

Titin resequencing and genetic discovery in human heart failure

Submitted to Imperial College London

for the degree of Doctor of Philosophy

by

Dr Gillian Rea

Cardiovascular Genetics and Genomics

NIHR Royal Brompton Cardiovascular Biomedical Research Unit

Imperial College Faculty of Medicine

Supervisors: Professor Stuart A Cook, Dr Paul JR Barton & Dr James Ware

February 2017

Word count: 42, 721

(Excluding bibliography and appendices)

Declaration of Originality

Everything contained within this thesis represents the work of Gillian Rea, other than where is duly acknowledged or appropriately referenced.

Copyright Declaration

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.

Table of Contents

DECLARATION OF ORIGINALITY	2
COPYRIGHT DECLARATION	2
ABSTRACT	13
ACKNOWLEDGEMENTS	14
GLOSSARY OF ABBREVIATIONS	17
1 INTRODUCTION	21
1.1 PREVALENCE AND AETIOLOGY OF HUMAN HEART FAILURE	21
1.1.1 CARDIOMYOPATHY	23
1.1.2 HYPERTROPHIC CARDIOMYOPATHY	23
1.1.3 DILATED CARDIOMYOPATHY	24
1.1.4 LEFT VENTRICULAR NON COMPACTION CARDIOMYOPATHY	24
1.1.5 PAEDIATRIC CARDIOMYOPATHY AND HEART FAILURE	25
1.2 IDENTIFYING THE GENOMIC BASIS OF DISEASE	26
1.2.1 NEXT GENERATION SEQUENCING	27
1.2.2 WHOLE EXOME SEQUENCING	29
1.3 VARIANT INTERPRETATION	30
1.3.1 DISTINGUISHING PATHOGENIC MUTATIONS FROM BACKGROUND GENETIC VARIATION ('NOISE')	30
1.3.2 CLASS OF VARIATION	31
1.3.3 POPULATION COHORTS	32
1.3.4 VARIANT CALL VALIDATION	32
1.4 APPROACHES TO PRIORITISE GENES, BASED ON THEIR PREDICTED PATHOGENICITY	33
1.4.1 PREDICTING HAPLOINSUFFICIENCY	33
1.4.2 RESIDUAL VARIATION INTOLERANCE SCORE (RVIS)	34

1.4.3	INTERPRETATION OF <i>DE NOVO</i> MUTATION	35
1.5	TYPES OF GENETIC DISEASE AND MODES OF INHERITANCE	36
1.5.1	THE NUCLEAR GENOME AND THE X CHROMOSOME	36
1.5.2	X INACTIVATION	37
1.5.3	MITOCHONDRIA	38
1.5.4	THE MITOCHONDRIAL GENOME	39
1.6	MITOCHONDRIAL DISORDERS	40
1.6.1	THE MITOCHONDRIAL RESPIRATORY CHAIN (MRC)	41
1.6.2	DIAGNOSIS OF MRC DISORDERS	42
1.6.3	INHERITANCE PATTERNS IN MRC DISORDERS	44
1.6.4	MITOCHONDRIAL DYSFUNCTION CAUSING CARDIAC DISEASE	45
1.6.5	COMPLEX 1 DEFICIENCIES	45
1.7	GENETIC DISCOVERY THROUGH PHENOTYPING	46
1.7.1	GENOTYPIC AND PHENOTYPIC HETEROGENEITY	46
1.7.2	LESSONS FROM WHOLE EXOME SEQUENCING IN RARE DISEASES	47
1.7.3	WHOLE EXOME SEQUENCING APPROACHES	48
1.8	KEY CARDIOMYOPATHY GENES	51
1.8.1	TITIN (<i>TTN</i>)	51
1.8.2	ADDITIONAL KEY DCM GENES	56
1.9	MYOCARDIAL INFARCTION COHORT	57
1.9.1	CARDIOVASCULAR MAGNETIC RESONANCE IMAGING	57
1.9.2	MYOCARDIAL INFARCTION	59
1.9.3	DIFFERENTIATING MYOCARDIAL FIBROSIS FROM MYOCARDIAL INFARCTION	60
1.9.4	LV REMODELING IN HF	61
1.9.5	MANAGEMENT OF HF AND IHD	63

2	<u>CHAPTER 2: THE ROLE OF TITIN TRUNCATING VARIANTS AS A MOLECULAR MECHANISM UNDERPINNING LEFT VENTRICULAR DYSFUNCTION IN PATIENTS FOLLOWING MYOCARDIAL INFARCTION.</u>	66
2.1	INTRODUCTION	66
2.1.1	KEY QUESTIONS	67
2.2	METHODS	68
2.2.1	STUDY SETTING AND DESIGN	68
2.2.1.3	COHORT OF SUBJECTS WITH END-STAGE ISCHAEMIC CARDIOMYOPATHY	71
2.2.1.3	HEALTHY VOLUNTEER COHORT	71
2.2.2	CARDIAC MAGNETIC RESONANCE (CMR) IMAGING	71
2.2.3	ADDITIONAL CMR ANALYSIS	72
2.2.4	TARGETED RESEQUENCING OF <i>TTN</i>	74
2.2.5	BIOINFORMATIC ANALYSIS OF TARGETED RESEQUENCING	75
2.2.6	VARIANT ANNOTATION	76
2.2.7	SANGER SEQUENCING OF <i>TTN</i> TRUNCATING VARIANTS	76
2.2.8	ADDITIONAL FILTERS APPLIED TO OTHER KEY DCM GENES	77
2.2.9	STATISTICAL ANALYSIS	77
2.3	RESULTS	77
2.3.1	QUALITY OF SEQUENCING	77
2.3.2	COHORT WITH CMR EVIDENCE OF MI: POPULATION AND CHARACTERISTICS	81
2.3.3	REGRESSION ANALYSIS	100
2.3.4	HEALTHY VOLUNTEER COHORT	106
2.3.5	END STAGE ISCHAEMIC CARDIOMYOPATHY COHORT	106
2.4	DISCUSSION	107
2.4.1	COHORT OF SUBJECTS WITH CMR EVIDENCE OF MI	107
2.4.2	SEVERE END STAGE ISCHAEMIC CARDIOMYOPATHY COHORT	109
2.4.3	STUDY DESIGN LIMITATIONS	111

2.4.4	LIMITATIONS OF PHENOTYPING	112
2.4.5	LIMITATIONS OF NGS SEQUENCE DATA	112
2.5	CONCLUSION	113
2.6	OUTLINE OF FURTHER WORK	114

3 CHAPTER 3: IDENTIFICATION OF THE GENETIC BASIS OF AN ULTRA-RARE CARDIAC PHENOTYPE: THE APPLICATION OF WHOLE EXOME SEQUENCING IN HISTIOCYTOID CARDIOMYOPATHY

3.1	ULTRA-RARE CARDIAC PHENOTYPES	115
3.1.1	INTRODUCTION TO HISTIOCYTOID CARDIOMYOPATHY	115
3.1.2	NATURAL HISTORY AND PREVALENCE OF HISTIOCYTOID CM	118
3.1.3	MULTISYSTEM INVOLVEMENT IN HISTIOCYTOID CM	119
3.1.4	FAMILIAL CASES	120
3.1.5	CLUES TO THE MOLECULAR BASIS OF HISTIOCYTOID CM	121
3.1.6	AIM	123
3.1.7	HYPOTHESIS	123
3.2	METHODS	123
3.2.1	ETHICS APPROVAL	123
3.2.2	WHOLE EXOME SEQUENCING APPROACHES	123
3.2.3	WHOLE EXOME SEQUENCING USING A 'TRIO' APPROACH	128
3.2.4	WHOLE EXOME SEQUENCING USING AN 'OVERLAP' APPROACH	129
3.2.5	MAPPING MITOCHONDRIAL DNA (MTDNA) VARIANTS	130
3.2.6	VARIANT PRIORITIZATION	131
3.2.7	SANGER SEQUENCING OF CANDIDATE GENES	131
3.2.8	IDENTIFYING VARIANTS IN CANDIDATE GENES IN PUBLICALLY AVAILABLE WES DATA	132
3.3	RESULTS	132
3.3.1	WHOLE EXOME SEQUENCING USING A 'TRIO' APPROACH	132
3.3.2	VALIDATION OF <i>NDUFB11</i> AS A NEW DISEASE GENE	137

3.3.3	WHOLE EXOME SEQUENCING USING AN 'OVERLAP' APPROACH	138
3.3.4	APPLYING A MITOCHONDRIAL DNA MODEL OF INHERITANCE	149
3.4	DISCUSSION	149
3.4.1	MERITS OF DIFFERENT WHOLE EXOME SEQUENCING APPROACHES	149
3.4.2	WES USING A 'TRIO' APPROACH: IDENTIFICATION OF <i>NDUFB11</i> AS THE CAUSATIVE GENE OF HISTIOCYTOID CM	151
3.4.3	EVIDENCE FOR PATHOLOGICAL CONSEQUENCE OF TRUNCATING VARIANTS IN <i>NDUFB11</i>	154
3.4.4	THE GENETIC BASIS OF MLS SYNDROME	156
3.4.5	DIAGNOSTIC CRITERIA FOR MICROPHTHALMIA AND LINEAR SKIN DEFECTS SYNDROME	159
3.4.6	EVIDENCE FOR A LINK BETWEEN HISTIOCYTOID CM AND MLS	161
3.4.7	DO MUTATIONS IN <i>HCCS</i> AND <i>COX7B</i> ALSO CAUSE HISTIOCYTOID CM?	162
3.4.8	A SINGLE MUTATION WITH MULTIPLE PHENOTYPES	162
3.4.9	WHOLE EXOME SEQUENCING USING AN OVERLAP APPROACH: CANDIDATE GENES IDENTIFIED	167
3.4.10	WHAT IS THE CAUSE OF HISTIOCYTOID CM IN THE OUTSTANDING CASES?	169
3.4.11	TREATMENT OPTIONS IN HISTIOCYTOID CM	171
3.4.12	THE ROLE OF WES OR WGS IN SEVERELY ILL INFANTS	174
3.5	CONCLUSION	176
3.5.1	OUTLINE OF FUTURE WORK	177
4	APPENDIX	190

Table of Figures

FIGURE 1-1 NEXT GENERATION SEQUENCING WORKFLOW	28
FIGURE 1-2 VARIANT CONSEQUENCE PREDICTIONS	31
FIGURE 1-3 MITOCHONDRIAL DNA HETEROPLASMY AND THE THRESHOLD EFFECT	41
FIGURE 1-4 THE HUMAN MITOCHONDRIAL RESPIRATORY CHAIN	42
FIGURE 1-5 WHOLE EXOME SEQUENCING APPROACHES USED TO IDENTIFY CANDIDATE GENES	50
FIGURE 1-6 THE ROLE OF <i>TTN</i> IN THE SARCOMERE	52
FIGURE 2-1 OUTLINE OF STUDY COHORTS	68
FIGURE 2-2 STUDY FLOW SHOWING INCLUSION CRITERIA FOR COHORT OF SUBJECTS WITH CMR EVIDENCE OF MI (N=335)	70
FIGURE 2-3 STANDARD 17 SEGMENT CARDIAC MODEL	73
FIGURE 2-4 CORONARY ARTERY TERRITORIES	74
FIGURE 2-5 QUALITY OF <i>TTN</i> SEQUENCING ACROSS COHORTS	79
FIGURE 2-6 CARDIAC MR IMAGES FROM COHORT OF SUBJECTS WITH CMR EVIDENCE OF MI (N=335) WITH <i>TTNTVS</i> (N=9)	82
FIGURE 2-7 SCATTERPLOT OF SUBJECTS WITH CMR EVIDENCE OF MI (N=335) SHOWING LVEF AND THE NUMBER OF INFARCTED SEGMENTS	98
FIGURE 2-8 SCATTERPLOT OF SUBJECTS WITH CMR EVIDENCE OF MI (N=335) SHOWING LVEDVI AND THE NUMBER OF INFARCTED SEGMENTS	98
FIGURE 2-9 BEESWARM PLOT OF SUBJECTS WITH CMR EVIDENCE OF MI (N=335) SHOWING LVEF AND PRESENCE OR ABSENCE OF A <i>TTNTV</i>	99
FIGURE 2-10 BEESWARM PLOT OF SUBJECTS WITH CMR EVIDENCE OF MI (N=335) SHOWING LVEF AND PRESENCE OR ABSENCE OF VARIANTS IN KEY DCM GENES	99
FIGURE 2-11 BEESWARM PLOT OF SUBJECTS WITH CMR EVIDENCE OF MI (N=335) SHOWING LVEF GROUPED BY THE PRESENCE OR ABSENCE OF <i>TTN NON-TVS</i>	100
FIGURE 2-12 OVERVIEW OF THE PROCESS OF MULTIVARIABLE REGRESSION ANALYSIS	101
FIGURE 3-1 HISTOLOGY OF HISTIOCYTOID CARDIOMYOPATHY	117
FIGURE 3-2 MACROSCOPIC FINDINGS IN HISTIOCYTOID CARDIOMYOPATHY	118
FIGURE 3-3 PEDIGREE OF FAMILY USED IN 'TRIO' APPROACH TO WHOLE EXOME SEQUENCING	125
FIGURE 3-4 HISTOLOGY OF NATIVE HEART FROM PROBAND (SUBJECT 1) WITH HISTIOCYTOID CM	128

