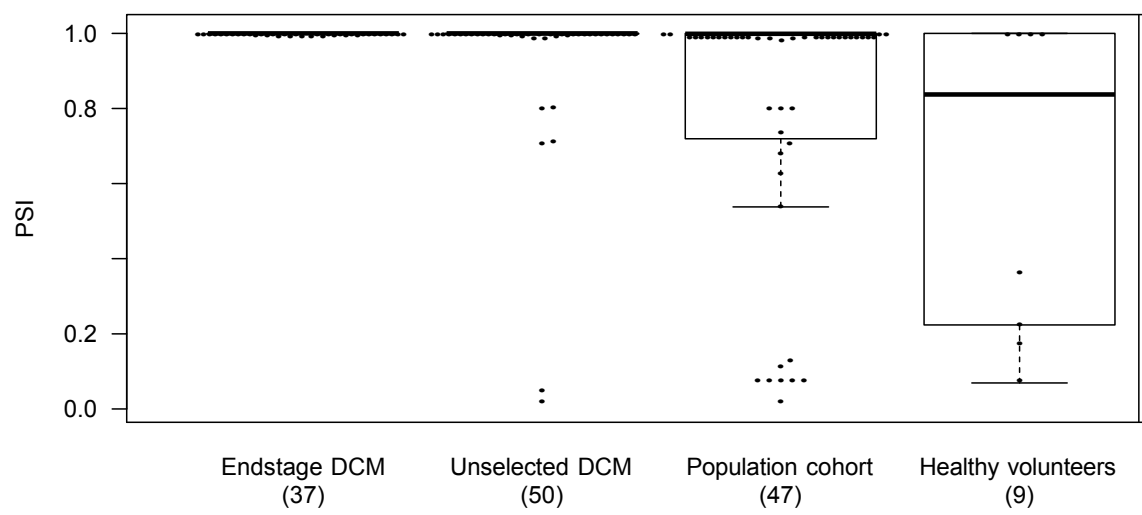


Table 6.6 Burden of TTN truncating variants in DCM patients and controls

Numbers of subjects with a TTNtv is shown for each group. TTNtv are classified by type, the affected transcript, and expression level of the variant-encoding exon. Comparisons between groups were assessed by Fisher's exact test.

^a3/374 subjects were excluded from these analyses due to relatedness to other subjects.

^bVariants that only impact the alternative terminal exon of novex-3 are excluded elsewhere.

^cExon usage levels are displayed categorically based on PSI

^dCanonical splice sites refer to the 2 intronic base pairs at the 5' and 3' splice junctions

^eVariants close to canonical splice sites that are predicted *in silico* to alter splicing

^ftotal number of individuals with TTNtv; 4 individuals with DCM (1 unselected, 3 end-stage) carry a second TTNtv which is a splice variant prediction in 3 cases.

	Discovery Cohorts					P value DCM vs. controls	
	Healthy volunteers n=308	FHS n=1623	JHS n=1980	Unselected DCM n=371 ^a	End-stage DCM n=155	Unselected DCM	All DCM
Transcript affected by truncation							
N2BA & N2B	4	11	20	42	34	1.7x10 ⁻²⁵	5.5x10 ⁻⁴⁶
N2BA only	4	2	8	7	0	0.0014	0.0089
Neither N2BA or N2B (Novex-3 terminal exon only ^b)	1 (0)	8 (5)	4 (1)	1 (1)	1 (1)	1	0.7
Totals	9	21	32	50	35		
Sarcomere domain							
A-band (18235aa)	2	7	12	32	29	0.011	0.00014
Non A-band (17756aa)	7	9	19	18	8		
Totals	9	16	31	49	34		
Usage^c of exon containing truncation							
Low (PSI <0.15)	1	4	4	2	0	0.25	0.38
Intermediate (PSI 0.15- 0.9)	4	1	7	5	0	0.012	0.044
High (PSI>0.9)	4	11	20	42	34	1.7x10 ⁻²⁵	5.5x10 ⁻⁴⁶
Totals	9	16	31	49 ^f	34 ^f		
Variant type							
Frameshift variant	2	3	7	22	15	4.9x10 ⁻¹⁶	4.0x10 ⁻²⁵
Stop gained	3	7	6	15	12	7.3x10 ⁻⁰⁹	2.3x10 ⁻¹⁵
Canonical splice sites ^d	2	3	7	9	5	2.9x10 ⁻⁰⁵	2.3x10 ⁻⁰⁷
Splice variant predictions ^e	2	3	11	4	5	0.089	0.0017
Totals	9	16	31	49 ^f	34 ^f	1.4x10 ⁻²⁵	2.8x10 ⁻⁴³

Table 6.7 TTN truncating variants in publicly available control populations

Abbreviations: ESP = exome sequencing project, 1kg = 1000 genome project. Variants are described with respect to the meta-transcript (LRG_391_t1), with the exception of variants only affecting the novex-3 alternative terminal exon that are reported with respect to novex-3 (LRG_391_t2).

*Several variants were found with considerable frequency in one population only, raising the possibility that these may be artefacts. The 3 marked variants were excluded from further analyses, as the observed allele frequencies suggest population stratification, allele ascertainment bias, or artefact ($p < 0.05$, Fisher test). None of these 3 variants are restricted to one ethnic group in the relevant population, and so would be expected to be detected in both projects if true variants. None are found in >1500 individuals sequenced at RBHT. The distributions of the remaining variants, with ≤ 23 observations in ESP or ≤ 2 observations in 1kg, are not statistically incompatible with stochastic sampling of rare variants, and so are included in analyses.

Variant effect	Transcript effect	Protein effect	ESP variant count	1KG variant count	Exon	Sarcomere domain	Exon expression	Transcripts affected
nonsense	c.325C>T	p.R109X	1	0	4	Z-band	100	NOVEX-3,NOVEX-1,N2B,N2A,META,NOVEX-2,N2BA
essential splice site	c.2370+1G>T		1	0	14	Z-band	100	N2B,NOVEX-3,N2BA,NOVEX-2,META,NOVEX-1,N2A
nonsense	c.6123G>A	p.W2041X	1	0	28	Z-band	100	NOVEX-1,N2A,N2B,NOVEX-3,N2BA,NOVEX-2,META
nonsense	c.5067G>A	p.W1689X	1	0	28	Z-band	100	N2BA,NOVEX-2,META,NOVEX-1,N2A,N2B,NOVEX-3
nonsense	c.6820C>T	p.Q2274X	1	0	30	I-band	100	NOVEX-2,N2BA,NOVEX-3,NOVEX-1,N2B,N2A,META
frameshift	c.7871dupA	p.Pro2625AlafsTer9	4	0	34	I-band	100	META,NOVEX-3,N2BA,NOVEX-1,NOVEX-2,N2B,N2A
frameshift	c.9132delA	p.Ala3045GlnfsX14	10	0	38	I-band	100	N2BA,NOVEX-2,NOVEX-1,NOVEX-3,N2A,N2B,META

Variant effect	Transcript effect	Protein effect	ESP variant count	1KG variant count	Exon	Sarcomere domain	Exon expression	Transcripts affected
nonsense	c.9448C>T	p.R3150X	1	1	40	I-band	100	NOVEX-3,NOVEX-1,N2B,N2A,META,NOVEX-2,N2BA
frameshift	c.9749_9750delTG	p.V3250Afs*40	1	0	42	I-band	100	N2A,NOVEX-1,META,N2B,N2BA,NOVEX-2,NOVEX-3
essential splice site	c.10303+2T>C		1	0	44	I-band	100	N2BA,NOVEX-2,META,NOVEX-1,N2A,N2B,NOVEX-3
nonsense	c.11190C>G	p.Y3730X	1	0	46	I-band	7	NOVEX-2,META
nonsense	c.10799C>A	p.S3600X	1	0	46	I-band	7	META,NOVEX-2
essential splice site	c.11254+2T>C		17	0	46	I-band	7	META,NOVEX-2
essential splice site	c.11254+1G>C		0	2	46	I-band	7	META,NOVEX-2
frameshift	c.11183dupG	p.Leu3729ThrfsTer9	5	0	46	I-band	7	NOVEX-2,META
frameshift	c.11113delA	p.R3705Dfs*2	2	0	46	I-band	7	NOVEX-2,META
nonsense	c.13243C>T	p.Q4415X	0	1	48	I-band	83	NOVEX-3
nonsense	c.13568C>G	p.S4523X	1	0	48	I-band	83	NOVEX-3
nonsense	c.12586G>T	p.G4196X	1	0	48	I-band	83	NOVEX-3
nonsense	c.13254T>G	p.Y4418X	0	1	48	I-band	83	NOVEX-3
nonsense	c.11752G>T	p.E3918X	1	0	48	I-band	83	NOVEX-3
nonsense	c.13705G>T	p.E4569X	1	0	48	I-band	83	NOVEX-3
nonsense	c.14122C>T	p.Q4708X	1	0	48	I-band	83	NOVEX-3
nonsense	c.15313C>T	p.R5105X	1	0	48	I-band	83	NOVEX-3
nonsense	c.16321C>T	p.R5441X	1	0	48	I-band	83	NOVEX-3
nonsense	c.16516G>T	p.E5506X	6	0	48	I-band	83	NOVEX-3
nonsense	c.16744G>T	p.G5582X	0	1	48	I-band	83	NOVEX-3
frameshift	c.10782delA	p.Leu3595CysfsTer13	1	0	48	I-band	83	NOVEX-3
frameshift	c.13410dupA	p.Gly4471ArgfsTer14	2	0	48	I-band	83	NOVEX-3
frameshift	c.13939delG	p.Glu4647LysfsTer54	4	0	48	I-band	83	NOVEX-3

Variant effect	Transcript effect	Protein effect	ESP variant count	1KG variant count	Exon	Sarcomere domain	Exon expression	Transcripts affected
frameshift	c.14415delA	p.Asp4806MetfsTer23	1	0	48	I-band	83	NOVEX-3
frameshift	c.14727delA	p.Leu4909PhefsTer35	1	0	48	I-band	83	NOVEX-3
frameshift	c.15305delC	p.Thr5102IlefsTer2	1	0	48	I-band	83	NOVEX-3
frameshift	c.16529delT	p.Val5510GlyfsTer18	1	0	48	I-band	83	NOVEX-3
nonsense	c.12587C>A	p.S4196X	1	0	49	I-band	99	META,NOVEX-1,N2B,N2BA,NOVEX-2
nonsense	c.12643C>T	p.Q4215X	1	0	49	I-band	99	N2BA,NOVEX-2,NOVEX-1,META,N2B
nonsense	c.15796C>T	p.R5266X	1	0	55	I-band	10	N2A,META,N2BA
essential splice site	c.24227-2A>G		1	0	85	I-band	8	N2BA,N2A,META
frameshift	c.25383delG	p.K8461Nfs*5	1	0	89	I-band	11	N2BA,N2A,META
frameshift	c.28016dupA	p.Pro9340AlafsTer23	23	0	98	I-band	70	META,N2A,N2BA
nonsense	c.29071A>T	p.K9691X	1	0	102	I-band	80	N2A,META,N2BA
essential splice site	c.29042-2A>C		6	0	102	I-band	80	META,N2BA,N2A
essential splice site*	c.30683-2delA		374	0	113	I-band	85	N2BA,N2A,META
essential splice site	c.31763-1G>A		5	0	123	I-band	83	N2A,META,N2BA
nonsense	c.31966A>T	p.K10656X	1	0	125	I-band	33	N2BA,N2A,META
essential splice site	c.32471-1G>A		2	0	131	I-band	17	N2A,META,N2BA
frameshift	c.33535delG	p.E11179Sfs*3	7	0	141	I-band	1	N2A,META,N2BA
essential splice site*	c.34291+2T>C		0	193	147	I-band	2	N2BA,N2A,META
essential splice site	c.34855+1G>A		1	0	154	I-band	16	N2BA,N2A,META
nonsense	c.35692A>T	p.R11898X	0	2	162	I-band	2	META
frameshift	c.35739dupT	p.Pro11914SerfsTer7	2	0	163	I-band	1	META

Variant effect	Transcript effect	Protein effect	ESP variant count	1KG variant count	Exon	Sarcomere domain	Exon expression	Transcripts affected
frameshift	c.37262delA	p.K12421Sfs*526	19	0	181	I-band	2	META
essential splice site	c.39043+1G>T		1	0	202	I-band	3	META
essential splice site	c.38960-3_38960-1delAAG		1	0	202	I-band	3	META
essential splice site	c.40558+1G>A		2	0	220	I-band	100	N2A,NOVEX-2,META,NOVEX-1,N2B,N2BA
essential splice site	c.40723+1G>T		1	0	222	I-band	100	NOVEX-1,N2A,N2B,META,NOVEX-2,N2BA
nonsense	c.43319G>A	p.W14440X	1	0	236	I-band	100	META,NOVEX-2,N2BA,NOVEX-1,N2A,N2B
nonsense	c.44272C>T	p.R14758X	0	1	240	I-band	100	META,NOVEX-2,N2BA,NOVEX-1,N2B,N2A
nonsense	c.48283C>T	p.R16095X	1	0	258	A-band	100	NOVEX-1,N2A,N2B,META,NOVEX-2,N2BA
nonsense	c.48405T>A	p.C16135X	1	0	259	A-band	100	META,NOVEX-2,N2BA,NOVEX-1,N2A,N2B
frameshift	c.48382_48383insT	p.K16128lfs*6	1	0	259	A-band	100	N2BA,NOVEX-2,META,NOVEX-1,N2A,N2B
essential splice site	c.50551+1G>A		0	1	269	A-band	100	N2B,N2A,META,NOVEX-2,N2BA,NOVEX-1
nonsense	c.60726T>A	p.Y20242X	0	1	305	A-band	100	N2BA,NOVEX-1,N2B,N2A,META,NOVEX-2
frameshift	c.66523_66524delCT	p.L22175lfs*8	2	0	317	A-band	100	META,NOVEX-1,N2A,N2B,N2BA,NOVEX-2
nonsense	c.81943G>T	p.E27315X	1	0	327	A-band	100	NOVEX-2,N2BA,NOVEX-1,N2A,N2B,META
nonsense	c.82273C>T	p.Q27425X	1	0	327	A-band	100	META,NOVEX-2,N2BA,NOVEX-1,N2A,N2B

Variant effect	Transcript effect	Protein effect	ESP variant count	1KG variant count	Exon	Sarcomere domain	Exon expression	Transcripts affected
frameshift	c.70775delT	p.V23592Gfs*4	1	0	327	A-band	100	N2BA,NOVEX-2,META,NOVEX-1,N2A,N2B
frameshift	c.94167delT	p.F31389Lfs*7	1	0	340	A-band	100	META,NOVEX-1,N2A,N2B,N2BA,NOVEX-2
nonsense	c.96892C>T	p.Q32298X	1	0	348	A-band	99	N2B,N2BA,N2A,NOVEX-2,META,NOVEX-1
essential splice site	c.100766-1G>T		0	1	359	A-band	100	NOVEX-2,N2BA,NOVEX-1,N2B,N2A,META
frameshift	c.105814delA	p.T35272Hfs*21	2	0	359	M-band	100	NOVEX-2,META,NOVEX-1,N2A,N2B,N2BA
frameshift	c.105739_105742dupGTGA	p.Lys35248SerfsTer2	1	0	359	M-band	100	NOVEX-1,META,N2B,N2BA,NOVEX-2,N2A
nonsense	c.107149C>T	p.Q35717X	1	0	361	M-band	100	N2A,N2B,META,NOVEX-2,N2BA,NOVEX-1
nonsense	c.106959T>A	p.Y35653X	1	0	361	M-band	100	META,NOVEX-2,N2BA,NOVEX-1,N2A,N2B
frameshift*	c.106998dupA	p.Ala35667SerfsTer6	57	0	361	M-band	100	N2BA,NOVEX-2,N2A,NOVEX-1,META,N2B
frameshift	c.107351delC	p.S35784*fs*1	3	0	362	M-band	100	N2BA,NOVEX-2,META,NOVEX-1,N2A,N2B

Table 6.8 Burden, type and distribution of *TTN* truncating variants in publicly available control populations

The number of alleles for each variant is shown. No variants were homozygous. Individuals are assumed to carry no more than one truncating variant, so that the allele prevalence approximates to the population prevalence.

7 / 12 1kg alleles (5 / 10 distinct variants) affect only novex or other low expression exons.

83 / 168 ESP alleles (28 / 64 distinct variants) affect these exons.

Abbreviations: 1kg = 1000 genomes project, ESP = exome sequencing project.

	Total TTNtv alleles		Excluding variants with low probability of pathogenicity (affecting only novex and other low expression exons)	
	1kg	ESP	1kg	ESP
Total TTN	12	168	5	85
Population size	1092	6504	1092	6504
Prevalence	1.1%	2.6%	0.46%	1.3%
Essential splice site	4	39	2	19
Frameshift	0	96	0	49
Nonsense	8	33	3	17
Z-band	0	2	0	2
Z-band / I-band	0	2	0	2
I-band	9	146	2	63
A-band	2	10	2	10
A-band / M-band	1	3	1	3
M-band	0	5	0	5

6.3.5 Diagnostic interpretation of TTNtv in DCM

Based on these differentiating factors, it can be estimated that TTNtv produced by nonsense, frameshift, or canonical splice site mutations (excluding predicted splice variants outside canonical splice sites) that affect highly expressed exons have at least a 93% probability of pathogenicity (likelihood ratio (LR) = 14) when identified in an unselected patient with DCM, and an even higher probability of pathogenicity in end-stage disease ($\geq 96\%$, LR= 24)(see methods for calculation of probability of pathogenicity and LR). These estimates are conservative, in that they assume all variants in controls are benign, giving an upper limit of the background noise.

6.3.6 Stratification of DCM by TTN genotype

The high accuracy and reproducibility of quantitative CMR imaging is ideally suited for genotype-phenotype studies^{172,173}. To delineate clinical phenotypes associated with TTNtv-positive DCM, CMR data for patients meeting diagnostic criteria^{169,173} (n=319) in the unselected DCM cohort was integrated with TTN genotype data. No subjects with a TTNtv had CMR features suggestive of a previous episode of myocarditis that could have caused DCM²²⁶, compared to 9% of subjects without TTNtv, supportive of a primary genetic cause in all TTNtv cases. TTNtv-positive DCM patients had more severely impaired LV function, lower stroke volumes and thinner LV walls (Table 6.9) than TTNtv-negative patients.

Consistent with these adverse intermediate phenotype associations, a difference in the composite endpoint of LV assist device implantation, listing for cardiac transplantation, and all cause mortality was observed. TTNtv-positive DCM patients reached this endpoint at earlier ages ($P = 0.015$, Figure. 6.6) and sooner after prospective enrolment ($P = 0.05$, Figure. 6.6). Midwall fibrosis, an important prognostic factor in DCM,^{174,180} was similar in patients with and without TTNtv, but sustained ventricular tachycardia was more common in TTNtv-positive patients (OR = 6.7; $P = 0.001$) and robust to adjustment for LV ejection fraction.

Figure 6.6 TTNtv and survival in DCM

Outcomes in unselected DCM patients with (red) and without (blue) TTNtv. The left panel shows age censored at adverse event (death, cardiac transplant or left-ventricular assist device (LVAD) or at age 70 years. The right panel shows adverse events after enrolment, to control for ascertainment (interval censored from time of enrolment to age 70 years or adverse event). Event free survival is reduced in TTNtv-positive DCM ($P = 0.015$) due to altered disease progression both before and after presentation. A trend to younger presentation (Table 6.9), and worse outcomes after enrolment ($P = 0.05$) combine to give reduced survival overall.

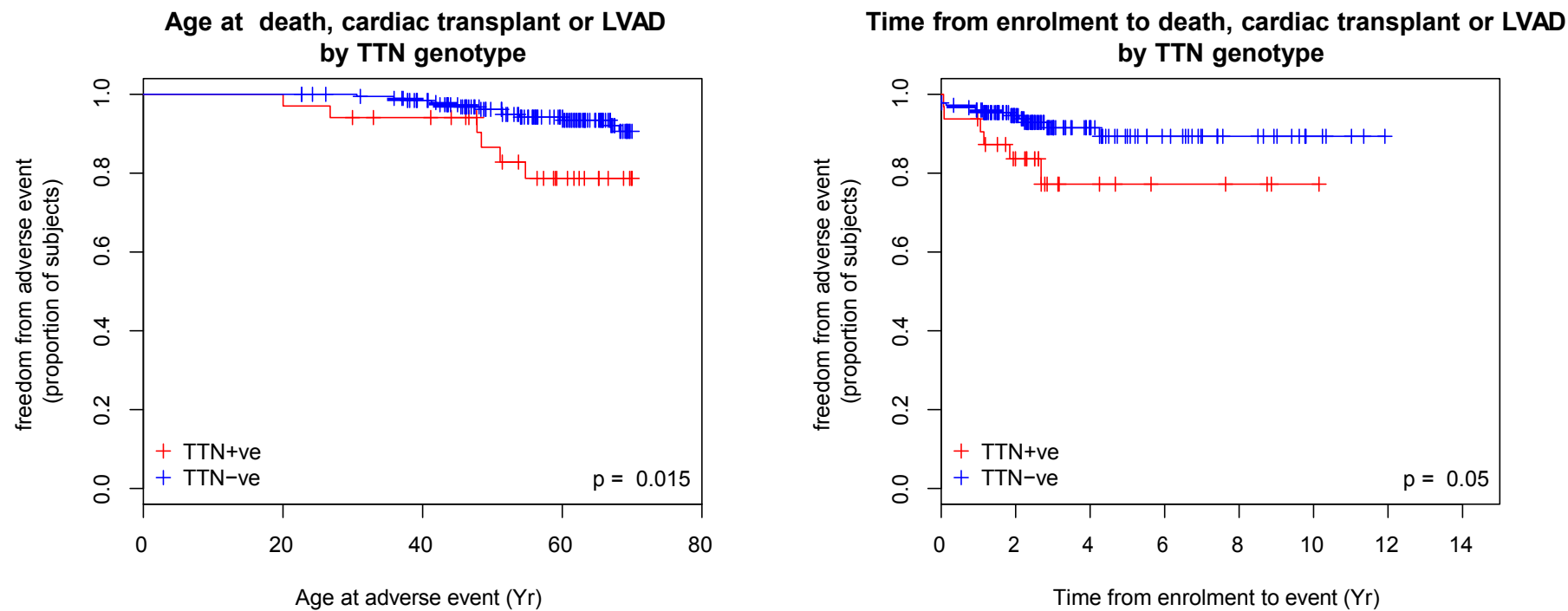


Table 6.9 Clinical characteristics in DCM patients with and without TTNtv

Unselected DCM cohort. Values are means $\pm$ SD. Volume measurements are indexed to body surface area. LV, left ventricle; RV right ventricle; EDVi/ESVi end diastolic/systolic volume; SVi stroke volume; EF ejection fraction; LVMi LV mass; WTi wall thickness; VT ventricular tachycardia; NYHA New York Heart Association functional class. P values, not corrected for multiple testing as not independent. False discovery rates (FDR)

CMR and Clinical Data		TTNtv-negative (n=277)	TTNtv-positive (n=42)	P value	False Discovery Rate
LV	EDVi	136 $\pm$ 38	137 $\pm$ 34.1	0.81	0.677
	ESVi	87.5 $\pm$ 39.1	93.7 $\pm$ 36.4	0.30	0.205
	SVi	48.2 $\pm$ 12.9	43.4 $\pm$ 14.1	0.11	0.041
	EF	37.5 $\pm$ 12.2	33.3 $\pm$ 13.3	0.11	0.047
RV	EDVi	89.4 $\pm$ 24.7	89.6 $\pm$ 26.1	0.97	0.972
	ESVi	45.2 $\pm$ 22.3	50.4 $\pm$ 25.9	0.31	0.234
	SVi	44.4 $\pm$ 12.6	39.2 $\pm$ 15.8	0.11	0.031
	EF	51.6 $\pm$ 14.1	45.2 $\pm$ 16	0.11	0.036
LVMi		95.4 $\pm$ 27.6	87.1 $\pm$ 18.3	0.106	0.20
Lateral WTi		3.13 $\pm$ 0.73	2.77 $\pm$ 0.713	0.003	0.02
Mid-wall fibrosis		94 / 270 (35%)	13 / 41 (32%)	0.869	0.81
Age at Diagnosis (years)		53.4 $\pm$ 13.4	49.3 $\pm$ 13.7	0.115	0.20
NYHA status 1/2/3/4		116/104/36/1	19/16/4/1	0.692	0.81
Sustained VT		20 / 97 (21%)	9 / 14 (64%)	0.001	0.02
Conduction Disease		82 / 227 (36%)	8 / 36 (22%)	0.130	0.30
Family history of DCM		24 / 218 (11%)	9 / 37 (24%)	0.034	0.09