FIGURE 3-5 SEQUENCE ELECTROPHEROGRAM OF AFFECTED CHILD SHOWING THE <i>NDUFB11</i>	
MUTATION	133
FIGURE 3-6 SEQUENCE ELECTROPHEROGRAM FROM UNAFFECTED MOTHER SHOWING NO EVIDENCE	
OF <i>NDUFB11</i> MUTATION	134
FIGURE 3-7 SEQUENCE ELECTROPHEROGRAM FROM THE UNAFFECTED FATHER SHOWING NO	
EVIDENCE OF <i>NDUFB11</i> MUTATION	134
FIGURE 3-8 SEQUENCE ELECTROPHEROGRAM FROM THE AFFECTED CHILD SHOWING THE <i>FAM135A</i>	
MUTATION	135
FIGURE 3-9 SEQUENCE ELECTROPHEROGRAM FROM UNAFFECTED MOTHER SHOWING NO EVIDENCE	
OF <i>FAM135A</i> MUTATION	135
FIGURE 3-10 SEQUENCE ELECTROPHEROGRAM FROM UNAFFECTED FATHER SHOWING NO EVIDENCE	
OF <i>FAM135A</i> MUTATION	136
FIGURE 3-11 FILTERING STRATEGY TO IDENTIFY CANDIDATE GENES FOR HISTIOCYTOID CM	
CONSISTENT WITH A MONO-ALLELIC CAUSE OF DISEASE	141
FIGURE 3-12 FILTERING STRATEGY TO IDENTIFY CANDIDATE GENES FOR HISTIOCYTOID CM	
CONSISTENT WITH A BI-ALLELIC CAUSE OF DISEASE	143
FIGURE 3-13 THE KEGG PATHWAY OF OXIDATIVE PHOSPHORYLATION (MITOCHONDRIAL	
RESPIRATORY CHAIN)	153
FIGURE 3-14 ILLUSTRATION OF <i>NDUFB11</i> SHOWING REPORTED TRUNCATING MUTATIONS AND	
ASSOCIATED PHENOTYPES	154
FIGURE 3-15 TYPICAL SKIN LESIONS AND EYE FEATURES OF PATIENTS WITH MLS SYNDROME	156
FIGURE 3-16 IDEOGRAM OF THE X CHROMOSOME WITH ANNOTATIONS OF GENES IMPLICATED IN MLS	
SYNDROME	157
FIGURE 3-17 SCHEMATIC REPRESENTATION OF X CHROMOSOME INACTIVATION IN FEMALE SOMATIC	
CELLS IN MLS SYNDROME	165
FIGURE 4-1 DEFINITION OF MYOCARDIAL INFARCTION	190
FIGURE 4-2 SCREENSHOTS OF PROFORMA USED TO CAPTURE INFORMATION ON MI COHORT	191
FIGURE 4-3 FORMAL ASSESSMENT OF MULTIPLE REGRESSION MODEL PREDICTING LEFT	
VENTRICULAR EJECTION FRACTION (LVEF)	204

Table of Tables

TABLE 1-1 FEATURES ASSOCIATED WITH HAPLOINSUFFICIENT GENES	34
TABLE 1-2 TYPES OF NUCLEAR DNA MUTATIONS WHICH CAUSE MITOCHONDRIAL DISEASE	44
TABLE 1-3 ADDITIONAL KEY DCM GENES	57
TABLE 1-4 UNIVERSAL CLASSIFICATION OF MI	60
TABLE 1-5 CONDITIONS ASSOCIATED WITH MYOCARDIAL FIBROSIS.....	61
TABLE 2-1 ADDITIONAL FILTERS APPLIED TO VARIANTS OTHER THAN <i>TTNTVS</i>	77
TABLE 2-2 CALLABILITY OF KEY DCM GENES ACROSS COHORTS.....	78
TABLE 2-3 <i>TTNTVS</i> IDENTIFIED IN EACH COHORT THAT DID NOT VALIDATE BY SANGER SEQUENCING	80
TABLE 2-4 CHARACTERISTICS OF <i>TTNTVS</i> (N=9) IDENTIFIED IN MI COHORT (N=335).....	81
TABLE 2-5 BASELINE DEMOGRAPHIC AND CLINICAL CHARACTERISTICS OF COHORT OF SUBJECTS WITH CMR EVIDENCE OF MI (N=335), GROUPED BY PRESENCE OR ABSENCE OF A <i>TTNTV</i>	91
TABLE 2-6 CMR CHARACTERISTICS OF COHORT OF SUBJECTS WITH CMR EVIDENCE OF MI, GROUPED BY THE PRESENCE OR ABSENCE OF A <i>TTNTV</i>	92
TABLE 2-7 CHARACTERISTICS OF CLINICAL PRESENTATION IN PATIENTS WITH A <i>TTNTV</i> (N=9) FROM THE COHORT OF SUBJECTS WITH CMR EVIDENCE OF MI	94
TABLE 2-8 CMR CHARACTERISTICS OF PATIENTS WITH A <i>TTNTV</i> (N=9) IN THE COHORT OF SUBJECTS WITH CMR EVIDENCE OF MI	95
TABLE 2-9 PREVALENCE OF <i>TTNTVS</i> IN SUBJECTS WITH CMR EVIDENCE OF MI, GROUPED BY LEVEL OF LVEF COMPARED TO BACKGROUND POPULATION.....	96
TABLE 2-10 MUTATION BURDEN, VARIANT TYPE AND DISTRIBUTION OF <i>TTNTV</i> IN SUBJECTS WITH CMR EVIDENCE OF MYOCARDIAL INFARCT (N=335), GROUPED BY LEVEL OF LEVEL OF LEFT VENTRICULAR EJECTION FRACTION (LVEF).....	97
TABLE 2-11 MULTIPLE REGRESSION MODEL PREDICTING LVEF	102
TABLE 2-12 MULTIPLE REGRESSION MODEL PREDICTING LVEDVI	103
TABLE 2-13 MULTIPLE REGRESSION MODEL PREDICTING LVESVI.....	104
TABLE 2-14 MULTIVARIABLE LINEAR MODELING OF THE RELATIONSHIP BETWEEN <i>TTNTV</i> GENOTYPE AND PHENOTYPE FOR LVEF, LVEDVI AND LVESVI IN SUBJECTS WITH MI	105
TABLE 2-15 <i>TTNTVS</i> IN THE HEALTHY VOLUNTEER COHORT (N=431)	106

TABLE 2-16 PREVALENCE OF <i>TTNTVS</i> IN SUBJECTS WITH END STAGE ISCHAEMIC CARDIOMYOPATHY COMPARED TO BACKGROUND POPULATION	107
TABLE 2-17 <i>TTNTVS</i> IN THE END-STAGE ISCHAEMIC CARDIOMYOPATHY COHORT (N=95)	107
TABLE 3-1 EXTRA-CARDIAC ANOMALIES REPORTED IN ASSOCIATION WITH HISTIOCYTOID CM	120
TABLE 3-2 INITIAL GENETIC INVESTIGATIONS UNDERTAKEN IN THE PROBAND.....	127
TABLE 3-3 SUMMARY FIGURES OF COVERAGE FOR WES IN EACH SAMPLE OF THE FAMILY ‘TRIO’	129
TABLE 3-4 SUMMARY FIGURES OF COVERAGE FOR WES USING AN ‘OVERLAP’ APPROACH	130
TABLE 3-5 PERCENTAGE CALLABILITY FOR MTDNA REGIONS INCLUDED IN WES.....	130
TABLE 3-6 PUTATIVE <i>DE NOVO</i> PROTEIN ALTERING VARIANTS IDENTIFIED IN THE AFFECTED CHILD	136
TABLE 3-7 CANDIDATE GENES FOR HISTIOCYTOID CARDIOMYOPATHY.....	137
TABLE 3-8 GENES (N=15) WITH RARE PROTEIN ALTERING VARIANTS IDENTIFIED IN ALL FOUR UNRELATED INDIVIDUALS AFFECTED BY HISTIOCYTOID CM	144
TABLE 3-9 SUMMARY OF VARIANTS IDENTIFIED IN ANKRD20A4.....	146
TABLE 3-10 SUMMARY OF VARIANTS IDENTIFIED IN <i>FRG1</i>	147
TABLE 3-11 SUMMARY OF VARIANTS IDENTIFIED IN IGHJ6	148
TABLE 3-12 SUMMARY OF CANDIDATE VARIANT IDENTIFIED IN MT-ND1.....	149
TABLE 3-13 MAJOR DIAGNOSTIC CRITERIA FOR DIAGNOSIS OF MICROPHthalmia AND LINEAR SKIN DEFECTS SYNDROME	160
TABLE 3-14 ADDITIONAL FEATURES REPORTED IN ASSOCIATION WITH MLS SYNDROME	160
TABLE 3-15 DETAILED COMPARISON OF THE CLINICAL FEATURES IN THE CASE OF HISTIOCYTOID CM AND A PREVIOUSLY REPORTED CASE WITH A PHENOTYPE OF MLS SYNDROME AND AN IDENTICAL <i>NDUFB11</i> MUTATION (C.262C>T; P. ARG88TER)	166
TABLE 3-16 CLINICAL DETAILS FROM PREVIOUS REPORTS OF THE M.3308T>G VARIANT.....	167
TABLE 4-1 STANDARDIZED CLINICAL DEFINITIONS USED FOR DATA CAPTURE IN MI COHORT	193
TABLE 4-2 BRU ID’S OF COHORT OF SUBJECTS WITH CMR EVIDENCE OF MI (N=335)	194
TABLE 4-3 BRU ID’S OF COHORT OF HEALTHY VOLUNTEERS (HVOLS) INCLUDED IN STUDY (N= 431)	195
TABLE 4-4 BRU ID’S OF COHORT WITH END STAGE ISCHAEMIC CARDIOMYOPATHY (N=95).....	197
TABLE 4-5 DETAILS OF AMPLICON PCR PRIMERS USED FOR SANGER SEQUENCING OF <i>TTNTVS</i>	198

TABLE 4-6 MAXIMUM, MINIMUM AND MEAN CALLABILITY OF ALL GENES (N=202) SEQUENCED USING SURESELECT AND SOLID ACROSS COHORTS.....	200
TABLE 4-7 METABOLIC INVESTIGATIONS OF SUBJECT 1 (AFFECTED CHILD WITH HISTIOCYTOID CM) SHOWED SOME MINOR NON-SPECIFIC ABNORMALITIES	205
TABLE 4-8 INVESTIGATIONS FOR MITOCHONDRIAL DNA MUTATIONS (MTDNA) IN SKELETAL MUSCLE OF SUBJECT 1 (AFFECTED CHILD WITH HISTIOCYTOID CM), SHOWED NO EVIDENCE OF ANY ABNORMALITY	205
TABLE 4-9 FURTHER HISTOPATHOLOGY, IMMUNOHISTOCHEMISTRY AND IMMUNOBLOTTING INVESTIGATIONS IN THE SKELETAL MUSCLE OF SUBJECT 1 (AFFECTED CHILD WITH HISTIOCYTOID CM) SHOWING NO SPECIFIC ABNORMALITY	206
TABLE 4-10 PRIMERS USED FOR SANGER SEQUENCING OF <i>NDUFB11</i> AND <i>FAM135A</i>	206
TABLE 4-11 THE TWO APPARENTLY <i>DE NOVO</i> PROTEIN ALTERING VARIANTS* IDENTIFIED IN THE AFFECTED CHILD (SUBJECT 1) WITH HISTIOCYTOID CM AND ABSENT IN THE UNAFFECTED PARENTS WHICH DID NOT VALIDATE, FROM THE 'TRIO' APPROACH TO WES	207
TABLE 4-12 ALL NON-SYNONYMOUS PROTEIN CODING MTDNA VARIANTS IDENTIFIED IN MTDNA GENES WHICH HAD EITHER VARIANTS IN EITHER TWO, THREE OR FOUR OF THE OVERLAP COHORT OF FOUR UNRELATED INDIVIDUALS WITH HISTIOCYTOID CM	208
TABLE 4-13 PERMISSIONS FOR FIGURES REPRODUCED WITHIN THESIS	209

Abstract

The overarching goal of the studies presented in this thesis was to study and better understand the genes and mechanisms underlying human rare and common heart muscle disease in early- and late-onset disease, respectively, using a range of genomic and informatic approaches. Although largely attributed to Ischaemic Heart Disease (IHD)-associated muscle damage, many non-IHD genes and pathways have been associated with HF, including those associated with inherited cardiomyopathies. The importance of the role of Titin (*TTN*), a major determinant of myocardial and ventricular function in both familial and “idiopathic” Dilated Cardiomyopathy (DCM), has recently been ascertained. Identifying the contribution of *TTN* variants, in a population of well phenotyped patients with Cardiac Magnetic Resonance (CMR) evidence of Myocardial Infarction (MI) and disproportionately reduced Left Ventricular (LV) function (‘dual-pathology’), was undertaken here to test the hypothesis that variants in *TTN* may contribute to heart failure in the context of IHD. In addition, Whole Exome Sequencing and applied informatics methodologies were used to prioritize a causative (new) gene as the underlying genetic aetiology of an ultra rare cardiac phenotype (Histiocytoid Cardiomyopathy). Taken together these studies show genetic approaches for studying rare variant effect in heart muscle disease of the young and in the general population and identify new genes and biology that broaden our understanding and propose possible clinical applications.

Acknowledgements

- Professor Stuart A Cook, Dr Paul JR Barton & Dr James S Ware supervised this work from conception to completion. Discussions on all aspects of this research were extremely helpful throughout. I am deeply indebted for their advice, patient encouragement and support throughout.
- The British Heart Foundation provided support with a Clinical Research Fellowship. This period of research has been a wonderful opportunity, for which I am very grateful.
- Discussions with Dr Sanjay Prasad regarding the Biobank database and Myocardial Infarction (MI) cohort were always constructive, educational and extremely enjoyable. I am grateful for his help and encouragement throughout.
- Stuart Cook assigned Myocardial Infarction (MI) patients to the Biobank database.
- The reporting clinicians performed basic Cardiac Magnetic Resonance (CMR) measurements.
- Joanna Petryka and Miguel Silva Vieira performed the additional CMR analyses as outlined in the methods section.
- Patient recruitment to the Cardiovascular BRU, patient questionnaires including follow-up data collection were undertaken by the research nurses, lead by Annashyl West.
- Gillian Rea completed the electronic IHD proforma collating data from the patient questionnaire, other patient notes and hospital records as available.
- Simon Newsome provided advice on statistical methods.
- Rachel Buchan and Alicia Wilk undertook NGS sequencing. Rachel Buchan was endlessly patient and willing with her teaching on all aspects of DNA sequencing amongst many other things.