Distribution and functional correlates of TTNtv

TTNtv in unselected DCM patients were considered as an allelic series in order to dissect positional effects and disease mechanism. In exploratory analysis, the location of the variant (distance from the N-terminus of the *TTN* protein to the variant) appeared to correlate with CMR indices (Figure 6.7). Distal truncations were associated with lower biventricular EF and SVi, and similar trends were seen for other parameters of cardiac dysfunction. Multivariate linear modelling (considering age, sex, TTNtv presence, and TTNtv position) was used to robustly assess the relationship between TTN genotype and cardiac phenotype. Regression models showed that TTNtv location was significantly correlated with principle indices of heart function: ejection fraction and stroke volume ($P < 0.006$; Tables GP10-11). This positional effect on cardiac parameters was large, such that a C-terminal TTNtv would be associated with substantially reduced ejection fraction (absolute reductions: left ventricle: $-18 \pm 7\%$, $P = 0.006$; right ventricle: $-21 \pm 9\%$, $P = 7.3 \times 10^{-6}$) and indexed stroke volumes (absolute reductions: left ventricle $-22 \pm 8 \text{ ml/m}^2$, $P = 0.0017$; right ventricle $-23 \pm 8 \text{ ml/m}^2$, $P = 0.0013$), as compared with an N-terminal TTNtv. In subjects with TTNtv, the variant position explained 19-23% of the observed variation (R^2) in these phenotypic indices. These positional data, together with observation that the burden of distal TTNtv increases in populations with increasingly severe disease, imply that TTNtv may not cause disease by simple haploinsufficiency. No positional effect was observed for wall thickness but indexed lateral wall thickness correlated strongly with simple presence of a TTNtv.

Figure 6.7 Truncated transcript length is correlated with indices of cardiac impairment severity in DCM.

The relationships between TTNtv location and cardiac morphology and function assessed by CMR imaging in an allelic series of cases: we assessed 14 CMR sub-phenotypes in 43 prospectively recruited DCM subjects with TTNtv and strict CMR diagnostic criteria^{173,176}. The TTNtv location (X axis) is plotted from the amino- (N) to carboxyl- (C) end of the protein. Distal (C-terminal) TTNtv were associated with worse cardiac contractile performance and associated with diminished indexed stroke volume (SVi) and ejection fraction (EF) of both left and right ventricles as compared to proximal truncations. These relationships were formally assessed using linear modelling. A regression line is shown for each variable. LV, left ventricle; RV, right ventricle; EDVi, indexed end diastolic stroke volume (ml/m²); ESVi, indexed end systolic volume (ml/m²), SVi, indexed stroke volume (ml/m²), EF ejection fraction (%).

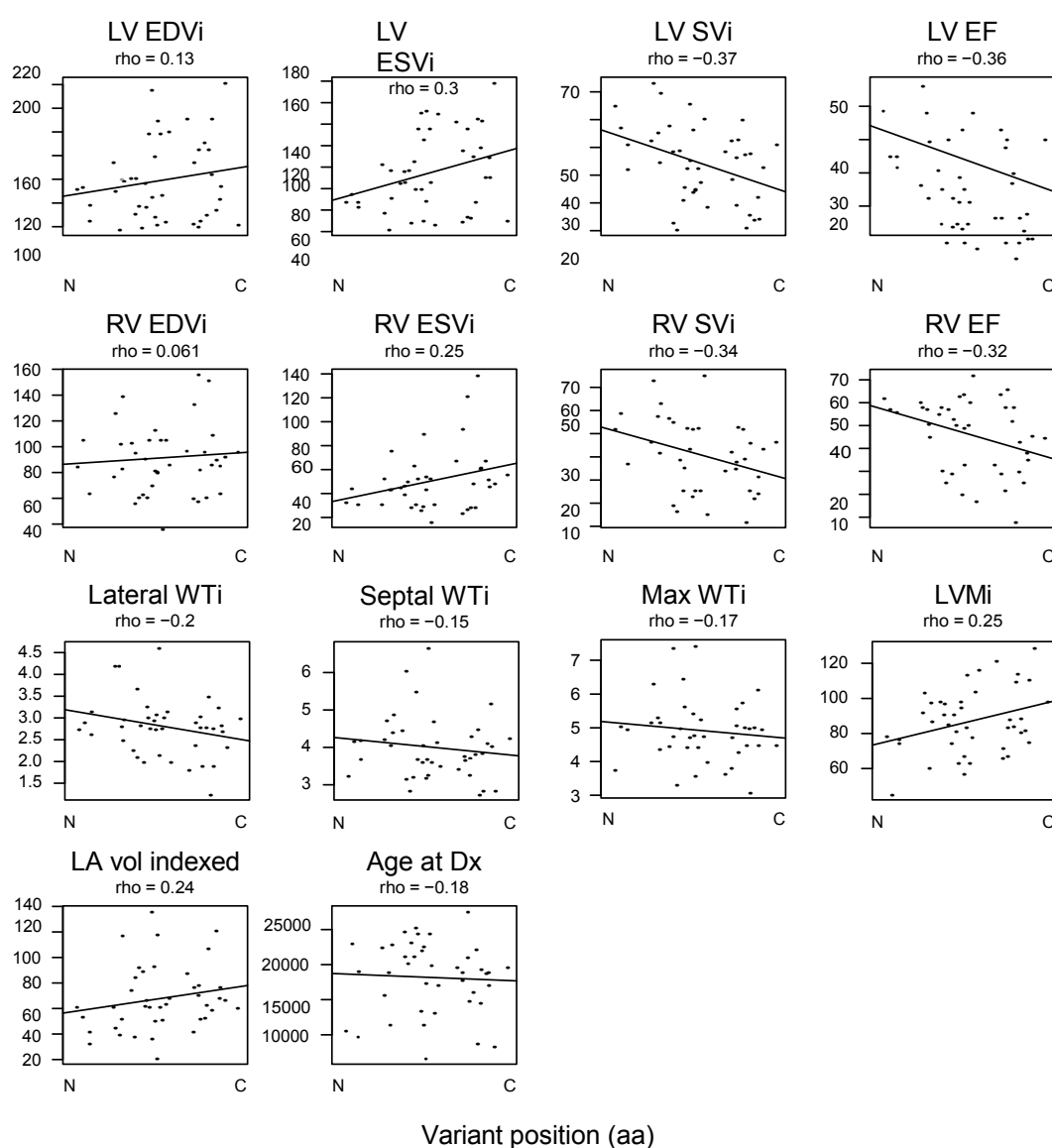


Table 6.10 Linear Modelling of the relationship between TTN genotype and phenotype for 14 continuous variables in the unselected DCM cohort

For each phenotype, 2 models are compared. First, the phenotype is modelled as a function of age & sex, and the model optimized. Statistically significant predictors are retained and listed in column 3. Next, TTN genotype is added to the model. A TTNtv may be present or absent (TTN), and if present is described by its protein coordinate (p.coord) and the level of use of the truncated exon (PSI). The model is re-optimised again, significant predictors are retained, and the two models compared using ANOVA. If no terms are retained in the model after optimization, the model terms are given as "None". If all TTN genotype terms are rejected during optimization of the second model then both models will be identical, and no P-value for comparison is returned.

Phenotype	N	Model terms	Whole population			Fitted to TTNtv positive samples		P value for model comparison
			R ²	Adjusted R ²	Delta R ²	Adjusted R ²	Delta R ²	
LV EF	318	None	0	0	0	0	0	0.006
	318	TTN, TTN:p.coord	0.03	0.03	0.03	0.19	0.19	
RV EF	314	Sex	0.05	0.04	0.05	0.05	0.05	0.003
	314	Sex, TTN, TTN:p.coord	0.08	0.07	0.04	0.24	0.19	
Age at diagnosis	292	None	0	0	0	0	0	
	292	None	0	0	0	0	0	
LAVi	315	Age_at_scan	0.05	0.04	0.05	-0.14	-0.14	
	315	Age_at_scan	0.05	0.04	0	-0.14	0	
LV EDVi	317	Sex	0.03	0.03	0.03	0.08	0.08	
	317	Sex	0.03	0.03	0	0.08	0	
LV ESVi	317	Sex	0.03	0.03	0.03	0.10	0.10	
	317	Sex	0.03	0.03	0	0.10	0	
LV SVi	317	None	0	0	0	0	0	0.002
	317	TTN, TTN:p.coord	0.04	0.03	0.04	0.23	0.23	
LVMi	316	Sex	0.07	0.07	0.07	0.03	0.03	
	316	Sex	0.07	0.07	0	0.03	0	
RV EDVi	313	Sex	0.04	0.04	0.04	-0.02	-0.02	
	313	Sex	0.04	0.04	0	-0.02	0	
RV ESVi	313	Sex	0.05	0.05	0.05	0.04	0.04	
	313	Sex	0.05	0.05	0	0.04	0	
RV SVi	313	None	0	0	0	0	0	0.001
	313	TTN, TTN:p.coord	0.04	0.04	0.04	0.20	0.20	
Septal WTi	312	Age_at_scan	0.05	0.05	0.05	0.07	0.07	
	312	Age_at_scan	0.05	0.05	0	0.07	0	
Lateral WTi	312	None	0	0	0	0	0	0.004
	312	TTN	0.03	0.02	0.03	0.16	0.16	
Max WTi	310	Age_at_scan	0.08	0.08	0.08	0.11	0.11	
	310	Age_at_scan	0.08	0.08	0	0.11	0	

Table 6.11 Full Linear Model describes impact of multivariate TTN genotype on phenotype for 14 continuous variables in the unselected DCM cohort

Definitive models were constructed, with TTN genotype defined by presence/absence of TTNtv, and the distance of the TTNtv from the N-terminus, together with a TTNtv:sex interaction term, as previous studies have suggested a more extreme phenotype in males with TTNtv. Here the distance of the truncation from the N-terminus was scaled from 0-1, so the model coefficients are readily interpretable

Phenotype	Model terms	Coefficient	Standard Error	Component P value	Full model P value
LV EF	(Intercept)	37	0.74	3.9×10^{-154}	0.006
	TTN	6.2	4.6	0.18	
	TTN:p.coord	-18	7.4	0.013	
RV EF	(Intercept)	56	1.5	1.3×10^{-114}	7.3×10^{-06}
	SexM	-6.7	1.8	0.00022	
	TTN	6.1	5.5	0.27	
	TTN:p.coord	-21	8.8	0.017	
Age at diagnosis	(Intercept)	52	0.79	5.3×10^{-179}	1
LAVi	(Intercept)	40	6.3	5.4×10^{-10}	0.00012
	Age_at_scan	0.0012	0.00031	0.00012	
LV EDVi	(Intercept)	130	4	8.7×10^{-98}	0.0016
	SexM	15	4.7	0.0016	
LV ESVi	(Intercept)	78	4.2	1.2×10^{-52}	0.0025
	SexM	15	4.9	0.0025	
LV SVi	(Intercept)	48	0.78	4.2×10^{-178}	0.0017
	TTN	7.7	4.8	0.11	
	TTN:p.coord	-22	7.8	0.0045	
LVMi	(Intercept)	83	2.8	3.6×10^{-92}	2.2×10^{-06}
	SexM	16	3.3	2.2×10^{-06}	
RV EDVi	(Intercept)	81	2.7	2.5×10^{-95}	0.00019
	SexM	12	3.1	0.00019	
RV ESVi	(Intercept)	37	2.4	3.1×10^{-40}	4.7×10^{-05}
	SexM	12	2.8	4.7×10^{-05}	
RV SVi	(Intercept)	44	0.78	1.8×10^{-166}	0.0013
	TTN	7.9	5.1	0.12	
	TTN:p.coord	-23	8.1	0.0049	
Septal WTi	(Intercept)	3.4	0.2	4.4×10^{-44}	3.4×10^{-05}
	Age_at_scan	4.2×10^{-05}	1.0×10^{-05}	3.4×10^{-05}	
Lateral WTi	(Intercept)	3.1	0.044	3.4×10^{-193}	0.0041
	TTN	-0.35	0.12	0.0041	
Max WTi	(Intercept)	4	0.25	5.2×10^{-44}	5.9×10^{-07}
	Age_at_scan	6.1×10^{-05}	1.2×10^{-05}	5.9×10^{-07}	

6.4 Discussion

6.4.1 Limitations of sequence data

Errors in NGS data can stem from experimental steps and from the bioinformatic analysis²²⁷. Ideally all samples from all cohorts would be collected, prepared, sequenced and analysed in the same way, in the same facility at the same time. However this is not always possible, particularly when large cohorts with detailed phenotyping are required. A pragmatic approach has been used here to combine 5 cohorts of large size and deep data with the understanding that TTN was target enriched, sequenced and analysed at two different centres, and that ~50 samples were whole genome sequenced by a third party (Complete Genomics).