- Sam Wilkinson provided initial training in the design of primers for polymerase-chain-reaction amplification and in dideoxy sequencing “Sanger Sequencing” and was helpful, enthusiastic and kind with his help with wet lab work. Primer design and Sanger validation of variants was performed by Gillian Rea and Sam Wilkinson.
- Gillian Rea and Rachel Buchan manually inspected variants of interest using Integrative Genome Viewer (IGV).
- The bioinformatics pipelines used here for sequence alignment, variant calling and variant annotation were established and maintained by Shibu John, Sungsam Gong, and Roddy Walsh.
- The proband (subject 1) with Histiocytoid Cardiomyopathy was identified, characterized and referred for Whole Exome Sequencing by Tessa Homfray, Jan Till, Ferran Roses-Noguer, A John Baksi, Piers Daubeney & Sanjay Prasad. Dr Michael Ashworth, Department of Histopathology, Great Ormond Street, made the initial diagnosis of Histiocytoid Cardiomyopathy on routine Histology.
- Additional cases of Histiocytoid Cardiomyopathy (subjects 2-5) cases were referred by Shane McKee, Fiona J Stewart, Victoria Murday & Robert W Taylor.
- Initial investigations in skeletal muscle of subject 1 included testing for mitochondrial DNA variants along with additional histopathology, immunohistochemistry and immunoblotting investigations; undertaken by the Mitochondrial Diagnostic Service and Neuromuscular Genetics Services, Newcastle, UK.
- Interrogation of sequence data in a cohort of SIDS cases for variants in *NDUFB11* was kindly undertaken by Dr Elijah Behr and his team.
- Prof H Gelberg provided slides of archive material from a Savannah Kitten with an identical condition to Histiocytoid cardiomyopathy.

- Clinical colleagues within the Northern Ireland Regional Genetics Service were extremely supportive in allowing me a dedicated period of research time and in welcoming me back to the department as a Consultant colleague.
- Finally, my thanks to my family (old, new and still to arrive) for the endless patience and support.

Glossary of abbreviations

ACEIs: Angiotensin Converting Enzyme Inhibitor

ACS: Acute Coronary Syndrome

ACMG: American College of Medical Genetics and Genomics

AD: Autosomal Dominant

AHA: American Heart Association

AF: Atrial Fibrillation

Array-CGH: Array Comparative Genomic Hybridization

AR: Autosomal Recessive

ARBs: Angiotensin II receptor blockers

ARVC/AVC: Arrhythmogenic Right Ventricular Cardiomyopathy/ Arrhythmogenic Ventricular Cardiomyopathy

ATP: Adenosine Triphosphate

BSA: Body surface area

CAD: Coronary artery disease

CBRU: Cardiovascular Biomedical Research Unit

CM: Cardiomyopathy

CMR: Cardiac Magnetic Resonance

CNV: Copy Number Variation

CPR: Cardiopulmonary resuscitation

CRS: Cambridge Reference Sequence

CRT: Cardiac resynchronization therapy

DCM: Dilated cardiomyopathy

DNA: Deoxyribonucleic acid

DNM(s): *de novo* mutation(s)

ECG: Electrocardiogram

EDV: End Diastolic Volume

EHRA: European Heart Rhythm Association

ESC: European Society of Cardiology

ESP: Exome sequencing project

ESV: End-Systolic Volume

ExAC: Exome Aggregation Consortium

GATK: Genome Analysis Toolkit

GTEx: Genotype Tissue Expression Project

HCM: Hypertrophic Cardiomyopathy

HI: Haploinsufficient

HF: Heart Failure

HS: Haplosufficient

IHD: Ischaemic Heart Disease

ICD: Implantable Cardiac Defibrillator

IGV: Integrative Genome Viewer

LBBB: Left bundle branch block

LGE: Late Gadolinium Enhancement

LMM: Laboratory for Molecular Medicine, Partners Healthcare

LV: Left ventricle/ventricular

LVAD: Left Ventricular Assist Device

LVH: Left Ventricular Hypertrophy

LVEF: Left Ventricular Ejection Fraction

LVNC: Left Ventricular Non-Compaction

MELAS: Mitochondrial Encephalomyopathy, Lactic Acidosis and Stroke-like episodes syndrome

MI: Myocardial Infarction

MRC: Mitochondrial Respiratory Chain

MRI Magnetic Resonance Imaging

mtDNA: Mitochondrial DNA

NGS: Next Generation Sequencing

NHLBI: National Heart Lung and Blood Institute

NSVT: Non-Sustained Ventricular Tachycardia

OMGL: Oxford Molecular Genetics Laboratory

OXPHOS: Oxidative phosphorylation

PCR: Polymerase Chain Reaction

PICU: Paediatric Intensive Care Unit

PSI: Proportion Spliced In

RC: Respiratory chain

RNA: Ribonucleic acid

RV: Right Ventricle/Ventricular

RVIS: Residual Variation Intolerance Score

PCR Polymerase Chain Reaction

rCRS: Revised Cambridge Reference Sequence

SCD Sudden Cardiac Death

SID: Sudden Infant Death

SNP: Single Nucleotide Polymorphism

SVAS: Supra Valvular Aortic Stenosis

TTN: Titin

SOLiD: Sequencing by Oligonucleotide Ligation and Detection

SVT: Supra Ventricular Tachycardia

USA: United States of America

VF: Ventricular Fibrillation

VUS: Variant of Uncertain Significance

VT: Ventricular Tachycardia

WBS: Williams-Beuren Syndrome

WES: Whole Exome Sequencing

WGS: Whole Genome Sequencing

WPW: Wolf-Parkinson-White Syndrome

XCI: X Chromosome Inactivation

1 Introduction

1.1 Prevalence and aetiology of human heart failure

Human Heart failure (HF) is a disease of epidemic proportions and a major cause of morbidity and mortality (1). It is frequently associated with cardiomyopathy due to myocardial ischaemia following myocardial infarction and muscle loss or due to chronic muscle hypo-perfusion and myocardial hibernation (2). HF is estimated to affect ~1-2% of the population in Europe and the United states (3), resulting in a disease burden in the developed world which is rivaled only by cancer (4). HF can be defined as

‘an abnormality of cardiac structure or function leading to failure of the heart to deliver oxygen at a rate commensurate with the requirements of the metabolizing tissues, despite normal filling pressures (or only at the expense of increased filling pressures)’(5).

It is a complex, progressive condition and the aetiology involves both inherited and environmental factors. The diagnosis can be difficult as many symptoms are non-discriminatory (5). HF is a complex syndrome representing a final common pathway (4), demonstration of an underlying cardiac cause is central to HF diagnosis (5). In more than two thirds of cases, the cause is attributed to Coronary Artery Disease (CAD) (6), and at least half of all cases of HF have a reduced Left Ventricular Ejection Fraction (LVEF) (5) with other cases of HF associated with grossly preserved global ventricular function (heart failure with preserved ejection fraction (HFpEF)). Several pathogenic mechanisms are implicated in HF, including ischaemia related dysfunction, increased haemodynamic overload, ventricular remodeling, excessive neurohormonal stimulation, abnormal myocyte calcium cycling, excessive or inadequate proliferation of the extracellular matrix and accelerated apoptosis (7). Genetic variants are likely to have a modifying and/or susceptibility effect in combination with many of these pathological stimuli.

Recent decades have seen a dramatic improvement in the outcomes of patients with CAD (8). In acute coronary syndromes (ACS), the widespread use of thrombolytic therapy, percutaneous coronary interventions and antithrombotic agents have all contributed to significant morbidity and mortality reductions (6). While overall survival has improved, a “HF Paradox” has emerged, whereby there has been a remarkable improvement in the prognosis of an individual cardiac condition, such as ACS, yet a growing prevalence of HF (7). This is likely to be attributable to several factors. Firstly, the issue is exacerbated by an aging population and higher prevalence of co-morbidities which confer an increased risk of CAD and congestive heart failure, such as diabetes mellitus (6), chronic renal disease and hypertension. Secondly, the slow but progressive improvements in HF prognosis may simply increase the prevalence of the condition (7). Finally, while the mortality associated with conditions such as ACS have strikingly improved; the patients are not “cured” as such and remain at risk for further episodes of ischaemic myocardial damage (7). Furthermore, the reduced mortality associated with ACS results in an increased number of individual’s with residual Left Ventricular (LV) dysfunction, undergoing progressive LV remodeling and developing HF (6).

The underlying genetic architecture of HF ranges from susceptibility to environmental causes which are modified by multiple genetic or epigenetic loci which are individually low-penetrance, through to monogenic HF syndromes which are caused largely by single underlying pathogenic mutations (4), such as occurs in inherited cardiomyopathies. Given the magnitude of the health costs involved in HF (the direct and indirect costs in the US are projected, in 2030 to be in the region of \$77.7 billion), attention is appropriately being directed to identify individuals at higher risk of HF (7).

1.1.1 Cardiomyopathy

The American Heart Association (AHA) defines cardiomyopathies as

‘a heterogeneous group of diseases of the myocardium associated with mechanical and /or electrical dysfunction that usually but not invariably exhibit inappropriate ventricular hypertrophy or dilatation and are due to a variety of causes that frequently are genetic. Cardiomyopathies either are confined to the heart or are part of generalized systemic disorders, often leading to cardiovascular death or progressive heart- failure related disability’ (9).

Typically these disorders show variable expression (manifestations of the disease may range from subtle abnormalities detected on screening, to severe heart failure or sudden death) and incomplete penetrance (where a proportion of individuals with a mutation, do not manifest the disorder). The classifications of cardiomyopathies have undergone multiple revisions in recent years and in some areas consensus has still not been reached between European and American groups. Current classification is based on relatively gross functional and structural changes (4), with molecular insights adding to the clinical classification systems, rather than replacing them (9, 10). Cardiomyopathies can be considered as primary when the sole or predominant involvement is restricted to heart muscle and as secondary where the cardiac involvement is part of a generalized systemic disorder (9). Each of the phenotypes of Hypertrophic cardiomyopathy (HCM), DCM and Left Ventricular Non-Compaction (LVNC) show extensive phenotypic overlap and each may cause HF (4).

1.1.2 Hypertrophic Cardiomyopathy

HCM is relatively common, affecting ~1:500 of the general population (9). It was the first cardiomyopathy where a specific genetic basis was ascertained (11) with the identification of mutations in cardiac beta (β) myosin heavy chain (*MYH7*) over twenty-five years ago. Since then hundreds of causative rare variants have been identified with >80% of known genetic causes occurring in two genes (*MYH7*, *MYBPC3*). Both *MYH7* and *MYBPC3*, in addition to further genes commonly implicated (*TNNT2* and *TNNI3*), all encode contractile proteins.

HCM is therefore considered a genetic disease of contractile proteins (12) or sarcomeres. A role for TTN may be emerging (13). LV hypertrophy closely resembling an HCM phenotype may also occur as a consequence of infiltrative (glycogen storage) or mitochondrial diseases and must be differentiated from that caused by sarcomeric HCM mutations (12). Molecular genetic testing for the purposes of cascade testing in families is recommended in European guidelines on HCM (14).

1.1.3 Dilated Cardiomyopathy

DCM has been associated with genetic, infectious, toxic and autoimmune causes; where no underlying aetiology is identified, cases remain “idiopathic” (9). Idiopathic DCM has a prevalence of ~1:250 (15). However, family-based studies involving clinical screening of family members have demonstrated that 20%-50% of patients diagnosed with idiopathic DCM can be shown to have familial DCM (12) suggesting a significant genetic contribution. Unlike HCM, which is considered a genetic disease of the contractile proteins (sarcomeres), unexpected genetic complexity has been shown for DCM. The gene ontology of DCM includes a number of different protein cascades which, when perturbed, will result in DCM. More than 50 genes (16) are involved including those encoding proteins of the sarcomere, Z-disc, cytoskeleton, RNA binding, ion channels, sarcoplasmic reticulum, transcription factors and mitochondria (12). All of these causes lead to the final common pathway of ventricular dilation and systolic dysfunction (9).

1.1.4 Left Ventricular Non Compaction Cardiomyopathy

The characteristic morphological findings in LVNC include a severely thickened, two layered myocardium, with numerous prominent trabeculations, and deep intertrabecular recesses (17), which are thought to occur due to failure of normal LV compaction during embryogenesis (18). There are widely varying estimates of prevalence, but in HF patients it is evident in 3% - 4% (19). Controversy persists both in differentiating normal LV

trabeculation and differentiating LVNC from DCM and HCM (18). In one study, 8.3% of normal controls (n=60) fulfilled one or more criteria for the diagnosis of LVNC (20). The difficulty in differentiating a normally trabeculated LV from LVNC is even more challenging in black patients in whom phenotypic expression of LV hypertrophy and trabeculation may differ both qualitatively and quantitatively as compared with white patients (21). Although recognized as a distinct primary cardiomyopathy by the AHA (9), the European Society of Cardiology (ESC) still consider it as an unclassified cardiomyopathy, stating that rather than a separate disease it may represent a trait shared by different cardiomyopathies (10). Advances in cardiovascular imaging, particularly Cardiac Magnetic Resonance (CMR) and more widespread availability have lead to an increase in its diagnosis, though controversy persists regarding in diagnostic criteria (22). In particular, it is unclear whether diagnostic criteria for LVNC need to be modified depending on race (21). As with DCM and HCM, autosomal dominant (AD) transmission is common. Genetic inheritance is thought to occur in at least 30%-50% of cases (19).

1.1.5 Paediatric Cardiomyopathy and Heart Failure

The most comprehensive survey in the United Kingdom and Ireland found heart muscle diseases associated with heart failure in children (below 16 years of age) had an incidence of 0.87/100,000 of the population, with a median age of presentation of one year (23). The annual incidence specifically of paediatric cardiomyopathy is between 1.13 and 1.24 cases per 100,000 children (24, 25). There are significant differences in the aetiology and clinical associations of cardiomyopathies and HF presenting in the paediatric age group than in adults, with no evidence that Titin truncating variants are significantly enriched in paediatric patients with dilated cardiomyopathy(26, 27). The aetiology of most cases of paediatric cardiomyopathy is largely unknown (9), with a prognosis, which is currently often poor (28). This group contains a large number of rare and disparate conditions; including genetic

syndromes such as Noonan syndrome, which is associated with a variety of cardiac defects including HCM (9), metabolic myopathies representing Adenosine Triphosphate (ATP) production and utilization defects that involve abnormalities of fatty acid oxidation (acyl CoA dehydrogenase deficiencies) and carnitine deficiency, as well as infiltrative myopathies, i.e., glycogen storage diseases such as Pompe disease (9).