Each sequencing platform can introduce unique biases and errors. However technical sensitivity is high given adequate read depth and the significance of platform-specific errors in the data presented here can be expected to be trivial, particularly as variants are validated. The population cohorts (FHS and JHS) were sequenced and analysed by the Seidman Laboratory at Harvard Medical School and TTNtv were checked by manual inspection of reads in the Integrative Genomics Viewer (IGV) (www.broadinstitute.org/igv/). The DCM cases (both unselected and end-stage cohorts) and healthy volunteers were sequenced and analysed at Imperial College London and Sanger validation was used for all variants. As we see an increased burden of TTNtv in DCM and Sanger validation is more stringent than manual inspection of reads, it is likely that any potential bias introduced by using different validation strategies would decrease rather than increase the difference between cases and controls. In support of the robustness of the findings presented here, the frequencies of variants across these 3 control cohorts and across two publically available control cohorts (ESP and 1000 genomes) are consistent. The frequency of TTNtv in end-stage DCM was consistent with previously published frequencies for similarly selected cohorts enriched for severe and familial DCM. The unselected DCM cohort with less severe and sporadic DCM is not equivalent to end-stage or familial DCM and, perhaps not unexpectedly, a lower prevalence of TTNtv is observed in this population as compared to previous reports.

6.4.2 Limitation of novex-3 interpretation

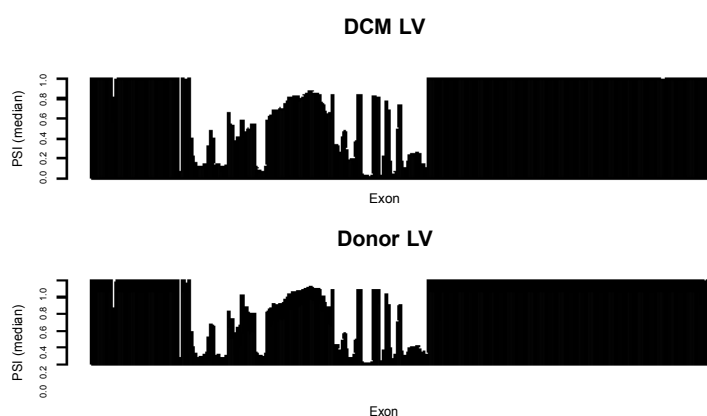
Only two novex-3 variants were observed in 529 DCM cases and confidence

intervals for the true frequency are therefore wide. However the frequency of novex-3 and low expression variants in ESP and 1KG cohorts is consistent with that seen in the population and healthy volunteer cohorts presented here suggesting these values are likely to be accurate. Additionally there are biological data to suggest these variants are unlikely to be pathogenic in that they are expressed at very low level in the heart and the novex-3 protein does not span sarcomere and so performs a different function to other TTN isoforms. It therefore seems reasonable to suggest that novex-3 variants are unlikely to be pathogenic despite lacking power to draw a definitive conclusion.

6.4.3 Limitations of deriving and applying exon usage values

All heart samples used for molecular studies reported here were from patients with advanced heart failure (n=84) The expression profile in a failing heart might be expected to differ from a healthy heart as many genes will be differentially regulated, either causing or in response to heart failure. However LV samples from donor hearts that were acquired and stored in the same manner in the same facility (n=7, sequenced after preparation of this thesis) show that the source of the tissue sample does not affect the profile of TTN exon usage (Figure 6.8).

Figure 6.8 TTN exon usage in DCM and control hearts



Using RNA-sequencing, PSI is derived from inclusion reads mapping to a specific exon and exclusion reads which do not map to any part of the exon in question. Since PSI is calculated as inclusion/(inclusion+exclusion), a low PSI value is only obtained if a large number of exclusion reads are observed. These derive from transcripts where the exon was specifically not included. Therefore as opposed to

reflecting a lack of data, a low PSI represents the presence of confirmatory exclusion reads. Thus the PSI profile would not be impacted by performance of library preparation or sequencing. Using a large number of samples to derive a mean exon usage also reduced potential for bias due to outliers.

6.4.4 Clinical stratification of DCM by TTN genotype

Patient stratification is a cornerstone of precision medicine¹⁹⁴. Data presented here suggest that DCM caused by TTNtv defines patients who may benefit from more tailored clinical management. Sustained ventricular tachycardia, LV wall thickness²²⁸ and LV ejection fraction influence DCM outcomes²²⁹. In comparison to TTNtv-negative DCM subjects, TTNtv-positive DCM subjects had increased risk for sustained ventricular tachycardia, perhaps related to increased wall stress²³⁰, an observation that may warrant a lower threshold for device therapy, as has been recommended for DCM caused by LMNA mutations²³¹. As distal TTNtv were over-represented in the end-stage DCM cohort (Figure 6.2) and were associated with worse LV function in the unselected DCM cohort, more aggressive clinical management may be appropriate in patients carrying these mutations.

6.4.5 Interpretation of TTNtv in individuals without DCM

Cardiomyopathy genes feature prominently in the American College of Medical Genetics and Genomics' list of genes in which mutations should be reported regardless of the primary indication for sequencing²³². Accurate interpretation of clinically actionable incidental findings in cardiomyopathy genes is both difficult and medically important because of the considerable population prevalence of protein-altering variants,^{232,233} a population prevalence of DCM of approximately 0.4%, and the associated medical consequences including heart failure and sudden death.²³⁴⁻²³⁶ Previous studies provided limited information about the prevalence of TTNtv in the general population and their impact on cardiovascular health^{134,224,237,238}. Here *TTN* was sequenced in over 3900 subjects from population and healthy volunteer cohorts with detailed cardiac phenotyping and clinical data.

TTNtv status does not seem to reliably predict DCM in the population cohorts: the majority of subjects with TTNtv in the general population do not appear to have DCM (population frequency of TTNtv 1:70, estimated frequency of DCM 1:250). Certain

subtypes of TTNtv appear clinically silent, approximately half of all TTNtv identified in control subjects occurred in Novex or other infrequently used exons, or altered non-canonical splice site sequences. These types of TTNtv were not associated with DCM, and as such it is likely that most of these variants are not pathogenic. Despite these insights, half of TTNtv found in population subjects without DCM were in highly expressed exons and appear comparable to TTNtv enriched in DCM subjects (population frequency of likely pathogenic TTNtv 1:140, estimated frequency of DCM 1:250; thus a frequency disparity remains). These TTNtv may cause DCM with reduced penetrance that is dependent on characteristics of the variant itself and/or the effects of additional protective and exacerbating genetic or environmental modifiers that are recognized in DCM^{72,75}. Such factors may also account for the mismatch between population prevalence of mutations in hypertrophic cardiomyopathy genes and overt disease.²³³ Alternatively it may be that these individuals will go on to develop DCM.

6.4.6 Insights into TTN DCM disease mechanism

The molecular mechanisms underlying TTN cardiomyopathy have been the subject of speculation²²⁴. It has been previously reported that variants impacting only a subset of gene transcripts are less likely to cause loss-of-function than variants impacting all isoforms²²⁵. The ability to determine which of the positional factors described are contributing to differences in phenotype is limited by their co-linearity. While a relationship with exon usage is potentially of profound clinical importance for diagnostic variant interpretation, a linear positional effect as seen within the DCM cohort would also have mechanistic implications favouring a dominant negative effect of an aberrant transcript or protein, rather than haploinsufficiency. In support of this, a 2q31q32 deletion encompassing the entire TTN locus was not found to cause cardiac muscle disease¹⁵⁶, and immunohistochemical studies have shown truncated *TTN* in the sarcomeres of skeletal muscles from patients with recessive TTN mutations¹⁵². The position-dependent functional effects reported here strongly suggest a dominant negative effect, as occurs with *MYBPC3* truncations that cause hypertrophic cardiomyopathy^{239,240}, with increasing severity associated with longer mutant proteins. Whether longer mutant TTN proteins cause more severe disease by perturbing force generation, placing greater energetic demands on the cell, by overwhelming protein degradation machinery, or indeed by acting through some

other mechanism, remains to be explored. Further elucidation of pathogenic mechanisms will be important to direct the development of appropriate therapies.

This proposed dominant negative mechanism of TTNtv has also been questioned as at least half of all DCM is 'sporadic', i.e. no affected family members are identified. However, familial DCM has incomplete penetrance and highly variable expressivity. Moreover, family history may not be robust: DCM cannot reliably be excluded in any individual family member, even with the most thorough clinical and imaging assessment, until an advanced age ($\geq$ age 50), given the late onset in many cases. Data presented here shows that TTNtv are associated with familial DCM (see Table 6.9), and is supported by the fact that TTNtv only account for ~13% unselected DCM but ~22% of severe and familial DCM.

6.5 Summary

TTNtv were found in 1-2% of the general population, in 13% of unselected DCM patients and in 22% of patients with end-stage DCM. These data show that TTNtv are the most common cause of DCM and illuminate important characteristics that distinguish TTNtv pathogenicity. Nonsense, frameshift, and canonical splice site TTNtv, particularly those that alter both principal isoforms of *TTN* and/or reside in the carboxyl A-band, cause DCM with severely impaired LV function and life-threatening ventricular arrhythmias. In contrast, truncations that occur in novel-specific exons or other infrequently used *TTN* exons are less likely to be deleterious. An immediate clinical utility of these findings is improved variant interpretation that enables cascade screening of relatives with pathogenic mutations and gene and genotype-guided stratified management of DCM^{233,241}. However, further study is necessary before TTN sequencing in the general population, in the absence of substantial clinical suspicion, will be sufficient to predict DCM. Further elucidation of the myocyte processes altered by large dominant negative mutant *TTN* proteins will be important to direct the development of therapies that prevent or attenuate the progression of TTNtv related DCM.

7 Concluding remarks

7.1 Overview of thesis

DCM is genetically and clinically heterogeneous, presenting a real impediment to genotype-phenotype correlation and the application of genetic information to augment patient management. TTNtv were recently identified as a common cause of severe DCM, but their contribution to the full spectrum of DCM including mild disease was unknown. This, together with the fact that TTNtv sometimes occur in healthy individuals, has posed significant challenges for the interpretation of these variants in an era of accessible genome sequencing. The mechanism by which TTNtv impact clinical outcomes is poorly understood, leading some to speculate that all titinopathies, including TTN+ve DCM, are recessive.

In this thesis, the power of quantitative CMR has been integrated with the capacity of NGS to assess the relationship between TTN genotype and cardiac phenotype.

First, a prospectively recruited DCM cohort was established following CMR assessment. Cardiac dilatation and impaired contraction define DCM but are also a common final pathway in many cardiac pathologies including ischaemic cardiomyopathy. Deep phenotyping by CMR and collation of detailed clinical data enabled curation of an unalloyed cohort, with greater power to detect genotype-phenotype associations.

An NGS assay was produced that captures all TTN coding exons and splice sites. Refinement and testing established optimal parameters for sequencing and analysis of variants. Genotype data was generated for 374 prospective, unselected DCM cases, 155 end-stage retrospective cases, and 308 MRI phenotyped healthy volunteers. Integration of these data with that from 3603 population subjects identified molecular signatures that aid interpretation of TTN truncations both in DCM and as incidental findings.

Overall, TTNtv were identified in 1.4% of controls, 13% of unselected and 22% of end-stage DCM cases confirming TTN as the commonest cause of genetic DCM in all patient groups. TTNtv-containing exons in DCM have higher usage than those in

controls. Nonsense, frameshift and essential splice site TTNtv in frequently used exons are estimated to have >93% probability of pathogenicity (likelihood ratio 14). The findings in this thesis do not support pathogenicity of TTN variants predicted in silico to alter non-essential splice sites outside the two conserved intronic base pairs, in the absence of supportive segregation or RNA sequencing data.

Finally, integration of TTN genotype with CMR sub-phenotypes and clinical outcomes in the DCM cohort revealed clinically important manifestations of TTNtv+ve DCM and suggest a length-dependent, dominant negative mechanism of disease. Compared to TTNtv-negative DCM, TTNtv-positive patients had lower LVEF ($P=0.02$), thinner LV walls ($P=0.02$), and a higher incidence of sustained ventricular tachycardia ($P=0.001$). C-terminus TTNtv were also associated with lower LVEF versus N-terminus ($\beta=-18\pm7\%$, $p=0.006$) and were more common in end-stage disease.

Incorporation of exon-specific usage annotation and variant position as described in this thesis could improve interpretation of TTNtv, enabling precision medicine.

7.2 Novelty

DCM phenotype data presented in this thesis relates to a population that is representative of the most commonly encountered form of DCM in the clinic. This population with less severe heart muscle disease and sporadic DCM is not equivalent to end-stage or familial DCM previously published ¹⁷ and the role of *TTN* in this patient group is largely undescribed. In addition, the depth of clinical data presented in this thesis for the unselected DCM cohort, goes far beyond that previously possible.

The volume of TTN genotype data for phenotyped DCM cases, population controls and healthy volunteers brought to bear in these analyses is far greater than that previously available.

7.3 Future work

The emergence of TTNtv as a major cause of DCM sparked a great deal of interest in other TTN variants as a potential cause. Most TTN missense variants do not cause

disease. That truncating variants are more likely to cause disease than missense variants is not surprising: this is commonly seen in other inherited cardiac conditions, and truncating variants cause much more dramatic protein alterations than point substitutions.

The problem of distinguishing those missense variants that are disease-causing from those that are not is likewise not unique, though the scale of the problem is perhaps unique for TTN given its size and the very significant burden of rare variants in the general population.

Relatively little is known about the mechanism of action of those TTN missense variants that do cause DCM. They are typically reported to follow a dominant inheritance pattern. At the time of writing, only one published report describes functional characterisation of two TTN missense variants¹⁴⁶, which suggest reduced binding of various interacting proteins, that would be unlikely to confer a dominant negative effect.

However, it is not unusual for truncating and missense variants to cause the same disease through different mechanisms. For example, long QT syndrome may be caused by truncating variants in KCNQ1 leading to haploinsufficiency whereas, various missense variants in KCNQ1 can produce either a partial loss of function (typically with milder phenotype), or exert a dominant negative effect (with more severe phenotype) depending on the precise location of the variant in the protein²¹.

Preliminary analyses by the Cardiovascular Genetics and Genomics group reveal that ~28% of both cases and healthy volunteers carry at least one novel TTN non-synonymous SNP (nsSNP). This number increases to ~45% when considering variants present in up to 0.1% of the Caucasian population (represented up to 10 times in ESP/1KG data). A detailed analysis of TTN missense mutations will require larger cohorts and domain-specific analyses using protein-modelling approaches.

References

1. Maron BJ, Towbin JA, Thiene G, et al. Contemporary definitions and classification of the cardiomyopathies: an American Heart Association Scientific Statement from the Council on Clinical Cardiology, Heart Failure and Transplantation Committee; Quality of Care and Outcomes Research and Functional Genomics and Translational Biology Interdisciplinary Working Groups; and Council on Epidemiology and Prevention. *Circulation* 2006;113:1807-16.
2. Elliott P, Andersson B, Arbustini E, et al. Classification of the cardiomyopathies: a position statement from the European Society Of Cardiology Working Group on Myocardial and Pericardial Diseases. *Eur Heart J* 2008;29:270-6.
3. Codd MB, Sugrue DD, Gersh BJ, Melton LJ. Epidemiology of idiopathic dilated and hypertrophic cardiomyopathy. A population-based study in Olmsted County, Minnesota, 1975-1984. *Circulation* 1989;80:564-72.
4. Hershberger RE, Hedges DJ, Morales A. Dilated cardiomyopathy: the complexity of a diverse genetic architecture. *Nature reviews Cardiology* 2013;10:531-47.
5. Taylor AL. The African American Heart Failure Trial: a clinical trial update. *The American journal of cardiology* 2005;96:44-8.
6. Echols MR, Felker GM, Thomas KL, et al. Racial differences in the characteristics of patients admitted for acute decompensated heart failure and their relation to outcomes: results from the OPTIME-CHF trial. *Journal of cardiac failure* 2006;12:684-8.
7. Coughlin SS, Myers L, Michaels RK. What explains black-white differences in survival in idiopathic dilated cardiomyopathy? The Washington, DC, Dilated Cardiomyopathy Study. *Journal of the National Medical Association* 1997;89:277-82.
8. Coughlin SS, Gottdiener JS, Baughman KL, et al. Black-white differences in mortality in idiopathic dilated cardiomyopathy: the Washington, DC, dilated cardiomyopathy study. *Journal of the National Medical Association* 1994;86:583-91.
9. Jefferies JL, Towbin JA. Dilated cardiomyopathy. *Lancet* 2010;375:752-62.
10. Taylor MR, Carniel E, Mestroni L. Cardiomyopathy, familial dilated. *Orphanet journal of rare diseases* 2006;1:27.
11. Members ATF, McMurray JJV, Adamopoulos S, et al. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. *European Heart Journal* 2012;33:1787-847.
12. Watkins H. Genetic clues to disease pathways in hypertrophic and dilated cardiomyopathies. *Circulation* 2003;107:1344-6.
13. Watkins H, Ashrafian H, Redwood C. Inherited cardiomyopathies. *The New England journal of medicine* 2011;364:1643-56.
14. Hershberger RE, Cowan J, Morales A, Siegfried JD. Progress With Genetic Cardiomyopathies. *Circulation: Heart Failure* 2009;2:253-61.
15. Framework TCHDNS. The Coronary Heart Disease National Service Framework: Building on excellence,maintaining progress - Progress report for 2008. In: Health Do, ed.2009.
16. DeMott K NL, Shaw EJ, Minhas R,, Humphries SE KM, Ritchie G,, Nunes V DD, Lee P, McDowell I,, Neil A QN, Rowlands P, Seed M,, Stracey H TM, Watson M. Clinical Guidelines and Evidence Review for Familial hypercholesterolaemia: the identification and management of adults and children with familial hypercholesterolaemia. . In: Practitioners NCCfPCaRCoG, ed.2008.

17. Herman DS, Lam L, Taylor MR, et al. Truncations of titin causing dilated cardiomyopathy. *The New England journal of medicine* 2012;366:619-28.
18. Girolami F, Ho CY, Semsarian C, et al. Clinical features and outcome of hypertrophic cardiomyopathy associated with triple sarcomere protein gene mutations. *J Am Coll Cardiol* 2010;55:1444-53.
19. Xu T, Yang Z, Vatta M, et al. Compound and digenic heterozygosity contributes to arrhythmogenic right ventricular cardiomyopathy. *J Am Coll Cardiol* 2010;55:587-97.
20. Hilfiker-Kleiner D, Knoll R. Disease-modifying mutations in familial hypertrophic cardiomyopathy: complexity from simplicity. *Circulation* 2008;117:1775-7.
21. Moss AJ, Shimizu W, Wilde AA, et al. Clinical aspects of type-1 long-QT syndrome by location, coding type, and biophysical function of mutations involving the KCNQ1 gene. *Circulation* 2007;115:2481-9.
22. Vincent GM, Schwartz PJ, Denjoy I, et al. High efficacy of beta-blockers in long-QT syndrome type 1: contribution of noncompliance and QT-prolonging drugs to the occurrence of beta-blocker treatment "failures". *Circulation* 2009;119:215-21.
23. Goldenberg I, Bradley J, Moss A, et al. Beta-blocker efficacy in high-risk patients with the congenital long-QT syndrome types 1 and 2: implications for patient management. *Journal of cardiovascular electrophysiology* 2010;21:893-901.
24. Liu JF, Moss AJ, Jons C, et al. Mutation-specific risk in two genetic forms of type 3 long QT syndrome. *Am J Cardiol* 2010;105:210-3.
25. Schwartz PJ, Spazzolini C, Crotti L. All LQT3 patients need an ICD: true or false? *Heart rhythm : the official journal of the Heart Rhythm Society* 2009;6:113-20.
26. Ackerman MJ, Priori SG, Willems S, et al. HRS/EHRA expert consensus statement on the state of genetic testing for the channelopathies and cardiomyopathies this document was developed as a partnership between the Heart Rhythm Society (HRS) and the European Heart Rhythm Association (EHRA). *Heart rhythm : the official journal of the Heart Rhythm Society* 2011;8:1308-39.
27. Bos JM, Towbin JA, Ackerman MJ. Diagnostic, prognostic, and therapeutic implications of genetic testing for hypertrophic cardiomyopathy. *J Am Coll Cardiol* 2009;54:201-11.
28. Villard E, Duboscq-Bidot L, Charron P, et al. Mutation screening in dilated cardiomyopathy: prominent role of the beta myosin heavy chain gene. *Eur Heart J* 2005;26:794-803.
29. Mogensen J, Murphy RT, Shaw T, et al. Severe disease expression of cardiac troponin C and T mutations in patients with idiopathic dilated cardiomyopathy. *J Am Coll Cardiol* 2004;44:2033-40.
30. Taylor MR, Fain PR, Sinagra G, et al. Natural history of dilated cardiomyopathy due to lamin A/C gene mutations. *J Am Coll Cardiol* 2003;41:771-80.
31. Luft FC. Monogenic hypertension: lessons from the genome. *Kidney Int* 2001;60:381-90.
32. McCarthy MI. Genomics, type 2 diabetes, and obesity. *N Engl J Med* 2010;363:2339-50.
33. Hindorff LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America* 2009;106:9362-7.
34. Cunnington MS, Mayosi BM, Hall DH, et al. Novel genetic variants linked to coronary artery disease by genome-wide association are not associated with carotid artery intima-media thickness or intermediate risk phenotypes. *Atherosclerosis* 2009;203:41-4.