Mitochondrial dysfunction is a leading cause (29) of cardiomyopathies presenting in infancy (the neonatal period or first year of life) and such disorders are often fatal. Mitochondrial DNA (mtDNA) mutations are causative in 10-30% of cases, but relatively few of the underlying molecular causes in nuclear encoded mitochondrial proteins had been identified until recently when Whole Exome Sequencing (WES) in an infant who died at ten months of age of HCM with combined cardiac respiratory chain complex I and IV deficiency, identified homozygous missense mutations in *AARS2*, which encode the mitochondrial alanyl-tRNA synthetase (mtAlaRS) (29). Infantile cardiomyopathy has been described as a heterogeneous “wastebasket” category and it is proposed that the identification of clinically or pathologically distinct conditions within this group is the first step in delineating distinctive conditions, their aetiology, and finally their underlying molecular defects (30).

1.2 Identifying the genomic basis of disease

The human genome project was completed in 2003, driving forward the current era of genomic discovery, with a pace which continues to accelerate, so that

‘Regardless of where medicine is practiced, genomics is inexorably changing our understanding of the biology of nearly all medical conditions’(31).

The scale and consequences of genomic variation can vary dramatically, from no phenotypic effect to life threatening disease, but may be categorized as secondary to one of three events; single base-pair changes (point variants), insertions or deletions of nucleotides from the DNA

and structural rearrangements which reshuffle the DNA sequence and therefore the order of the nucleotides (31).

1.2.1 Next Generation Sequencing

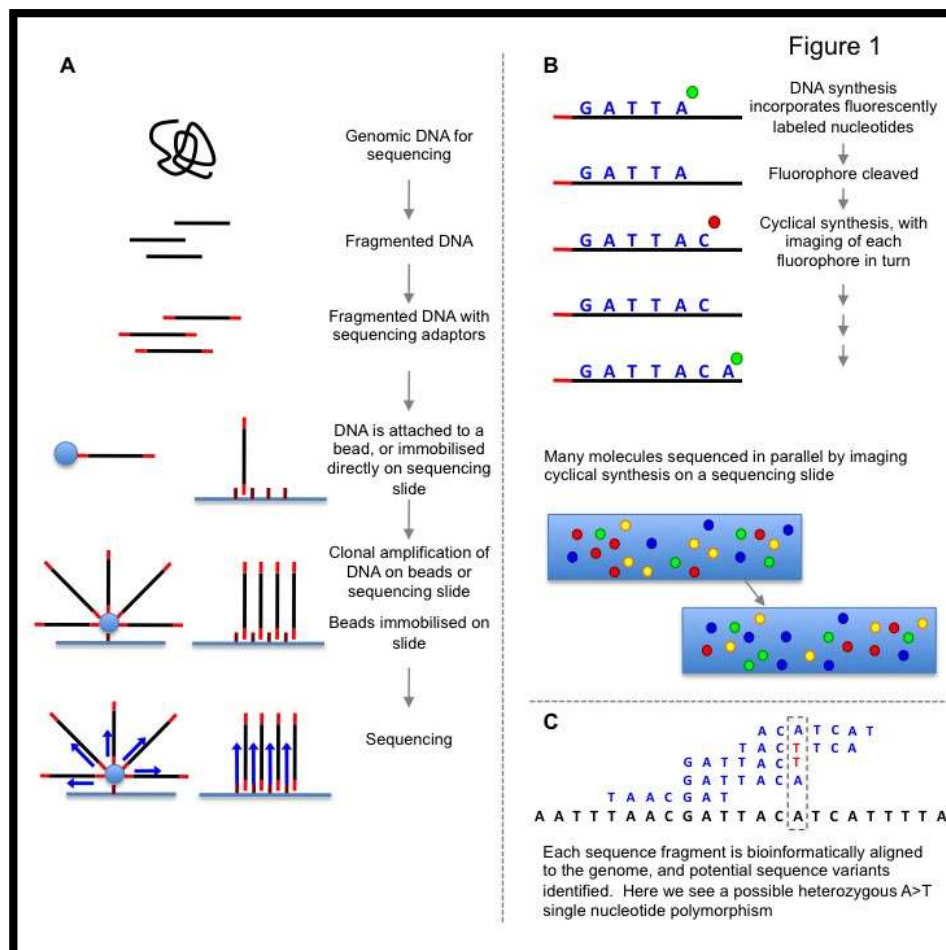
The automated Sanger sequencing method, referred to as ‘first generation’ technology had monopolized the Genomics industry for almost twenty years until the advent of newer technologies known as ‘Next Generation Sequencing’ (NGS) (32). The impact of NGS on genomics is immense and its effects translate to all areas of medicine. It has been described as the most powerful diagnostic tool since the roentgenogram (X-ray) (33). NGS encompasses several elements, which involve template preparation, sequencing, imaging, genome alignment and assembly methods (see Figure 1-1). The overall advantage of NGS over conventional methods is that it allows the inexpensive production of large volumes of sequence data (32), which is achieved through a combination of factors including sample size, speed, and accuracy.

Firstly, sample size; for Sanger sequencing several strands of template DNA are required for each base being sequenced i.e. for a 100bp sequence, several hundred copies are required as a strand that terminates each base is necessary for construction of the full sequence. In contrast, in NGS a sequence can be taken from a single strand. Multiple staged copies are then taken for contig construction and sequence validation in both types of sequencing.

Secondly, speed; there are two reasons why NGS is quicker than Sanger sequencing. The first is that in some forms of NGS the chemical reaction can be combined with the signal detection process (34). The second reason is that with Sanger sequencing, sequencing reactions are performed at a rate of just 1 to 96 at a time, while with NGS many thousands of millions or reactions may be performed simultaneously and it is therefore sometimes referred to as ‘massively parallel sequencing’(35).

Finally, accuracy: with NGS each read is amplified before sequencing and it relies on having many short overlapping reads, so that each section of DNA is sequenced many times. In addition it possible to do more repeats than with Sanger (as it is quicker and cheaper), and these additional repeats increase the coverage, increasing accuracy and reliability of NGS, even if the individual reads are less accurate for NGS (35). Overall the reductions in time, manpower and necessary reagents in NGS all contribute to the lower cost.

Figure 1-1 Next Generation Sequencing 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 immobilized on a sequencing slide. Different sequencing chemistries are employed by different NGS platforms.

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 images 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. (36). Image reproduced with permission of the rights holder, BMJ Publishing Group Ltd.

1.2.2 Whole Exome Sequencing

Historically, the discovery of genes implicated in Mendelian disease was through parametric linkage analyses of large families with multiple affected individuals, but such families are rare. More recently, NGS has been utilized to seek Mendelian disease genes in an unbiased manner by sequencing the entire exome (defined as the 1-2% of the human genome sequence which is protein-coding), referred to as Whole Exome Sequencing (WES), in smaller kindreds or unrelated individuals with a similar phenotype, who are not amenable to linkage analysis (12, 37). In a field which is rapidly changing, overall success rates are estimated between ~25% -60% (38, 39). WES has led to an explosion in new disease genes: in a case series of clinical WES in 250 patients Yang *et al* noted that their diagnoses in ~25% of their 62 patients with positive molecular results, were based on disease-gene discoveries made in the prior two years (39). In addition to the role of WES in identifying novel disease genes, in ten projects where no novel disease gene was identified, 13% of cases had variants identified which were in previously known disease genes (38). The increase in the number of genetic variants detected by WES, requires careful consideration to be given to strategies which efficiently and robustly prioritize pathogenic variants (38). Furthermore, WES (and similarly, Whole Genome Sequencing (WGS)) can generate results which are unrelated to the primary indication for sequencing and these findings may be clinically useful (40). Often described as incidental or secondary findings, the American College of Medical Genetics and Genomics (ACMG) has recommended that laboratories providing WES routinely seek and report to the referrer specific variants in a minimum set of 56 genes associated with 24 disorders for which potentially therapeutic interventions are available – termed ‘medically actionable’ (41). It is estimated that 1% to 3% of patients undergoing WES or WGS have such a finding (40).

Although positive WES findings are highly accurate, the false negative rate is variable across the genome and therefore the targeted sequencing of a panel of genes, which have been optimized for a specific disorder continues to have advantages (40).

1.3 Variant interpretation

1.3.1 Distinguishing pathogenic mutations from background genetic variation ('noise')

Advances in sequencing technology and plummeting costs of sequencing have resulted in variant identification becoming relatively straightforward, so that the main obstacle currently faced by the field is the challenge of interpreting the consequence of a given variant rather than data acquisition (42), particularly in disorders which are phenotypically and genetically heterogeneous (43, 44), such as cardiomyopathies. In recent years it has become clear that variation in individual human genomes is much greater than had been anticipated (4). Prior to the advent NGS, the true extent of genetic variation was systematically underestimated by genetic studies, largely due to ascertainment biases whereby, the focus of genetic testing was on well-studied families or individuals with clearly recognizable, relatively homogenous phenotypes. This advantaged highly penetrant mutations, which seriously affected protein function through straightforward functional consequences-that is, loss of function, gain of function, or dominant negative mutations. Such studies dramatically underestimated the real amount of genetic variation (45).

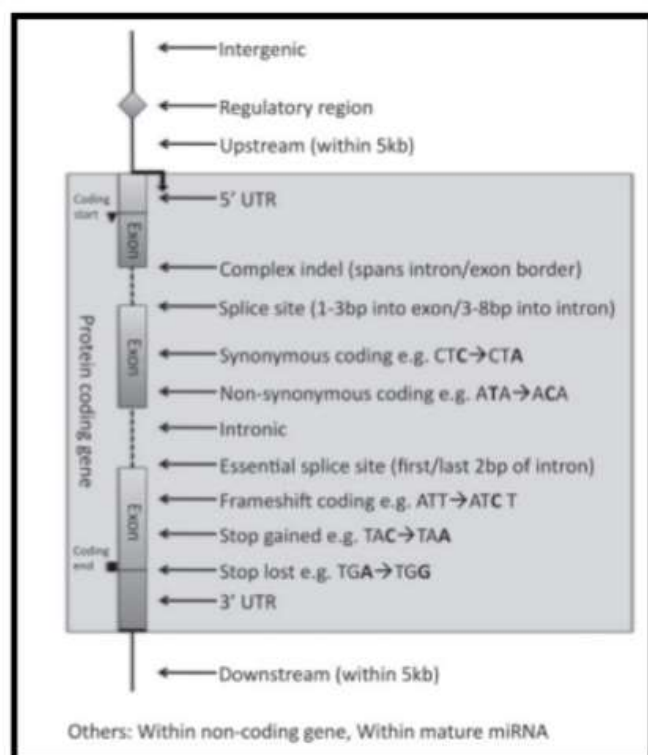
Newer strategies for detecting the genetic basis of disease, such as WES will identify multiple genetic variations, which must then be filtered on various levels in order to prioritize variants as causative for disease as none individually are evidence in itself. The identification of rare coding variants in genes for Mendelian disorders, in control subjects is well recognized, although the implications for increased disease risk are less clear (46). It is now evident that rarity alone is not sufficient to predict pathogenicity and the possibility of false attribution of pathogenicity is greatly increased in WES, due to the simultaneous testing of

thousands of genes (40). As it has become evident that the prevalence of putative causative rare variants may be at a prevalence, which is far in excess of the prevalence of disease (46), attention has focused on methods and tools to aid assessments of causality.

1.3.2 Class of variation

Initial prioritization for potentially casual variants may be on the basis of predicted impact of a variant on protein structure and function, such as nonsense mutations (stop mutations), frame-shifting mutations, mutations in the canonical (essential) splice sites or other non-synonymous variants (see Figure 1-2). The effectiveness of such a strategy, is not solely because it reduces the number of candidate variants, but because rare non-synonymous variants are intrinsically more plausible disease candidates than common, synonymous or non-coding variants (42).

Figure 1-2 Variant consequence predictions



Consequence types predicted by Ensemble in the context of transcript structure. The other types shown apply to non-protein coding genes (47). Image reproduced with permission of the rights holder, Oxford University Press.

1.3.3 Population Cohorts

For severe Mendelian disorders, prioritization assumes that the mutation has a large effect size, and is therefore novel or very rare in the general population. At the DNA sequence level any two individuals are 99.6% identical (48), but due to the size of the genome, this difference reflects differences in around 24 million bp (31). Multiple large population cohorts are now available and greatly enhance the power of determining how rare a variant may be. They include:

1. The National Heart, Lung, and Blood Institute Exome Sequencing Project (NHLBI-ESP): undertaken in 2009, has made available the exome sequences from more than 6500 individuals with cardiovascular, lung, or blood phenotypes (49).
2. The 1000 Genomes Project: This involved the genomes of 2,504 individuals from 26 populations and included a combination of low-coverage whole-genome sequencing, deep exome sequencing, and dense microarray genotyping (50).
3. The Exome Aggregation Consortium (ExAC): involving exome data from ~60,000 individuals from a range of disease cohorts including schizophrenia, diabetes heart disease and inflammatory bowel disease, but excluding cases of known severe paediatric disease (51).
4. A human mitochondrial genome database is available (<http://www.mitomap.org/>) which includes variants in human mitochondrial DNA, from 30589 human mitochondrial DNA sequences, both from published and unpublished data.