35. Janssens ACJW, van Duijn CM. Genome-based prediction of common diseases: advances and prospects. *Hum Mol Genet* 2008;ddn250.
36. Salari K, Watkins H, Ashley EA. Personalized medicine: hope or hype? *Eur Heart J* 2012;33:1564-70.
37. Bardy GH, Lee KL, Mark DB, et al. Amiodarone or an implantable cardioverter-defibrillator for congestive heart failure. *N Engl J Med* 2005;352:225-37.
38. Hulot JS, Bura A, Villard E, et al. Cytochrome P450 2C19 loss-of-function polymorphism is a major determinant of clopidogrel responsiveness in healthy subjects. *Blood* 2006;108:2244-7.
39. Simon T, Verstuyft C, Mary-Krause M, et al. Genetic determinants of response to clopidogrel and cardiovascular events. *N Engl J Med* 2009;360:363-75.
40. Mega JL, Close SL, Wiviott SD, et al. Cytochrome p-450 polymorphisms and response to clopidogrel. *N Engl J Med* 2009;360:354-62.
41. U.S. Food and Drug Administration. FDA Drug Safety Communication: Reduced effectiveness of Plavix (clopidogrel) in patients who are poor metabolizers of the drug. In: Services USDoHH, ed.: U.S. Department of Health & Human Services; 2010.
42. Pare G, Mehta SR, Yusuf S, et al. Effects of CYP2C19 genotype on outcomes of clopidogrel treatment. *N Engl J Med* 2010;363:1704-14.
43. Wallentin L, James S, Storey RF, et al. Effect of CYP2C19 and ABCB1 single nucleotide polymorphisms on outcomes of treatment with ticagrelor versus clopidogrel for acute coronary syndromes: a genetic substudy of the PLATO trial. *Lancet* 2010;376:1320-8.
44. Anderson JL, Horne BD, Stevens SM, et al. Randomized trial of genotype-guided versus standard warfarin dosing in patients initiating oral anticoagulation. *Circulation* 2007;116:2563-70.
45. Klein TE, Altman RB, Eriksson N, et al. Estimation of the warfarin dose with clinical and pharmacogenetic data. *N Engl J Med* 2009;360:753-64.
46. Brunham LR, Lansberg PJ, Zhang L, et al. Differential effect of the rs4149056 variant in SLCO1B1 on myopathy associated with simvastatin and atorvastatin. *Pharmacogenomics J* 2011.
47. Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends in genetics : TIG* 2001;17:502-10.
48. Lifton RP. Molecular genetics of human blood pressure variation. *Science (New York, NY)* 1996;272:676-80.
49. Holm H, Gudbjartsson DF, Arnar DO, et al. Several common variants modulate heart rate, PR interval and QRS duration. *Nature genetics*;42:117-22.
50. Iyengar SK, Elston RC. The genetic basis of complex traits: rare variants or "common gene, common disease"? *Methods in molecular biology (Clifton, NJ)* 2007;376:71-84.
51. Lander ES. The new genomics: global views of biology. *Science (New York, NY)* 1996;274:536-9.
52. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science (New York, NY)* 2008;322:881-8.
53. Moulik M, Vatta M, Witt SH, et al. ANKRD1, the gene encoding cardiac ankyrin repeat protein, is a novel dilated cardiomyopathy gene. *Journal of the American College of Cardiology* 2009;54:325-33.
54. Meyer T, Ruppert V, Ackermann S, et al. Novel mutations in the sarcomeric protein myopalladin in patients with dilated cardiomyopathy. *European journal of human genetics : EJHG* 2013;21:294-300.
55. Seidman JG, Seidman C. The genetic basis for cardiomyopathy: from mutation identification to mechanistic paradigms. *Cell* 2001;104:557-67.

56. Neale BM. Introduction to linkage disequilibrium, the HapMap, and imputation. Cold Spring Harbor protocols 2010;2010:pdb.top74.
57. Pritchard JK, Cox NJ. The allelic architecture of human disease genes: common disease–common variant...or not? Human Molecular Genetics 2002;11:2417-23.
58. Collins FS, Morgan M, Patrinos A. The Human Genome Project: lessons from large-scale biology. Science (New York, NY) 2003;300:286-90.
59. International HapMap C. A haplotype map of the human genome. Nature 2005;437:1299-320.
60. Al-Chalabi A. Genome-Wide Association Studies. Cold Spring Harbor protocols 2009;2009:pdb.top66.
61. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. Nature 2009;461:747-53.
62. Ehret GB, Morrison AC, O'Connor AA, et al. Replication of the Wellcome Trust genome-wide association study of essential hypertension: the Family Blood Pressure Program. European journal of human genetics : EJHG 2008;16:1507-11.
63. Ehret GB, O'Connor AA, Weder A, Cooper RS, Chakravarti A. Follow-up of a major linkage peak on chromosome 1 reveals suggestive QTLs associated with essential hypertension: GenNet study. European journal of human genetics : EJHG 2009;17:1650-7.
64. Govindaraju DR, Larson MG, Yin X, Benjamin EJ, Rao MB, Vasan RS. Association between SNP heterozygosity and quantitative traits in the Framingham Heart Study. Annals of human genetics 2009;73:465-73.
65. Holm H, Gudbjartsson DF, Arnar DO, et al. Several common variants modulate heart rate, PR interval and QRS duration. Nature genetics 2010;42:117-22.
66. Kathiresan S, Manning AK, Demissie S, et al. A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. BMC medical genetics 2007;8 Suppl 1:S17.
67. Kathiresan S, Voight BF, Purcell S, et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nature genetics 2009;41:334-41.
68. Newton-Cheh C, Guo CY, Wang TJ, O'Donnell C J, Levy D, Larson MG. Genome-wide association study of electrocardiographic and heart rate variability traits: the Framingham Heart Study. BMC medical genetics 2007;8 Suppl 1:S7.
69. Ozaki K, Tanaka T. Genome-wide association study to identify SNPs conferring risk of myocardial infarction and their functional analyses. Cellular and molecular life sciences : CMLS 2005;62:1804-13.
70. Vasan RS, Glazer NL, Felix JF, et al. Genetic variants associated with cardiac structure and function: a meta-analysis and replication of genome-wide association data. JAMA : the journal of the American Medical Association 2009;302:168-78.
71. Vasan RS, Larson MG, Aragam J, et al. Genome-wide association of echocardiographic dimensions, brachial artery endothelial function and treadmill exercise responses in the Framingham Heart Study. BMC medical genetics 2007;8 Suppl 1:S2.
72. Villard E, Perret C, Gary Fo, et al. A genome-wide association study identifies two loci associated with heart failure due to dilated cardiomyopathy. European Heart Journal 2011;32:1065-76.
73. Norton N, Li D, Rieder MJ, et al. Genome-wide studies of copy number variation and exome sequencing identify rare variants in BAG3 as a cause of dilated cardiomyopathy. American journal of human genetics 2011;88:273-82.
74. Meder B, Ruhle F, Weis T, et al. A genome-wide association study identifies 6p21 as novel risk locus for dilated cardiomyopathy. Eur Heart J 2014;35:1069-77.

75. Stark K, Esslinger UB, Reinhard W, et al. Genetic association study identifies HSPB7 as a risk gene for idiopathic dilated cardiomyopathy. *PLoS genetics* 2010;6:e1001167.
76. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. 1977. *Biotechnology (Reading, Mass)* 1992;24:104-8.
77. McPherson JD, Marra M, Hillier L, et al. A physical map of the human genome. *Nature* 2001;409:934-41.
78. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science* 2001;291:1304-51.
79. Ware JS, Roberts AM, Cook SA. Next generation sequencing for clinical diagnostics and personalised medicine: implications for the next generation cardiologist. *Heart (British Cardiac Society)* 2012;98:276-81.
80. McPherson JD. Next-generation gap. *Nature methods* 2009;6:S2-5.
81. ten Bosch JR, Grody WW. Keeping up with the next generation: massively parallel sequencing in clinical diagnostics. *The Journal of molecular diagnostics : JMD* 2008;10:484-92.
82. Ng SB, Buckingham KJ, Lee C, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nature genetics* 2009;advance online publication.
83. Haack TB, Danhauser K, Haberberger B, et al. Exome sequencing identifies ACAD9 mutations as a cause of complex I deficiency. *Nature genetics* 2010;42:1131-+.
84. Choi M, Scholl UI, Ji W, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 2009;106:19096-101.
85. Musunuru K, Pirruccello JP, Do R, et al. Exome Sequencing, ANGPTL3 Mutations, and Familial Combined Hypolipidemia. *New England Journal of Medicine* 2010;363:2220-7.
86. Ng SB, Bigham AW, Buckingham KJ, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nature genetics* 2010;42:790-3.
87. Wang JL, Yang X, Xia K, et al. TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. *Brain* 2010;133:3510-8.
88. Ng SB, Turner EH, Robertson PD, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009;461:272-6.
89. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews Genetics* 2009;10:57-63.
90. Haas BJ, Zody MC. Advancing RNA-Seq analysis. *Nat Biotech* 2010;28:421-3.
91. Abecasis GR, Altshuler D, Auton A, et al. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061-73.
92. Mills RE, Walter K, Stewart C, et al. Mapping copy number variation by population-scale genome sequencing. *Nature* 2011;470:59-65.
93. Harvey PA, Leinwand LA. The cell biology of disease: cellular mechanisms of cardiomyopathy. *The Journal of cell biology* 2011;194:355-65.
94. Kostin S, Heling A, Hein S, Scholz D, Klövekorn W-P, Schaper J. The Protein Composition of the Normal and Diseased Cardiac Myocyte. *Heart failure reviews* 1998;2:245-60.
95. Au Y. The muscle ultrastructure: a structural perspective of the sarcomere. *Cellular and molecular life sciences : CMLS* 2004;61:3016-33.
96. Schiaffino S, Reggiani C. Molecular diversity of myofibrillar proteins: gene regulation and functional significance. *Physiological reviews* 1996;76:371-423.
97. Luther PK. The vertebrate muscle Z-disc: sarcomere anchor for structure and signalling. *Journal of muscle research and cell motility* 2009;30:171-85.

98. Frank D, Kuhn C, Katus HA, Frey N. The sarcomeric Z-disc: a nodal point in signalling and disease. *Journal of molecular medicine (Berlin, Germany)* 2006;84:446-68.
99. Knoll R, Buyandelger B, Lab M. The sarcomeric Z-disc and Z-discopathies. *Journal of biomedicine & biotechnology* 2011;2011:569628.
100. Agarkova I, Perriard JC. The M-band: an elastic web that crosslinks thick filaments in the center of the sarcomere. *Trends in cell biology* 2005;15:477-85.
101. Agarkova I, Ehler E, Lange S, Schoenauer R, Perriard JC. M-band: a safeguard for sarcomere stability? *Journal of muscle research and cell motility* 2003;24:191-203.
102. Thompson BR, Metzger JM. Cell biology of sarcomeric protein engineering: disease modeling and therapeutic potential. *Anatomical record (Hoboken, NJ : 2007)* 2014;297:1663-9.
103. Konstam MA, Kramer DG, Patel AR, Maron MS, Udelson JE. Left ventricular remodeling in heart failure: current concepts in clinical significance and assessment. *JACC Cardiovascular imaging* 2011;4:98-108.
104. Schwarz F, Mall G, Zebe H, et al. Quantitative morphologic findings of the myocardium in idiopathic dilated cardiomyopathy. *The American journal of cardiology* 1983;51:501-6.
105. Schaper J, Froede R, Hein S, et al. Impairment of the myocardial ultrastructure and changes of the cytoskeleton in dilated cardiomyopathy. *Circulation* 1991;83:504-14.
106. Roberts WC, Siegel RJ, McManus BM. Idiopathic dilated cardiomyopathy: analysis of 152 necropsy patients. *The American journal of cardiology* 1987;60:1340-55.
107. Hughes SE, McKenna WJ. New insights into the pathology of inherited cardiomyopathy. *Heart (British Cardiac Society)* 2005;91:257-64.
108. Kuhl U, Noutsias M, Seeberg B, Schultheiss HP. Immunohistological evidence for a chronic intramyocardial inflammatory process in dilated cardiomyopathy. *Heart (British Cardiac Society)* 1996;75:295-300.
109. Noutsias M, Pauschinger M, Schultheiss H, U Kh. Phenotypic characterization of infiltrates in dilated cardiomyopathy - diagnostic significance of T-lymphocytes and macrophages in inflammatory cardiomyopathy. *Medical science monitor : international medical journal of experimental and clinical research* 2002;8:Cr478-87.
110. Ashrafian H, Frenneaux MP. Metabolic modulation in heart failure: the coming of age. *Cardiovascular drugs and therapy / sponsored by the International Society of Cardiovascular Pharmacotherapy* 2007;21:5-7.
111. Neubauer S. The failing heart--an engine out of fuel. *The New England journal of medicine* 2007;356:1140-51.
112. Ingwall JS, Weiss RG. Is the failing heart energy starved? On using chemical energy to support cardiac function. *Circulation research* 2004;95:135-45.
113. Stanley WC, Lopaschuk GD, McCormack JG. Regulation of energy substrate metabolism in the diabetic heart. *Cardiovascular research* 1997;34:25-33.
114. Davila-Roman VG, Vedala G, Herrero P, et al. Altered myocardial fatty acid and glucose metabolism in idiopathic dilated cardiomyopathy. *Journal of the American College of Cardiology* 2002;40:271-7.
115. Kalsi KK, Smolenski RT, Pritchard RD, Khaghani A, Seymour AM, Yacoub MH. Energetics and function of the failing human heart with dilated or hypertrophic cardiomyopathy. *European journal of clinical investigation* 1999;29:469-77.
116. Depre C, Taegtmeyer H. Metabolic aspects of programmed cell survival and cell death in the heart. *Cardiovascular research* 2000;45:538-48.