1.3.4 Variant call validation

Where a variant has been identified through NGS sequencing pipelines then consideration needs to be given to variant validation. A useful adjunct or preliminary step to Sanger sequencing is human review using Integrative Genomics Viewer (IGV) (<http://www.broadinstitute.org/igv/>). This visualization tool enables visualization and

interactive exploration of NGS sequence alignment and genome annotation viewing (52), and complements computational approaches to variant calling (53). Variant call validation by visual inspection in this manner can be an efficient and powerful tool, eliminating many false positives without the need for Sanger sequencing and aiding in confirmation of true findings (53). Homologous and repetitive regions, particularly those from pseudogenes and segmental duplications, may result in lack of coverage and alignment difficulties and therefore incorrect variant calls (54). Thus, IGV is especially useful in diagnosing misalignments, particularly in repeat regions (53). In addition, when multiple variants in the same gene are identified, the phase of the variants (i.e. on the same chromosome –in *cis* or on the homologous chromosome-in *trans*) may influence the interpretation, especially for autosomal recessive traits. By using manual visualization with IGV, if variants are within the same NGS fragment, then the phase may be determined without the need for parental samples (54).

1.4 Approaches to prioritise genes, based on their predicted pathogenicity

Available tools, which may aid the interpretation of sequence data, include those, which look at the genes in which those variants are found. A number of tools have emerged in recent years which aim to identify genes in the human genome which are most sensitive to mutational changes, as these genes which appeared to be most intolerant of variation genes may be judged most likely to contribute to disease (55, 56).

1.4.1 Predicting Haploinsufficiency

Where a single functional copy of a gene is insufficient for the maintenance of normal cellular function then the gene is termed Haploinsufficient (HI). Hundreds of HI genes have been identified and these represent a major cause of dominant disease (57). Through the systematic identification of genes which were known to be unambiguously and repeatedly comprised by CNV among 8,458 apparently healthy individuals; Huang *et al* (57) compiled a map of 1,079 haplosufficient (HS) genes and contrasted the genomic, evolutionary,

functional, and network properties between these HS genes and known HI genes (57). Key differences are highlighted in Table 1-1.

Table 1-1 Features associated with Haploinsufficient genes

- | |
|---|
| <ul style="list-style-type: none"> • Genes are typically longer, with more promoters • Coding sequences are more conserved • Genes exhibit higher levels of expression during early development and greater tissue specificity • Within a probabilistic human functional interaction network; HI genes have more interaction partners and greater network proximity to other known HI genes |
|---|

Features often associated with haplosufficient genes as compared to haplosufficient genes (57).
HI, Haploinsufficient; HS, Haplosufficient.

On the basis of these differences Huang *et al* built a predictive model and annotated 12,443 genes with their predicted probability of being HS. These predictions were validated by confirming that genes with a high-predicted probability of featuring haplosufficiency are enriched among genes which had previously been implicated in either dominant human diseases or alternatively among genes resulting in abnormal phenotypes in heterozygous knockout mice. Gene-based haplosufficiency predictions were then converted into haplosufficiency scores for genic deletions, which are superior at discriminating between pathogenic and benign deletions over conventional means, namely consideration of the deletion size or number of genes deleted. Predictions of haplosufficiency aid the prioritization and interpretation of novel loss-of-function variants and genes for further investigation (57).

1.4.2 Residual Variation Intolerance Score (RVIS)

Utilizing WES data of allele frequency from 6503 individuals from the NHLBI- ESP, Petrovksi *et al* developed a gene-based intolerance scoring system known as Residual Variation Intolerance Score (RVIS). This score ranks genes ‘on the basis of the strength and consistency of purifying selection acting against functional variation in the gene’ in effect, in terms of whether they ‘have relatively more or less functional genetic variation than expected

based on the apparently neutral variation in the gene' (56), which standardizes for gene size and mutational rate. All genes are ranked in order from most intolerant to least. Where $S=0$ then the gene has the average number of common functional variants given its total mutational burden. Where a gene has a positive score ($S \geq 0$), then it has more common functional variation, and a gene with a negative score ($S < 0$), has less and is referred to as 'intolerant'. This score was shown to correlate well with genes already known to cause Mendelian diseases ($P < 10^{-26}$) (56).

1.4.3 Interpretation of *de novo* mutation

The dramatic increase in the number of individuals undergoing sequencing in recent years is inevitably leading to identification of *de novo* mutations (DNMs) in genes, which have occurred by chance. The background rate of DNMs varies considerably between genes and is a major complicating factor in the interpretation of observed DNMs. Where DNMs are associated with a particular disease, then we would expect that genes related to disease should contain more DNMs than expected by chance (55).

Samocha *et al* developed a statistical model of DNMs, which establishes a framework for the evaluation of the rate of DNM on a per-gene basis, globally and by gene set. This model was then used to both predict the expected amount of rare standing variation per gene and also to identify which genes are significantly and specifically deficient in functional variation. It is hypothesized that variants in these genes are more likely to be deleterious (55).

Accurate estimation of the expected rate of DNM in a gene requires a precise estimate of each gene's mutability. Sources of influence on the rate of mutations in genes include not just gene length, but also local sequence context (58). Based on this knowledge, gene-specific probabilities for different types of mutation: synonymous, missense, nonsense, essential splice site and frameshift were estimated (55). There is a reasonable correlation between this metric of genes under selective constraint (based on the absence of rare functional variation)

and Petrovski's RVIS score, which considers the pattern of functional sequence variation (rare and common variants) (55).

1.5 Types of genetic disease and modes of inheritance

Mendelian diseases are commonly understood to be explained, principally by a single disease locus and are therefore often termed monogenic disorders; when familial, they exhibit patterns of Mendelian inheritance (autosomal, X-linked, dominant or recessive); such disorders are due to mutations in the nuclear genome. They may be further considered as syndromic or non-syndromic, where a syndrome is a recurring, recognizable pattern of abnormalities, which are characteristic of a specific disorder or disease.

Mendelian inheritance implies that half of the genetic material in the zygote is derived from each parent (59). An additional important cause of genetic disorders is disease resulting from mutations in the mitochondrial genome. Mitochondrial inheritance is described as maternal or matrilineal (60) as mtDNA is inherited only from the mother.

1.5.1 The Nuclear Genome and the X Chromosome

In addition to 22 pairs of autosomes, females possess two X chromosomes, males an X and a Y chromosome. The X and Y-chromosomes in humans (and all mammals) are believed to have diverged from an identical pair of ancestral chromosomes. During evolution, structural rearrangements successively limited recombination, providing independent evolution of each chromosome, eventually resulting in a Y chromosome which contains very few genes (61). Within the Y chromosome there is a psuedoautosomal region where genes show equivalent dosage with two active copies in males and females. In females, who normally have two X chromosomes, dosage compensation is achieved by inactivation of one of the X chromosomes in somatic cells.

Normal females therefore are mosaic for two cell populations, whereby each expresses the allele from one X chromosome or the other. This means that in female carriers of an X-

linked recession mutation, on average, in around 50% of cells the normal allele on the X chromosome will be expressed, whereas in the other 50% of the cells the abnormal allele is expressed. This balance is typically sufficient to ensure enough functionally normal cells so that females largely escape the clinical effects of an X-linked disease (62).

The mammalian X chromosome consists of ~ 160Mb of DNA and is estimated to contain ~1400 genes (including non-coding RNA) (63). X chromosome abnormalities are estimated to affect 1 in 650 live births (61) and it is evident from religious texts that the existence of X-linked disorders e.g. Colour blindness in humans has been recognized for many centuries (64). Of 316 phenotypes which show a distinctive X-linked inheritance pattern the molecular basis is currently known for only 187 of these (63). Although the X-Chromosome contains only 4% of all human genes, ~10% of all diseases with a Mendelian pattern of inheritance have been assigned to the X chromosome (65).

1.5.2 X Inactivation

Most of the genes on one of the X chromosomes in women are 'silenced' as a result of X inactivation, a process recognized by M Lyon over fifty years ago, whereby one of the X chromosomes become transcriptionally inactive in early embryonic life (66). Beginning at the time of late blastocyst or early gastrulation, one of the two X chromosomes, independent of the parental origin, is inactivated and this differential activity of the X chromosomes is then stably transmitted to all the descendants of single cells which gives rise to cellular mosaicism. Where the inactivation of the parental alleles is unequal (known as skewing) this can be the result of two different mechanisms. In the first, stochastic factors can cause non-random X Chromosome Inactivation (XCI) in the early embryo, particularly if the pool of precursor cells is limited. Alternatively, cell selection down-stream of the X-inactivation process causes secondary or acquired skewing (67). Evolution of the X chromosome has been actively altered by X chromosome inactivation with stringent regulation that needs gene

content to remain highly conserved (61). More recently, it has been appreciated that ~15% of X-linked genes escape X-inactivation to some degree and are expressed from both X chromosomes (the active and inactive), while a further 10% of X-linked genes exhibit variable patterns of inactivation, with expression from the inactive X chromosomes to at least some extent (61). While X-linked genes were recognized as likely candidates to contribute to sexually dimorphic traits and to clinical manifestations of disease in patients with abnormal X chromosomes, this variability in X inactivation suggests a dramatic degree of expression heterogeneity among females (61).

In addition to the random nature of X-inactivation, some deleterious X-linked mutations may cause a growth disadvantage for cells containing the mutation (on the active chromosome) and this may impact on cell proliferation and viability (68). Studies have shown this to be a feature of X-linked mental retardation (MR) (68) and immune-deficient syndromes (69). Sex chromosome mosaicism may therefore help to prevent the development of phenotypic manifestations and partially explain the high degree of inter- and intra-familial phenotypic variability seen in X-linked disorders.

1.5.3 Mitochondria

Normal function of human cells requires energy in the form of Adenosine Triphosphate (ATP), which is primarily generated by mitochondria (70) through the process of oxidative phosphorylation (OXPHOS). Mitochondria (from the Greek *mitos*- meaning thread like and *khondros*- meaning granule) are bacterium-sized organelles found in all nucleated cells (71). Additional key roles of mitochondria include adaptive thermogenesis, ion homeostasis, innate immune responses, the production of reactive oxygen species and programmed cell death (apoptosis) (70).

1.5.4 The Mitochondrial Genome

Mitochondrial DNA (mtDNA) is a circular molecule of only 16.6kbs (71), largely devoid of introns, with ~80% encoding proteins, meaning that the majority of mtDNA mutations potentially have functional consequences (59). Unlike the diploid nuclear genome, which has two homologous copies of each chromosome, the mitochondrial genome is polyploid, containing one to ten identical molecules of DNA within its matrix. This variable copy number has important implications for phenotypic expression of a mutation, referred to as heteroplasmy (see Figure 1-3). Generally, the higher the percentage of mutated mtDNA, the more severe the phenotype of the disorder. Typically, disease manifestations will be apparent where at least 70% of mtDNA is mutated (72). In addition to 13 genes encoding proteins, which are all components of the OXPHOS system, mtDNA encodes for 22 transfer RNAs and two ribosomal RNAs, which are used for translation of the 13 proteins giving a total of 37 mitochondrial genes (71). Proteins encoded by nuclear DNA are involved in controlling replication, transcription, translation and repair of mtDNA (70). The mutation rate in mtDNA genes is 10-20 fold higher than nuclear DNA genes, resulting from a lack of histone protection, an insufficient repair mechanism and the highly compact structure of the mtDNA(72). Reactive Oxygen Species (ROS) which are generated throughout the lifespan of an organism are likely to play a role in the higher mutation rate, both because mtDNA is located in the organelle which generates the most ROS and due to the relative lack of protection mechanisms of mtDNA described above(73).

The human mtDNA genome was first sequenced in 1981 (74), and this version is referred to as the Cambridge Reference Sequence (CRS). It was subsequently upgraded to the revised version known as rCRS in 1999 (75) by re-sequencing the same sample to remove the initial sequencing errors. (76). These sequences represent an arbitrary extant sequence, which was selected for notational purposes, but are frequently misunderstood as a wild type or consensus sequence (76). The rCRS has been incorporated into the most recent version of

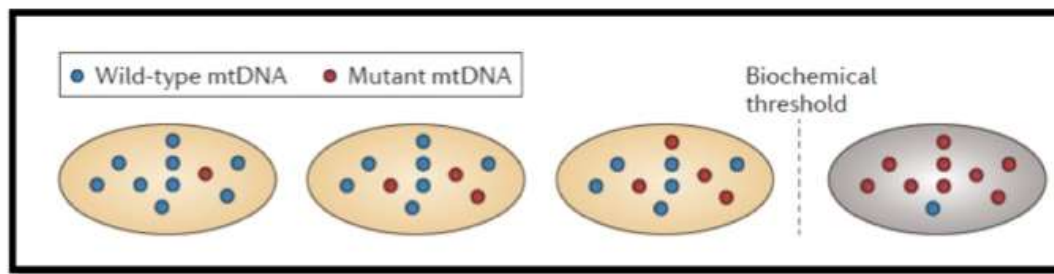
the human genome (hg38), but where mtDNA has been sequenced and bioinformatic tools have generated variations relative to a reference sequence different from rCRS, then an additional automatic tool is needed that translates this output into standard rCRS information before this information is publishable (76). Typically each individual harbors an average of 30 mtDNA variants when compared to the CRS, though this is dependent upon ethnic origin and some individuals may have over 100 variants (60).

1.6 Mitochondrial Disorders

Mitochondrial disorders were first recognized by Luft *et al* over 50 years ago, with a description of a young woman with euthyroid hyper- metabolism and muscle weakness in whom muscle biopsy and enzyme analysis revealed an uncoupling of mitochondria (77). They are now recognized as a clinically heterogeneous group of disorders with a wide range of clinical expression. They are caused by pathological dysfunction of the final common pathway of energy metabolism, oxidative phosphorylation (OXPHOS) i.e. ATP synthesis by the oxygen consuming Mitochondrial Respiratory Chain (MRC) and are therefore also known as MRC or RC disorders.

The process of oxidative phosphorylation is functional before birth and therefore mitochondrial disorders may present early in life. All age groups are affected (78). All human cells except mature red blood cells (which rely exclusively on anaerobic metabolism) contain mitochondria, typically hundreds to thousands in each cell. The number is correlated with cellular respiratory demand, with highly energetic cells containing more. Cells may contain a mixture of mutant and normal mtDNA (72) which affect disease expression (see Figure 1-3).

Figure 1-3 Mitochondrial DNA heteroplasmy and the threshold effect



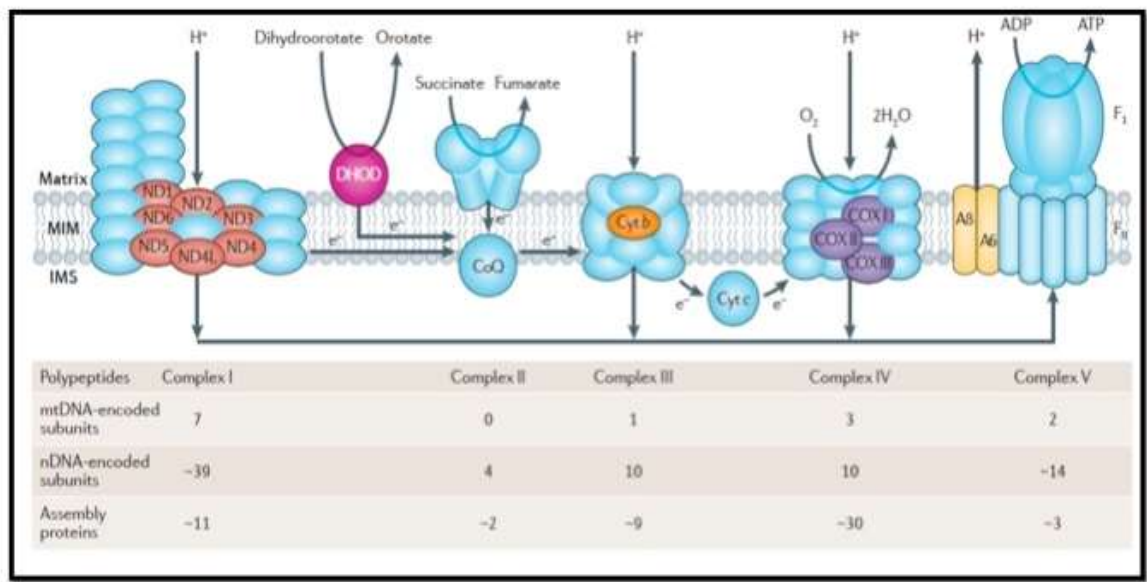
Mitochondrial DNA (mtDNA) mutations that have occurred within approximately three human generations are usually heteroplasmic, and the same cells can contain varying proportions of mutated and wild type mtDNA. If a mutation is pathogenic, the cell can usually tolerate a high percentage level of this variant before the biochemical threshold is exceeded and a defect in the respiratory chain is detected. Typically, this threshold is >80%, suggesting that most mtDNA mutations are haplosufficient or recessive. (79). Image reproduced with permission of the rights holder, Nature Publishing Group.

The clinical and genetic heterogeneity of mitochondrial disease, mean that the prevalence has proven difficult to ascertain. Gorman *et al* comprehensively assessed the prevalence of all forms of adult mitochondrial disease (including both nuclear and mtDNA pathogenic mutations) and established the minimum prevalence rate for mtDNA was 1 in 5,000 (20 per 100,000) and nuclear mutations were responsible for clinically overt adult mitochondrial disease in 2.9 per 100,000 adults. Combined this suggests a prevalence of adult mitochondrial disease affecting ~1 in 4,300 (80).

1.6.1 The Mitochondrial Respiratory Chain (MRC)

The MRC is responsible for producing ATP and consists of five enzymatic components (known as complexes I-IV) which each contain multiple subunits organized within the inner mitochondrial membrane (see Figure 1-4). Complexes I, III, IV, and V are encoded by both nuclear and mitochondrial genes (78) and this bi-genomic control of the mitochondrial proteome is a key feature of mitochondrial biology (81). In addition to their principle function of ATP synthesis in OXPHOS, mitochondria are involved in other essential cellular processes including calcium signaling, apoptosis, and generation of reactive oxygen species (ROS) (81).

Figure 1-4 The Human Mitochondrial Respiratory Chain



Oxidative phosphorylation (OXPHOS) complexes. The system is comprised of complex I (NADH dehydrogenase-CoQ reductase), complex II (succinate dehydrogenase), complex III (ubiquinone-Cyt c oxidoreductase), complex IV (COX) and complex V (ATP synthase; F₀ and F₁ denote the proton-transporting and catalytic sub-complexes, respectively), plus two electron carriers, coenzyme Q (CoQ: also called ubiquinone) and Cyt c. Complexes I-IV pump NADH- and FADH₂-derived protons (produced by the tricarboxylic acid (TCA) cycle and the beta-oxidation 'spiral') from the matrix across the mitochondrial inner membrane to generate a proton gradient while at the same time transferring electrons to molecular oxygen to produce water. The proton gradient, which makes up most of the mitochondrial transmembrane potential, is used to do work by being dissipated across the inner membrane in the opposite direction through the fifth complex (ATP synthase), thereby generating ATP from ADP and free phosphate. Complexes I, III, IV and V contain both mtDNA and nuclear DNA (nDNA)-encoded subunits, whereas complex II, which is also part of the TCA cycle, has only nDNA-encoded subunits. Note that CoQ also received electrons from dihydroorotate dehydrogenase (DHOD), an enzyme of pyrimidine synthesis. Polypeptides encoded by nDNA are in blue (except DHOD, which is in pink); those encoded by mtDNA are in colours corresponding to the colours of the polypeptide-coding genes. The 'assembly' proteins are all nDNA-encoded. IMS, intermembrane space; MIM, mitochondrial inner membrane (71). Image reproduced with permission of the rights holder, Nature Publishing Group.

1.6.2 Diagnosis of MRC disorders

MRC disorders have been increasingly recognized in recent years, with the realization that they account for a large variety of clinical symptoms in childhood. Often the diagnosis will not be considered when the first symptoms occur, but becomes easier to recognize when second, seemingly unrelated symptoms are observed (82). However, the clinical and genetic heterogeneity of mitochondrial disease, mean that the prevalence has proven difficult to ascertain. In one study Gorman *et al* comprehensively assessed the prevalence of all forms of adult mitochondrial disease (including both nuclear and mtDNA pathogenic mutations) and established the minimum prevalence rate for mtDNA was 1 in 5,000 (20 per 100,000) and nuclear mutations were responsible for clinically overt adult mitochondrial disease in 2.9 per

100,000 adults. Combined this suggests a prevalence of adult mitochondrial disease affecting ~1 in 4,300 (80). In a different epidemiological based study in the north east of England mtDNA defects are estimated to have caused disease in 6.57 per 100,000 individuals in the adult population of working age (aged >16-<60 years for female, <65 years for male)(83). Overall, common putative pathogenic mtDNA mutations have been identified in ~1 out of 300 of the general population, therefore any sequencing result must be interpreted in the correct clinical context (84).

Traditional evaluations for MRC disorders are not trivial and in some cases there are no universally accepted diagnostic criteria (85). Typically, diagnosis involves metabolic screening protocols for determination of plasma levels of lactate, pyruvate, and ketone bodies and their molar ratios, as indices of the oxidation/reduction status in cytoplasm and mitochondria, respectively (82). Independent to metabolic screening, biochemical defects evident from polarographic and spectrophotometric studies may provide clues to the diagnosis, the tissue most suited for investigation is typically that which clinically expresses the disease (which may not be readily accessible) (82). The very marked clinical and genetic heterogeneity characteristic of mitochondrial disorders leads to appreciable diagnostic challenges, including patients being referred to a wide range of specialists depending on presenting features and the relative rarity meaning few centers with significant experience. Diagnosis is further hampered by the involvement of two genomes (85), although early mutation analysis may be preferable to metabolic and biochemical testing.

Using mitochondrial proteome analysis, over 1158 genes encoding mitochondrial proteins have been identified in humans (86). Mitochondrial dysfunction can arise from a mutation in one of these genes (causing a primary mitochondrial disorder) or from an outside influence on mitochondria (causing a secondary mitochondrial disorder). Causes of secondary disorders include viral infections and off-target drug effects. To date, mutations in 228 protein-encoding nDNA genes and 13 mtDNA genes have been linked to a human disorder

(70). Nuclear DNA mutations causing mitochondrial disease may be considered in four main categories (see Table 1-2).

Table 1-2 Types of Nuclear DNA mutations which cause mitochondrial disease

Category	Comment
Mutations in genes encoding respiratory chain proteins	
Mutations in ancillary proteins	Mutations in nuclear genes encoding proteins needed for the proper assembly or function of the respiratory chain proteins
Defects of intergenomic signaling	Nuclear gene defects causing multiple deletions or depletion of mitochondrial DNA
Disorders of lipid milieu	Mutations affecting cardiolipin, an integral part of the inner mitochondrial membrane)

Nuclear DNA mutations causing mitochondrial disease may be considered in four categories. Such mitochondrial diseases clinically resemble those caused by mtDNA mutations but exhibit Mendelian inheritance.

1.6.3 Inheritance patterns in MRC disorders

Many mitochondrial diseases are inherited. As the respiratory enzymes have a unique double genetic ('bi-genomic') origin (involving both nuclear DNA and mtDNA), mutations can be located in either (59) and therefore any mode of inheritance is possible. In sporadic cases mtDNA or nuclear gene mutations may be causative. In inherited cases there may be maternal transmission of mtDNA mutations, autosomal recessive, autosomal dominant or X-linked inheritance. mtDNA mutations may include mtDNA genes, which directly encode respiratory chain proteins or mutations in mitochondrial transfer RNA or ribosomal RNA or large-scale deletions affecting the synthesis of multiple mitochondrial genes.

Following the sequence and annotation of the human mitochondrial genome (74), the first disease causing mutations in mitochondrial DNA were described in 1998 (87, 88)

For many years it was assumed that MRC disorders solely originated from mutations of mtDNA as only mutations or deletions of mtDNA were identified (82). Although mtDNA mutations account for only 10-15% of cases of paediatric mitochondrial disease (82) they are the commonest cause of mitochondrial disease in adults, identified in ~70% of patients (81)

In 1995 the first mutation of a nuclear gene giving rise to a MRC disorder was reported (89) and since then mutations in nuclear genes affecting the MRC have been identified at an escalating pace (71); they are expected to underlie the vast majority of MRC disorders (82). Correct functioning of the MRC requires not only the subunits of each complex of the respiratory chain proteins but also multiple ancillary proteins involved in the many stages of proper assembly and function of the respiratory chain proteins including genes involved in inter-genomic signaling between mtDNA and nuclear DNA, holoenzyme biogenesis, including transcription, translation, chaperoning, addition of prosthetic groups, assembly of the proteins and enzyme involved in mtDNA metabolism. The proteins involved in all of these processes are not yet known (82).

1.6.4 Mitochondrial dysfunction causing cardiac disease

Mitochondrial disorders are typically multisystem disorders and affect organs with the highest demand for ATP (namely cardiac and skeletal muscle, the central nervous system and eyes) (78). The heart is one of the most frequently affected organs in mitochondrial disorders, due to its high aerobic energy demand. Cardiac manifestations are diverse and may overlap; they include cardiomyopathy, arrhythmias, heart failure, pulmonary hypertension, pericardial effusion, coronary heart disease, autonomous nervous system dysfunction, congenital heart defects and dilation of the aortic root, or sudden cardiac death. The leading forms of cardiomyopathy associated with MRC disorders are HCM, DCM and LVNC (90). When the heart is affected the overall survival of children with mitochondrial disorders decreases substantially (91).

1.6.5 Complex 1 Deficiencies

Complex 1 is the first and largest enzyme of the MRC complexes, involving seven mitochondrial encoded and 38 nuclear encoded subunits, in addition to a number of assembly factors implicated in the correct biosynthesis of complex 1 within the inner mitochondrial

membrane (85). See Figure 1-4. Complex 1 deficiency is the most frequent cause of mitochondrial disorders presenting in childhood, accounting for up to 30% (85). Common clinical presentations currently recognized include multisystem disorders such as Leigh syndrome (a form of fatal congenital lactic acidosis) and Mitochondrial Encephalomyopathy, Lactic Acidosis and Stroke-like episodes syndrome (MELAS), Single organ involvement e.g. with HCM or isolated optic neuropathy are also recognized (85). Overall, HCM is found as a major clinical feature in 19% of cases of complex 1 deficiency (85). The underlying molecular abnormality is known in the minority of cases, usually in one of the genes (nuclear or mtDNA) encoding the structural subunits of the complex (92). Fassone *et al* report that ~25% of cases with a complex 1 deficiency have mtDNA mutations with a further 25% having mutations a nuclear subunit or in one of the known assembly factors (85).

1.7 Genetic discovery through phenotyping

1.7.1 Genotypic and Phenotypic Heterogeneity

The large majority of Mendelian disorders have significant phenotypic heterogeneity; with variability between individuals, within families and between families in both expressivity and penetrance of phenotypic manifestations (33). Three of the key determinants suggested to explain phenotypic variability include firstly, environmental factors; which may play a major role in disease manifestations e.g. in individuals with Phenylketonuria, an autosomal recessive metabolic disorder, the clinical effects can be largely ameliorated by lifelong exclusion of phenylalanine from the diet. The second factor is genetic background; whereby additional genes modify the clinical outcome of the genetic mutation and lastly there is stochasism or chance (93), although this is likely a gross simplification of the true complexity underlying penetrance and expressivity of disease-associated variants.

Recently Lu *et al* has referred to the ‘promiscuity of genotype-phenotype association’ (45) regarding to the high level of allelic heterogeneity (different mutations in one gene) and locus

heterogeneity (mutations in different genes) associated with even simple Mendelian diseases (45) which is becoming increasingly apparent. Studies of paediatric diseases such as Kabuki syndrome and Schinzel-Giedion syndrome have revealed that allelic combinations of missense, nonsense, and compound heterozygous mutations within different genes can have comparable functional effects resulting in overlapping clinical phenotypes. While in diseases such as laminopathies allelic heterogeneity produces diverse phenotypic outcomes due to the specific functional effects of each particular variant, in different tissues (45).

1.7.2 Lessons from Whole Exome Sequencing in rare diseases

While Mendelian diseases may be individually rare, they are collectively common with eight percent of people identified as having a genetic disorder before reaching adulthood (94).