117. Depre C, Taegtmeyer H. Metabolic aspects of programmed cell survival and cell death in the heart. *Cardiovascular Research* 2000;45:538-48.
118. Mughal W, Dhingra R, Kirshenbaum LA. Striking a balance: autophagy, apoptosis, and necrosis in a normal and failing heart. *Current hypertension reports* 2012;14:540-7.
119. Chiong M, Wang ZV, Pedrozo Z, et al. Cardiomyocyte death: mechanisms and translational implications. *Cell death & disease* 2011;2:e244.
120. van Empel VP, Bertrand AT, Hofstra L, Crijns HJ, Doevendans PA, De Windt LJ. Myocyte apoptosis in heart failure. *Cardiovascular research* 2005;67:21-9.
121. Mani K. Programmed cell death in cardiac myocytes: strategies to maximize post-ischemic salvage. *Heart failure reviews* 2008;13:193-209.
122. Kung G, Konstantinidis K, Kitsis RN. Programmed necrosis, not apoptosis, in the heart. *Circulation research* 2011;108:1017-36.
123. Konstantinidis K, Whelan RS, Kitsis RN. Mechanisms of cell death in heart disease. *Arteriosclerosis, thrombosis, and vascular biology* 2012;32:1552-62.
124. Nakayama H, Chen X, Baines CP, et al. Ca²⁺- and mitochondrial-dependent cardiomyocyte necrosis as a primary mediator of heart failure. *The Journal of clinical investigation* 2007;117:2431-44.
125. Elrod JW, Wong R, Mishra S, et al. Cyclophilin D controls mitochondrial pore-dependent Ca(2+) exchange, metabolic flexibility, and propensity for heart failure in mice. *The Journal of clinical investigation* 2010;120:3680-7.
126. Mizushima N, Levine B, Cuervo AM, Klionsky DJ. Autophagy fights disease through cellular self-digestion. *Nature* 2008;451:1069-75.
127. Zhu Y, Zhao L, Liu L, et al. Beclin 1 cleavage by caspase-3 inactivates autophagy and promotes apoptosis. *Protein & cell* 2010;1:468-77.
128. Kontogianni-Konstantopoulos A, Ackermann MA, Bowman AL, Yap SV, Bloch RJ. Muscle giants: molecular scaffolds in sarcomerogenesis. *Physiological reviews* 2009;89:1217-67.
129. Gregorio CC, Trombitas K, Centner T, et al. The NH2 terminus of titin spans the Z-disc: its interaction with a novel 19-kD ligand (T-cap) is required for sarcomeric integrity. *The Journal of cell biology* 1998;143:1013-27.
130. Linke WA, Granzier H. A spring tale: new facts on titin elasticity. *Biophysical journal* 1998;75:2613-4.
131. Buyandelger B, Ng KE, Miodic S, et al. Genetics of mechanosensation in the heart. *Journal of cardiovascular translational research* 2011;4:238-44.
132. Leinwand LA, Tardiff JC, Gregorio CC. Mutations in the sensitive giant titin result in a broken heart. *Circulation research* 2012;111:158-61.
133. Puchner EM, Alexandrovich A, Kho AL, et al. Mechanoenzymatics of titin kinase. *Proceedings of the National Academy of Sciences of the United States of America* 2008;105:13385-90.
134. LeWinter MM, Granzier HL. Titin is a major human disease gene. *Circulation* 2013;127:938-44.
135. Kellermayer MS, Smith SB, Bustamante C, Granzier HL. Complete unfolding of the titin molecule under external force. *Journal of structural biology* 1998;122:197-205.
136. Granzier H, Wu Y, Siegfried L, LeWinter M. Titin: physiological function and role in cardiomyopathy and failure. *Heart failure reviews* 2005;10:211-23.
137. LeWinter MM, Granzier H. Cardiac titin: a multifunctional giant. *Circulation* 2010;121:2137-45.
138. Udd B, Haravuori H, Kalimo H, et al. Tibial muscular dystrophy--from clinical description to linkage on chromosome 2q31. *Neuromuscular disorders* : NMD 1998;8:327-32.

139. Haravuori H, Makela-Bengs P, Udd B, et al. Assignment of the tibial muscular dystrophy locus to chromosome 2q31. *American journal of human genetics* 1998;62:620-6.
140. Hackman P, Vihola A, Haravuori H, et al. Tibial muscular dystrophy is a titinopathy caused by mutations in TTN, the gene encoding the giant skeletal-muscle protein titin. *American journal of human genetics* 2002;71:492-500.
141. Fougereousse F, Anderson LV, Delezoide AL, Suel L, Durand M, Beckmann JS. Calpain3 expression during human cardiogenesis. *Neuromuscular disorders : NMD* 2000;10:251-6.
142. de Seze J, Udd B, Haravuori H, et al. The first European family with tibial muscular dystrophy outside the Finnish population. *Neurology* 1998;51:1746-8.
143. Siu BL, Niimura H, Osborne JA, et al. Familial dilated cardiomyopathy locus maps to chromosome 2q31. *Circulation* 1999;99:1022-6.
144. Gerull B, Gramlich M, Atherton J, et al. Mutations of TTN, encoding the giant muscle filament titin, cause familial dilated cardiomyopathy. *Nature genetics* 2002;30:201-4.
145. Hein S, Schaper J. Weakness of a giant: mutations of the sarcomeric protein titin. *Trends in molecular medicine* 2002;8:311-3.
146. Itoh-Satoh M, Hayashi T, Nishi H, et al. Titin mutations as the molecular basis for dilated cardiomyopathy. *Biochemical and biophysical research communications* 2002;291:385-93.
147. Norton N, Li D, Rampersaud E, et al. Exome sequencing and genome-wide linkage analysis in 17 families illustrate the complex contribution of TTN truncating variants to dilated cardiomyopathy. *Circulation Cardiovascular genetics* 2013;6:144-53.
148. Satoh M, Takahashi M, Sakamoto T, Hiroe M, Marumo F, Kimura A. Structural analysis of the titin gene in hypertrophic cardiomyopathy: identification of a novel disease gene. *Biochemical and biophysical research communications* 1999;262:411-7.
149. Lopes LR, Zekavati A, Syrris P, et al. Genetic complexity in hypertrophic cardiomyopathy revealed by high-throughput sequencing. *Journal of medical genetics* 2013;50:228-39.
150. Taylor M, Graw S, Sinagra G, et al. Genetic variation in titin in arrhythmogenic right ventricular cardiomyopathy-overlap syndromes. *Circulation* 2011;124:876-85.
151. Carmignac V, Salih MA, Quijano-Roy S, et al. C-terminal titin deletions cause a novel early-onset myopathy with fatal cardiomyopathy. *Annals of neurology* 2007;61:340-51.
152. Ceyhan-Birsoy O, Agrawal PB, Hidalgo C, et al. Recessive truncating titin gene, TTN, mutations presenting as centronuclear myopathy. *Neurology* 2013;81:1205-14.
153. Palmio J, Evila A, Chapon F, et al. Hereditary myopathy with early respiratory failure: occurrence in various populations. *Journal of neurology, neurosurgery, and psychiatry* 2014;85:345-53.
154. Manolagos E, Vetro A, Kefalas K, et al. Deletion 2q31.2-q31.3 in a 4-year-old girl with microcephaly and severe mental retardation. *American journal of medical genetics Part A* 2011;155a:1476-82.
155. Mitter D, Chiaie BD, Ludecke HJ, et al. Genotype-phenotype correlation in eight new patients with a deletion encompassing 2q31.1. *American journal of medical genetics Part A* 2010;152a:1213-24.
156. Prontera P, Bernardini L, Stangoni G, et al. 2q31.2q32.3 deletion syndrome: report of an adult patient. *American journal of medical genetics Part A* 2009;149A:706-12.

157. Rifai L, Port-Lis M, Tabet AC, et al. Ectodermal dysplasia-like syndrome with mental retardation due to contiguous gene deletion: further clinical and molecular delineation of del(2q32) syndrome. *American journal of medical genetics Part A* 2010;152a:111-7.
158. Seeley M, Huang W, Chen Z, Wolff WO, Lin X, Xu X. Depletion of zebrafish titin reduces cardiac contractility by disrupting the assembly of Z-discs and A-bands. *Circulation research* 2007;100:238-45.
159. Xu X, Meiler SE, Zhong TP, et al. Cardiomyopathy in zebrafish due to mutation in an alternatively spliced exon of titin. *Nature genetics* 2002;30:205-9.
160. Garvey SM, Rajan C, Lerner AP, Frankel WN, Cox GA. The muscular dystrophy with myositis (mdm) mouse mutation disrupts a skeletal muscle-specific domain of titin. *Genomics* 2002;79:146-9.
161. Charton K, Daniele N, Vihola A, et al. Removal of the calpain 3 protease reverses the myopathology in a mouse model for titinopathies. *Hum Mol Genet* 2010;19:4608-24.
162. Gotthardt M, Hammer RE, Hubner N, et al. Conditional expression of mutant M-line titins results in cardiomyopathy with altered sarcomere structure. *The Journal of biological chemistry* 2003;278:6059-65.
163. Peng J, Raddatz K, Labeit S, Granzier H, Gotthardt M. Muscle atrophy in titin M-line deficient mice. *Journal of muscle research and cell motility* 2005;26:381-8.
164. Weinert S, Bergmann N, Luo X, Erdmann B, Gotthardt M. M line-deficient titin causes cardiac lethality through impaired maturation of the sarcomere. *The Journal of cell biology* 2006;173:559-70.
165. Gramlich M, Michely B, Krohne C, et al. Stress-induced dilated cardiomyopathy in a knock-in mouse model mimicking human titin-based disease. *Journal of molecular and cellular cardiology* 2009;47:352-8.
166. Selvanayagam J, Robson M, Francis J, Neubauer S. *Cardiovascular Magnetic Resonance: Basic Principles, Methods, and Techniques*. Cardiac CT, PET and MR: Wiley-Blackwell; 2010:30-71.
167. Ridgway J. Cardiovascular magnetic resonance physics for clinicians: part I. *Journal of Cardiovascular Magnetic Resonance* 2010;12:71.
168. Ginat DT, Fong MW, Tuttle DJ, Hobbs SK, Vyas RC. Cardiac imaging: Part 1, MR pulse sequences, imaging planes, and basic anatomy. *AJR American journal of roentgenology* 2011;197:808-15.
169. Keenan NG, Pennell DJ. CMR of ventricular function. *Echocardiography (Mount Kisco, NY)* 2007;24:185-93.
170. Semelka RC, Tomei E, Wagner S, et al. Interstudy reproducibility of dimensional and functional measurements between cine magnetic resonance studies in the morphologically abnormal left ventricle. *American heart journal* 1990;119:1367-73.
171. Grothues F, Moon JC, Bellenger NG, Smith GS, Klein HU, Pennell DJ. Interstudy reproducibility of right ventricular volumes, function, and mass with cardiovascular magnetic resonance. *American heart journal* 2004;147:218-23.
172. Grothues F, Smith GC, Moon JC, et al. Comparison of interstudy reproducibility of cardiovascular magnetic resonance with two-dimensional echocardiography in normal subjects and in patients with heart failure or left ventricular hypertrophy. *The American journal of cardiology* 2002;90:29-34.
173. Parsai C, O'Hanlon R, Prasad SK, Mohiaddin RH. Diagnostic and prognostic value of cardiovascular magnetic resonance in non-ischaemic cardiomyopathies. *Journal of cardiovascular magnetic resonance : official journal of the Society for Cardiovascular Magnetic Resonance* 2012;14:54.