This translates in the USA alone to 25 million people affected by a Mendelian disorder, many of which are associated with morbidity, mortality and a high economic burden (95). Chong *et al* calculated that as of February 2015, 2,937 genes underlying 4,163 Mendelian phenotypes had been identified, leaving the genetic basis for ~50% of all known Mendelian phenotypes still to be elucidated, in addition to the as yet unrecognized Mendelian disorders (95).

For disorders where the genetic basis remained unknown, WES, which refers to the targeted sequencing of the subset of the human genome that is protein coding, has proved a powerful and cost-effective tool (96). Since the first report in 2009, WES (37) has been used to identify casual alleles for dozens of Mendelian disorders. In addition it has been suggested that as WES becomes more widespread, and it evolves to characterize more patients with atypical presentations of known genetic diseases then the phenotypic spectrum associated with genetic disorders will expand (39).

Bamshad *et al* suggest

“solving all Mendelian disorders should be an imperative for both the human genetics community and funding agencies, as such discoveries will be of enormous service to

families while also providing novel entry points for the investigation of the mechanism underlying the development of disease” (96).

One such initiative is the Deciphering Developmental Disorders (DDD) study, which involves patient recruitment from all 24 regional genetics services in the UK and Ireland of children with undiagnosed developmental disorders together with their (largely) unaffected parents. They have performed exome sequencing with adjunct microarray analysis and achieved a diagnostic yield of 27% among 1133 previously investigated yet undiagnosed children with developmental disorders (97). Such studies where there is an emphasis on detailed and systematic phenotyping are among those most likely to yield results. It is predicted that

“the value of NGS can only be optimally exploited if novel approaches to detailed phenotyping are integrated with NGS results”(33).

1.7.3 Whole Exome Sequencing approaches

The unbiased approach of Whole Exome Sequencing (WES) has successfully been applied to determine the underlying molecular aetiology of multiple Mendelian disorders in recent years (98). It has revealed Mendelian disease in general and sporadic disease especially to be commonly caused by rare *de novo* mutations (99). A *de novo* mutation refers to the presence of a mutation, which is present in an individual but absent from both parents. In recent years it has become clear that the per-generation mutation rate in humans is high. It is hypothesized that *de novo* mutations may compensate for allele loss which results from markedly reduced fecundity which is common in neurodevelopmental and psychiatric diseases (disorders in which the mutational target is large and comprise of many genes), resolving a major paradox in evolutionary genetic theory (100). In each generation, an average newborn is estimated to have acquired 50 to 175 *de novo* point mutations in his or her genome (101-103). Such spontaneous germ-line mutations can have serious phenotypic consequences where they affect functionally relevant bases in the genome (100). The

average exome of a newborn is expected to have up to three *de novo* protein-coding changes (38, 101, 102).

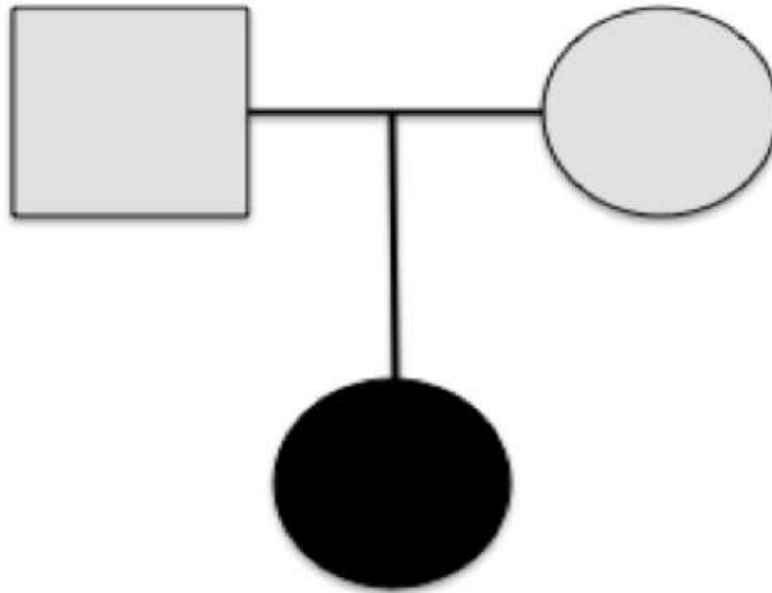
Where a disorder occurs mostly sporadically, and particularly if associated with reduced fecundity (such as early lethal disorders) then the cause may be hypothesized to lie within a *de novo* mutation (38), which can be identified by sequencing the exome of an affected individual and these parents and filtering for *de novo* variants (see Figure 1-5). This powerful strategy (known as a *de novo* strategy or a ‘trio’ or family-based approach) significantly limits the number of potentially pathogenic variants. Previous studies have identified an average of five (range two-seven) candidate *de novo* non-synonymous mutations per affected individual (100). Sequencing of trios has been successfully used to find *de novo* mutations responsible for rare Mendelian disorders (for example, Schinzel-Giedion syndrome) as well as for genetically heterogeneous disorders such as intellectual disabilities (96).

An alternative method is to use an “overlap” approach, which involves combining data from multiple phenotypically similar but unrelated patients (see Figure 1-5). This may be appropriate when parental samples are unavailable and can be used effectively to reduce the number of candidate variants generated by WES to a manageable number, but at the risk that patients with similar phenotypes have mutations in different genes (38). For relatively common disorders such as DCM, where the causes are genetically highly heterogeneous then the mutational target i.e. the amount of the genome which is occupied by genes which when mutated result in disease is too high to have a have a feasible chance of finding multiple patients with mutations in the same gene (38).

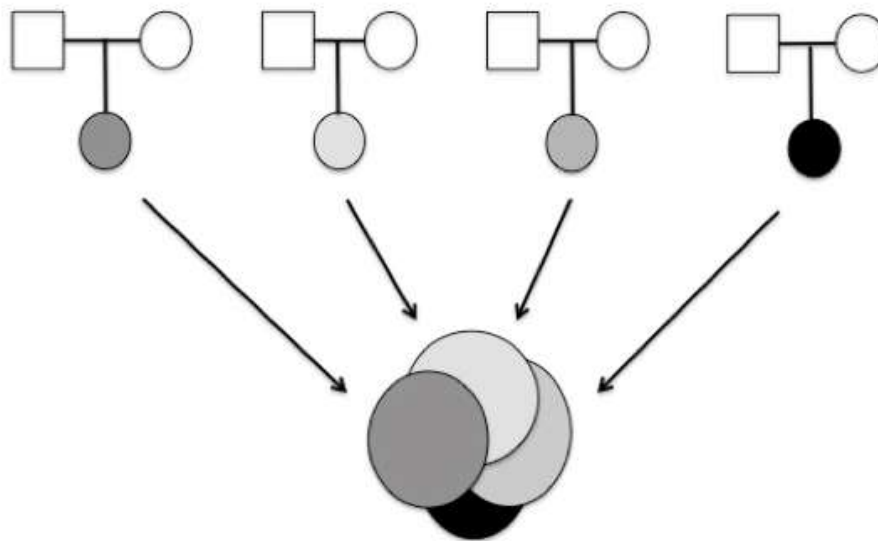
Figure 1-5 Whole Exome Sequencing approaches used to identify candidate genes

(A) Shows the 'trio' or family based approach of an affected child and both unaffected parents.
(B) Shows the 'overlap' approach of multiple unrelated affected individuals.

(A)



(B)



1.8 Key Cardiomyopathy genes

1.8.1 Titin (*TTN*)

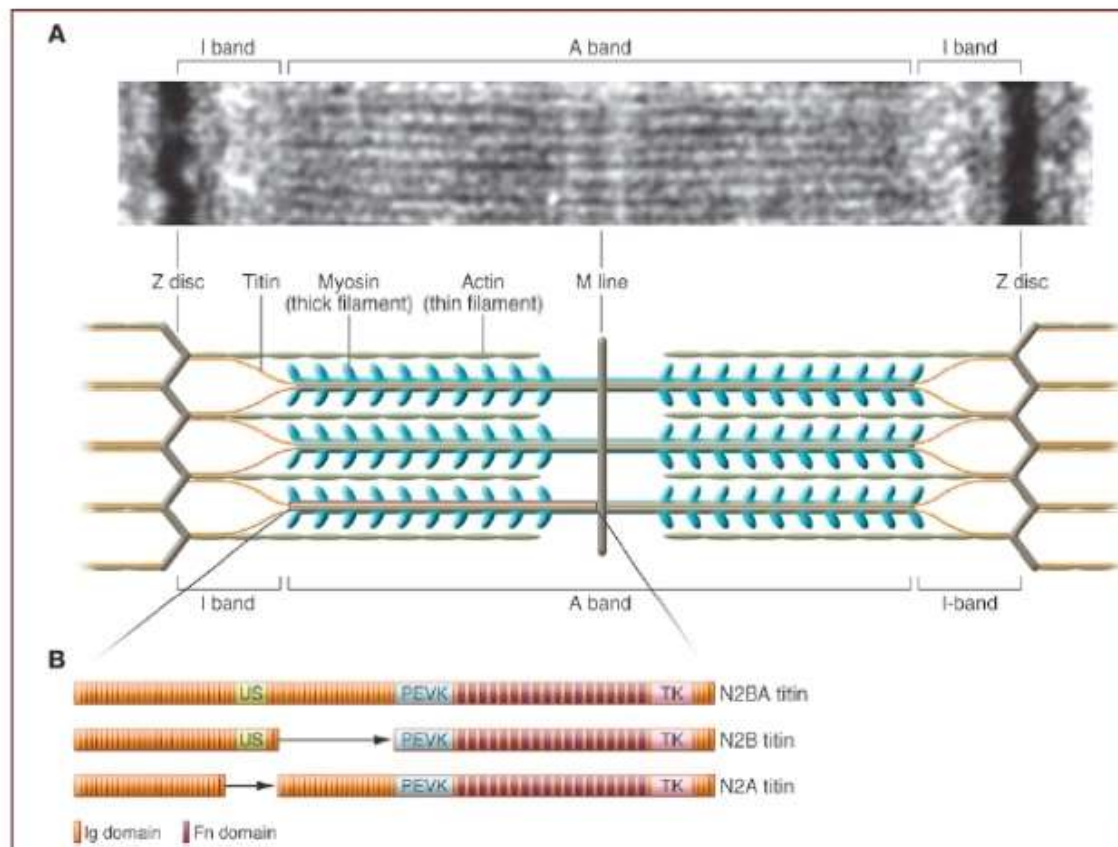
1.8.1.1 Titin Truncating Variants (*TTNtrvs*)

Although mutations in the gene *TTN*, which encodes the sarcomere protein titin have long been associated with dilated cardiomyopathy (104-106), the enormous size of the gene meant that until recently it was insufficiently analysed (107). An international collaborative study using high throughput sequencing involving The Royal Brompton & Harefield NHS Foundation Trust Cardiovascular Biomedical Research Unit (BRU) identified the role of Titin truncating variants (*TTNtrvs*) as a common cause of dilated cardiomyopathy, occurring in approximately 25% of cases of familial DCM and in 18% of sporadic cases (107). In addition to DCM, *TTN* mutations have been implicated in several other disorders with cardiac and skeletal muscle phenotypes, including as a rare cause of HCM (107), in ARVC overlap syndromes (108) and congenital myopathies such as tibial muscular dystrophy (109). Until recently mutations in known genes accounted for only a small proportion of genetic causes (110) in familial DCM, leaving the genetic basis frequently unresolved. The elucidation of the role of *TTNtrvs* and the introduction of large gene panels for diagnostic testing, coupled with NGS, mean the overall detection rate for sarcomeric gene mutations in genetic DCM is expected to increase to 35%-40% (16). NGS has overcome many of the technical challenges presented by the large coding sequence of *TTN* of ~100kb (107). The increased sensitivity compared to Sanger sequencing could be because NGS repeatedly analyses a given sequence (111), or that the techniques used to align the resulting sequences are more suited to the repetitive nature of *TTN* (111).

Titin is the largest human protein (~35,000 amino acids (111)), and the third most common striated-muscle protein (107, 112). It encoded by a single gene, located on the long arm of chromosome 2, region 2q31 (113). Titin is found within the sarcomeres, the contractile

protein structures in heart cells (111), and is necessary for their assembly (107). Together two titin molecules span the sarcomere and are anchored at the Z-disc and M-line (107), see Figure 1-6.

Figure 1-6 The role of *TTN* in the sarcomere



A. The sarcomere from Z disc to Z disc. Electron micrograph of a sarcomere from a human heart. The *TTN* gene encoding the giant protein titin is mutated in DCM. Titin's amino terminus anchors in the Z band, and its carboxy terminus ends in the ends in the M band. The titin kinase (TK) domain is found at the carboxy terminus and, when mutated, results in impaired stretch sensing and signaling. Titin interacts with both the thin and thick filaments. The thick-filament proteins (grey) are encoded by *MYH7* and *MYBPC3*, two genes commonly linked to HCM.

B. Isoforms of titin. Alternative splicing in the region of titin that encodes the I band gives rise to isoforms with varying spring properties. The N2B isoform is found exclusively in cardiac muscle and the N2A isoform in skeletal muscle. The N2BA isoform is also found in cardiac muscle and contains features found in both N2B and N2A titin. N2BA titin has a longer extensible I band region than N2B titin, making it more compliant. US, unique sequence; fn, fibronectin domains; PEVK, repeating units of amino acids (proline, glutamic acid, valine, and lysine) (16). Image reproduced with permission of the rights holder, American Society for Clinical Investigation.

The role of titin is to monitor the length of the sarcomere during muscle contraction and relaxation and to transmit this information to the rest of the cell (111). Titin is therefore sometimes referred to as the 'Molecular Ruler' (111). Titin is required for the passive forces

which maintain integrity of the sarcomere, and play a key role in LV mechanics, particularly filling during diastole (2). The titin molecule consists largely of tandem domains of the immunoglobulin and fibronectin-III types, along with specialized binding regions and a putative elastic region, the PEVK domain (114). Containing nearly 350 exons, *TTN* involves many repeated sequence motifs (111). The gene structure is organized to accommodate extensive splicing events. Earlier work by our group involved the definition of a meta-transcript for *TTN*, which for the first time, allows variants to be catalogued in an unequivocal manner, essential for variant classification, particularly in view of the enormous size and complexity of splicing in *TTN* (115). Of the multiple different titin isoforms; those found in adult cardiac muscle are predominantly N2B or N2BA (116). In addition, *TTN* encodes a separate cardiac isoform, known as novex-3 titin, which is only 5600 amino acids in length, and lacks the A-band and M-band segments of titin and therefore does not fully span the sarcomere. This isoform is less abundant in cardiac tissue than full-length titin (107). The lack of evidence for pathogenicity of novex-3 truncations means that mutations unique to this isoform can be discounted as a cause of DCM (117).