174. Assomull RG, Prasad SK, Lyne J, et al. Cardiovascular magnetic resonance, fibrosis, and prognosis in dilated cardiomyopathy. *Journal of the American College of Cardiology* 2006;48:1977-85.
175. McCrohon JA, Moon JC, Prasad SK, et al. Differentiation of heart failure related to dilated cardiomyopathy and coronary artery disease using gadolinium-enhanced cardiovascular magnetic resonance. *Circulation* 2003;108:54-9.
176. Maceira AM, Prasad SK, Khan M, Pennell DJ. Normalized left ventricular systolic and diastolic function by steady state free precession cardiovascular magnetic resonance. *Journal of cardiovascular magnetic resonance : official journal of the Society for Cardiovascular Magnetic Resonance* 2006;8:417-26.
177. Govindaraju DR, Cupples LA, Kannel WB, et al. Genetics of the Framingham Heart Study population. *Advances in genetics* 2008;62:33-65.
178. Wyatt SB, Diekelmann N, Henderson F, et al. A community-driven model of research participation: the Jackson Heart Study Participant Recruitment and Retention Study. *Ethnicity & disease* 2003;13:438-55.
179. Gulati A, Ismail TF, Jabbour A, et al. Clinical utility and prognostic value of left atrial volume assessment by cardiovascular magnetic resonance in non-ischaemic dilated cardiomyopathy. *European journal of heart failure* 2013;15:660-70.
180. Gulati A, Jabbour A, Ismail TF, et al. Association of fibrosis with mortality and sudden cardiac death in patients with nonischemic dilated cardiomyopathy. *JAMA : the journal of the American Medical Association* 2013;309:896-908.
181. Tewhey R, Nakano M, Wang X, et al. Enrichment of sequencing targets from the human genome by solution hybridization. *Genome biology* 2009;10:R116.
182. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics (Oxford, England)* 2009;25:1105-11.
183. Flicek P, Amode MR, Barrell D, et al. Ensembl 2012. *Nucleic acids research* 2012;40:D84-90.
184. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* 2009;25:2078-9.
185. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* 2010;20:1297-303.
186. Dagleish R, Flicek P, Cunningham F, et al. Locus Reference Genomic sequences: an improved basis for describing human DNA variants. *Genome medicine* 2010;2:24.
187. Rios D, McLaren WM, Chen Y, et al. A database and API for variation, dense genotyping and resequencing data. *BMC bioinformatics* 2010;11:238.
188. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics (Oxford, England)* 2010;26:2069-70.
189. Bang ML, Centner T, Fornoff F, et al. The complete gene sequence of titin, expression of an unusual approximately 700-kDa titin isoform, and its interaction with obscurin identify a novel Z-line to I-band linking system. *Circulation research* 2001;89:1065-72.
190. Houdayer C, Caux-Moncoutier V, Krieger S, et al. Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Human mutation* 2012;33:1228-38.
191. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *Journal of computational biology : a journal of computational molecular cell biology* 2004;11:377-94.
192. Freimer N, Sabatti C. The human phenome project. *Nature genetics* 2003;34:15-21.

193. Hennekam RC, Biesecker LG. Next-generation sequencing demands next-generation phenotyping. *Human mutation* 2012;33:884-6.
194. Robinson PN. Deep phenotyping for precision medicine. *Human mutation* 2012;33:777-80.
195. Lanktree MB, Hassell RG, Lahiry P, Hegele RA. Phenomics: expanding the role of clinical evaluation in genomic studies. *Journal of investigative medicine : the official publication of the American Federation for Clinical Research* 2010;58:700-6.
196. Greulich S, Deluigi CC, Gloekler S, et al. CMR imaging predicts death and other adverse events in suspected cardiac sarcoidosis. *JACC Cardiovascular imaging* 2013;6:501-11.
197. Lu M, Zhao S, Jiang S, et al. Fat deposition in dilated cardiomyopathy assessed by CMR. *JACC Cardiovascular imaging* 2013;6:889-98.
198. Masci PG, Schuurman R, Andrea B, et al. Myocardial fibrosis as a key determinant of left ventricular remodeling in idiopathic dilated cardiomyopathy: a contrast-enhanced cardiovascular magnetic study. *Circulation Cardiovascular imaging* 2013;6:790-9.
199. Neilan TG, Coelho-Filho OR, Danik SB, et al. CMR quantification of myocardial scar provides additive prognostic information in nonischemic cardiomyopathy. *JACC Cardiovascular imaging* 2013;6:944-54.
200. Wu KC, Weiss RG, Thiemann DR, et al. Late gadolinium enhancement by cardiovascular magnetic resonance heralds an adverse prognosis in nonischemic cardiomyopathy. *Journal of the American College of Cardiology* 2008;51:2414-21.
201. Singh JP, Evans JC, Levy D, et al. Prevalence and clinical determinants of mitral, tricuspid, and aortic regurgitation (the Framingham Heart Study). *The American journal of cardiology* 1999;83:897-902.
202. Karaoguz R, Maydanozcu S, Altun T, Guldal M, Akyurek O, Erol C. Appropriate ICD therapy in patients with idiopathic dilated cardiomyopathy: long term follow-up. *International heart journal* 2006;47:763-73.
203. Schmitto JD, Lee LS, Mokashi SA, Bolman RM, 3rd, Cohn LH, Chen FY. Functional mitral regurgitation. *Cardiology in review* 2010;18:285-91.
204. Blondheim DS, Jacobs LE, Kotler MN, Costacurta GA, Parry WR. Dilated cardiomyopathy with mitral regurgitation: decreased survival despite a low frequency of left ventricular thrombus. *American heart journal* 1991;122:763-71.
205. George A, Figueredo VM. Alcoholic cardiomyopathy: a review. *Journal of cardiac failure* 2011;17:844-9.
206. Iacovoni A, De Maria R, Gavazzi A. Alcoholic cardiomyopathy. *Journal of cardiovascular medicine (Hagerstown, Md)* 2010;11:884-92.
207. Foster BJ, Gao T, Mackie AS, et al. Limitations of expressing left ventricular mass relative to height and to body surface area in children. *Journal of the American Society of Echocardiography : official publication of the American Society of Echocardiography* 2013;26:410-8.
208. Marcus JT, DeWaal LK, Gotte MJ, van der Geest RJ, Heethaar RM, Van Rossum AC. MRI-derived left ventricular function parameters and mass in healthy young adults: relation with gender and body size. *International journal of cardiac imaging* 1999;15:411-9.
209. Shellock FG. Magnetic resonance safety update 2002: implants and devices. *Journal of magnetic resonance imaging : JMRI* 2002;16:485-96.
210. Hillege HL, Nitsch D, Pfeffer MA, et al. Renal function as a predictor of outcome in a broad spectrum of patients with heart failure. *Circulation* 2006;113:671-8.

211. Anselme F, Moubarak G, Savoure A, et al. Implantable cardioverter-defibrillators in lamin A/C mutation carriers with cardiac conduction disorders. *Heart rhythm : the official journal of the Heart Rhythm Society* 2013;10:1492-8.
212. Guo W, Schafer S, Greaser ML, et al. RBM20, a gene for hereditary cardiomyopathy, regulates titin splicing. *Nature medicine* 2012;18:766-73.
213. MacArthur JA, Morales J, Tully RE, et al. Locus Reference Genomic: reference sequences for the reporting of clinically relevant sequence variants. *Nucleic acids research* 2014;42:D873-8.
214. Bamshad MJ, Ng SB, Bigham AW, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nature reviews Genetics* 2011;12:745-55.
215. Ajay SS, Parker SC, Abaan HO, Fajardo KV, Margulies EH. Accurate and comprehensive sequencing of personal genomes. *Genome research* 2011;21:1498-505.
216. Gao R, Liu Y, Gjesing AP, et al. Evaluation of a target region capture sequencing platform using monogenic diabetes as a study-model. *BMC genetics* 2014;15:13.
217. Bodi K, Perera AG, Adams PS, et al. Comparison of commercially available target enrichment methods for next-generation sequencing. *Journal of biomolecular techniques : JBT* 2013;24:73-86.
218. Avila-Arcos MC, Cappellini E, Romero-Navarro JA, et al. Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Scientific reports* 2011;1:74.
219. Teer JK, Bonnycastle LL, Chines PS, et al. Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing. *Genome research* 2010;20:1420-31.
220. Guo W, Bharmal SJ, Esbona K, Greaser ML. Titin diversity--alternative splicing gone wild. *Journal of biomedicine & biotechnology* 2010;2010:753675.
221. Vikhlyantsev IM, Podlubnaya ZA. New titin (connectin) isoforms and their functional role in striated muscles of mammals: facts and suppositions. *Biochemistry Biokhimiia* 2012;77:1515-35.
222. Clark MJ, Chen R, Lam HY, et al. Performance comparison of exome DNA sequencing technologies. *Nature biotechnology* 2011;29:908-14.
223. Li X, Buckton AJ, Wilkinson SL, et al. Towards clinical molecular diagnosis of inherited cardiac conditions: a comparison of bench-top genome DNA sequencers. *PloS one* 2013;8:e67744.
224. McNally EM. Genetics: broken giant linked to heart failure. *Nature* 2012;483:281-2.
225. MacArthur DG, Balasubramanian S, Frankish A, et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science (New York, NY)* 2012;335:823-8.
226. Friedrich MG, Sechtem U, Schulz-Menger J, et al. Cardiovascular magnetic resonance in myocarditis: A JACC White Paper. *J Am Coll Cardiol* 2009;53:1475-87.
227. Robasky K, Lewis NE, Church GM. The role of replicates for error mitigation in next-generation sequencing. *Nature reviews Genetics* 2014;15:56-62.
228. Pahl E, Sleeper LA, Canter CE, et al. Incidence of and risk factors for sudden cardiac death in children with dilated cardiomyopathy: a report from the Pediatric Cardiomyopathy Registry. *Journal of the American College of Cardiology* 2012;59:607-15.
229. Solomon SD, Anavekar N, Skali H, et al. Influence of ejection fraction on cardiovascular outcomes in a broad spectrum of heart failure patients. *Circulation* 2005;112:3738-44.

230. Gjesdal O, Bluemke DA, Lima JA. Cardiac remodeling at the population level-risk factors, screening, and outcomes. *Nature reviews Cardiology* 2011;8:673-85.
231. Pasotti M, Klersy C, Pilotto A, et al. Long-term outcome and risk stratification in dilated cardiomyopathies. *Journal of the American College of Cardiology* 2008;52:1250-60.
232. Green RC, Berg JS, Grody WW, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genetics in medicine : official journal of the American College of Medical Genetics* 2013;15:565-74.
233. Bick AG, Flannick J, Ito K, et al. Burden of rare sarcomere gene variants in the Framingham and Jackson Heart Study cohorts. *American journal of human genetics* 2012;91:513-9.
234. Hershberger RE, Lindenfeld J, Mestroni L, Seidman CE, Taylor MR, Towbin JA. Genetic evaluation of cardiomyopathy--a Heart Failure Society of America practice guideline. *Journal of cardiac failure* 2009;15:83-97.
235. Kushner JD, Nauman D, Burgess D, et al. Clinical Characteristics of 304 Kindreds Evaluated for Familial Dilated Cardiomyopathy. *Journal of cardiac failure* 2006;12:422-9.
236. Richardson P, McKenna W, Bristow M, et al. Report of the 1995 World Health Organization/International Society and Federation of Cardiology Task Force on the Definition and Classification of cardiomyopathies. *Circulation* 1996;93:841-2.
237. Golbus JR, Puckelwartz MJ, Fahrenbach JP, Dellefave-Castillo LM, Wolfgeher D, McNally EM. Population-based variation in cardiomyopathy genes. *Circulation Cardiovascular genetics* 2012;5:391-9.
238. Roncarati R, Viviani Anselmi C, Krawitz P, et al. Doubly heterozygous LMNA and TTN mutations revealed by exome sequencing in a severe form of dilated cardiomyopathy. *European journal of human genetics : EJHG* 2013.
239. Yang Q, Sanbe A, Osinska H, Hewett TE, Klevitsky R, Robbins J. A mouse model of myosin binding protein C human familial hypertrophic cardiomyopathy. *The Journal of clinical investigation* 1998;102:1292-300.
240. Sarikas A, Carrier L, Schenke C, et al. Impairment of the ubiquitin-proteasome system by truncated cardiac myosin binding protein C mutants. *Cardiovascular research* 2005;66:33-44.
241. Norton N, Li D, Rampersaud E, et al. Exome Sequencing and Genome-Wide Linkage Analysis in 17 Families Illustrates the Complex Contribution of TTN Truncating Variants to Dilated Cardiomyopathy. *Circ Cardiovasc Genet* 2013.