Observed *TTN* truncating variants (*TTN*trvs) in DCM are not randomly distributed along the length of the gene, but are more frequent in the area coding for a region, which interacts with the protein myosin in the sarcomere. Many of these mutations would disrupt titin's kinase domain and therefore may impair the protein's ability to signal to the cell that the sarcomere has lengthened abnormally (111). Heart cells may then be susceptible to damage from muscle activity, unaware that the sarcomere has elongated too far; in addition, some mutations may also disrupt titin's ability to recoil during muscle contraction (111). It appears that men with *TTN* mutations may have worse DCM than women with *TTN* mutations; this may indicate that male hearts being more sensitive to *TTN* mutations as they are normally larger than female hearts and a larger heart faces greater pressure and strain (111). DCM patients with a

*TTN**tv* have been identified to have more severely impaired LV function, lower stroke volumes, and thinner LV than in *TTN**tv*-negative DCM patients (117).

Extensive work undertaken at The Royal Brompton & Harefield NHS Foundation Trust Cardiovascular Biomedical Research Unit (BRU) has assessed mutations in *TTN*, seeking to identify which cause DCM and which do not (117), increasing the clinical utility of these commonly encountered variants (115). Using RNA data from human DCM hearts and data derived from GTEx (<http://www.gtexportal.org/home/>), they calculated for each exon in *TTN*, the proportion of transcripts incorporating that exon (referred to as the ‘proportion spliced in’ or PSI). It was found that distal exons, including those encoding the A-band of *TTN*, are constitutively expressed, whereas many proximal exons, particularly I-band exons, are variably spliced in different isoforms (117).

Furthermore, *TTN**tv*-containing exons in controls were noted to have lower usage than the *TTN**tv*-containing exons in DCM patients ($P=2.5 \times 10^{-4}$). On the basis of this observation, it was suggested that many of the *TTN**tv* found in control individuals might be tolerated as they fall in exons that are spliced out of the majority of expressed transcripts (117, 118). A summary of these transcript annotations, including PSI values, is available at <http://cardiodb.org/titin>

The frequency of *TTN**tv*s observed in control individuals suggests that there may be additional genetic or environmental factors (“second hits”), which impact upon or ‘buffer’ the disease causing potential of such mutations. For example, asymptomatic individuals may carry other gene variants which suppress the effects of a *TTN* mutation, perhaps by driving increased degradation of mutated *TTN* messenger RNA, or of the truncated titin, both of which could reduce the amount of damaged protein in the heart (111). Alternatively, physiological perturbations e.g. pregnancy may reveal otherwise latent *TTN* mutation effects (119). In addition, it appears that variation within *TTN* may be important, not only as a cause of disease on its own, but it may also modify the phenotype of mutations in other genes (113,

120), including a potential role as a common genetic modifier in DCM (121). Watkins hypothesizes that each A-band truncation has pathogenic potential but consistently low penetrance and comments that this would fit the late-onset DCM phenotype (mean age at diagnosis was ~50) and the notable paucity of AD DCM pedigrees (115).

Functional recovery from end stage heart failure through optimal medical therapy or as a result of mechanical unloading of the heart using a Left Ventricular Assist Device (LVAD) is possible, with earlier suggestions that this was less likely in genetic DCM (which has an underlying irreversible cause), than in non genetic DCM (with an underlying cause which is potentiality reversible) (122). More recently other work within our group showed both that recovery was possible in DCM secondary to a *TTN*tv and that recovery was as likely with or without a *TTN*tv(123).

1.8.1.2 Titin Non Truncating Variants (*TTN*nonTVs)

In the first report of *TTN* as a causative gene in familial DCM (106) two different *TTN* variants (one truncating variant and one missense variant) were described, segregating with disease within separate families. While there is robust evidence for *TTN* missense variants as the pathogenic variants underlying recessive skeletal muscle myopathies (124) further reports of *TTN* missense variants causing DCM are absent. It is conceivable that the *TTN* missense variant segregating with disease, in the original report by Gerull *et al* (106) was co inherited with a further (pathogenic) variant. I am unaware of any follow-up study on the original kindred to identify if variants in genes more recently linked to DCM are present. More recently Hinson *et al* (125) modeled certain *TTN* missense variants in induced Pluripotent Stem Cells and found that, like *TTN*tv's there was diminished contractile performance suggesting pathogenicity.

However, *TTN* variants (mostly non-truncating) have been identified in virtually every individual from control populations (126) a frequency, which far exceeds that expected for

associated diseases. For example, MacArthur *et al* noted four recent exome sequencing studies, involving a total of 945 families with a child affected by autism, which together observed four independent *de novo* missense mutations in *TTN*. Applying a statistical model that accounts for gene size, mutation rate, number of trios and distribution of exome coverage, the authors comment that 1.96 *de novo* *TTN* missense or loss-of function mutations are predicted by chance, which is not significantly different ($P=0.14$) from the four observed in the independent studies suggesting no casual role for *TTN* in autism (44).

The interpretation of *TTN* non truncating variants currently remains challenging and evidence does not support a clear disease causing role at this stage (127), although they may play a disease modifying role (121).

1.8.2 Additional Key DCM Genes

More than 40 genes have been associated with DCM, mostly encoding cytoskeletal, sacromeric, or Z disc proteins (128). Some of these genes may have been linked to DCM without robust data, before the rate of background variation ('noise') in cardiomyopathy genes was appreciated (27). Other work within our group has evaluated these genes in light of recent population data and identified seven key genes (other than *TTN*), see Table 1-3 with a significant enrichment in DCM (129)

Table 1-3 Additional Key DCM genes

Gene	Gene name	Percentage of Familial DCM traditionally attributed to mutations in this gene*
<i>LMNA</i>	Lamin A/C	6 %
<i>MYH7</i>	β -Myosin heavy chain 7	4.2%
<i>TNNT2</i>	Cardiac Troponin T, type 2	2.9%
<i>DSP</i>	Desmoplakin	-
<i>TCAP</i>	Titin-cap (Telethonin)	1%
<i>TPM1</i>	α- Tropomyosin 1	1%-1.9%
<i>VCL</i>	Vinculin	1%

Key genes, which play a role in DCM. The frequency of rare variants (defined as a mean allelic frequency < 0.0001 in ExAC) was compared between ExAC reference population samples and DCM cases referred for clinical genetic screening at the Oxford Molecular Genetics Laboratory (OMGL), UK and the Laboratory for Molecular Medicine, Partners Healthcare (LMM), USA. Forty-eight genes were assessed with the number of DCM cases sequenced ranging from 121 to 1315. Truncating (nonsense, frameshift, essential splice site) and non-truncating (missense, inframe indels) were analysed separately. A Fisher's exact test was performed to test for significant enrichment of rare variants in cases with Bonferroni multiple testing correction applied ($p < 0.001$) (ref personal communication with R Walsh; <http://cardiodb.org/ACGV/>). *The proportion of familial DCM traditionally attributed to mutations in these genes is also included (130).

1.9 Myocardial Infarction cohort

1.9.1 Cardiovascular Magnetic Resonance Imaging

In recent years, Cardiovascular Magnetic Resonance Imaging (MRI or CMR) has emerged as a key imaging technique for clinical and research assessment of the heart and large vessels.

By combining high resolution cine imaging with tissue characterization, it provides information on global LV function, regional wall motion, and thickening (6). It has played a key role in both understanding the natural history of the LV remodeling process and the role of pharmacological and non-pharmacological interventions, which alter this (131).

Incorporating stress testing and delayed enhancement techniques means that CMR is a unique method for superior non-invasive comprehensive assessment of myocardial viability and ischaemia (131). The use of gadolinium-chelated contrast agents (which have paramagnetic properties that effect magnetic properties of tissues) (6), permits the detection of perfusion defects, micro-vascular obstruction and areas of scarred tissue or fibrosis (6). Such agents do not penetrate myocytes with intact membranes, but will diffuse and accumulate where there is rupture of the cell membrane, such as in acute myocardial infarction. Similarly they will accumulate in the extracellular space where there is an increased volume of distribution such

as where there is replacement fibrosis (6). Where MI has occurred, then the diagnostic accuracy of CMR is high for the delineation of extent (partial thickness or transmural) of myocardial scar tissue (132). The sensitivity of detection of MI was found to increase with increasing doses of gadolinium contrast in a multi-centre, double blind, randomized control trial evaluating the efficacy of infarct detection by late gadolinium enhancement (LGE) in both acute and chronic MI (133). Detection of MI and measurement of the extent of injury as well as the distinction between viable and non-viable myocardium, has important implications for management decisions (134). In patients where there is a substantial amount of dysfunction but viable myocardium, then they are likely to benefit from coronary revascularization, with improvements in regional and global contractile function, symptoms, exercise capacity, and long-term prognosis (132). The question of how much viable myocardium is required to achieve a significant improvement in LV function following revascularization is a current challenge (6), but the general consensus is that the changes in LVEF after revascularization are linearly correlated with the number of viable segments (6). Additional advantages of CMR over alternative imaging techniques include a greater sensitivity and comparable high specificity for detection of LV thrombi in the MI population (135, 136). With growing prognostic implications, increased availability and continual technical improvement, then CMR is assuming an increasing role in patient stratification and management (131).

Limitations of CMR include many types of pacemakers and implantable cardiac defibrillators (ICD's) being contra-indicated. Patients with claustrophobia are unlikely to tolerate the enclosed space and associated noise. In addition, the ECG gating which is used with the rapid imaging techniques and sequences mean that the quality of images is significantly degraded where there are uncontrolled rhythm disturbances such as fast atrial fibrillation (AF). In those with significantly impaired renal function, gadolinium contrast agents have

been associated with nephrogenic systemic fibrosis, an untreatable, progressive debilitating disorder (137).

1.9.2 Myocardial Infarction

The 2010 Heart Disease and Stroke Statistics update from the AHA reported 8.5 million persons in the United States with MI, a prevalence that increases with age for both sexes. It is defined as an event caused by myocardial ischemia (as opposed to other pathology such as myocarditis or trauma)(138) resulting in myocardial injury or necrosis (139). While the process of MI is characterized pathologically by coagulative necrosis of the myocardium (140), the clinical diagnosis is reliant upon typical symptoms, and supportive evidence such as characteristic electrocardiographic (ECG) findings and biochemical markers of myocyte injury (such as troponin, serum creatine kinase and creatine kinase –MB) (140).

In the majority of cases, MI will correspond with epicardial CAD, detectable at coronary angiography. There are multiple alternative coronary artery causes of MI and these include most commonly coronary artery occlusion due to plaque rupture but also coronary artery dissection, plaque rupture with subsequent downstream embolism leaving no residual narrowing, coronary artery embolism, vasospasm which may or may not be pharmacologically induced e.g. with cocaine use and flush occlusion of coronary branches which may be iatrogenic during invasive coronary angiography. The final end point, whatever the aetiology of MI, is tissue necrosis.

Criteria for diagnosis of acute MI have evolved in recent years, reflecting the role of newer, more sensitive cardiac biomarkers (troponin), the role of cardiac imaging and percutaneous coronary angiography and intervention. In addition to the release of a third universal definition of MI in 2012 (see appendix Figure 4-1); a clinical classification based on the suspected aetiology of the ischaemia (141) has been developed recognizing the diverse range of causes of MI. See Table 1-4.

Table 1-4 Universal Classification of MI

Type 1	Spontaneous MI	Secondary to pathology in the coronary artery wall e.g. dissection, plaque erosion/rupture resulting in intraluminal thrombus.
Type 2	Secondary to an ischemic imbalance	Secondary to an increase in oxygen demand or decreased supply e.g. coronary artery spasm, anaemia, or hypotension.
Type 3	Resulting in death when biomarker values are unavailable	Sudden unexpected cardiac death before samples for biomarkers could be taken or too early for their detection.
Type 4a	MI related to PCI	
Type 4b	MI related to stent thrombosis	
Type 5	MI related to CABG	

The Joint Task Force of the European Society of Cardiology, American College of Cardiology Foundation, the American Heart Association, and the World Heart Federation (ESC/ACCF/AHA/WHF) endorsed by the European Society of Cardiology (ESC), developed a clinical classification according to the suspected aetiology of the ischaemia (141).
CABG, Coronary Artery Bypass Grafting; MI, Myocardial Infarction; PCI, Percutaneous Coronary Intervention.

Silent ischemia is recognized to account for significant proportion (75%) of all ischemic episodes (MI or angina) (142), with ischemia potentially being revealed by exercise treadmill testing or serial ECG monitoring. MI may also be confirmed or even identified in previously unsuspected cases when cardiac imaging is undertaken, revealing either new loss of viable myocardium or regional wall motion abnormality. It is recommend that the term “Silent MI’ be applied where asymptomatic patients develop new pathologic Q wave criteria for MI during routine ECG follow-up, or where cardiac imaging reveals reveal evidence of MI which cannot be directly attributed to a coronary revascularization procedure (141).

1.9.3 Differentiating Myocardial Fibrosis from Myocardial Infarction

Myocardial fibrosis is integral to a variety of cardiovascular pathologies, arising as both a cause and consequence of heart failure, it is characterized by an accumulation of extracellular matrix within the myocardium (143). The extracellular matrix plays several roles in normal myocardium including the provision of a structural basis for the compact organization of myocytes, transmission of contractile force and electrical signals, protection against myocardial rupture and the prevention of myocyte overstretching or muscle fibre slippage (143). The accumulation of fibrosis and increase in intracellular space may lead to abnormalities in cardiac muscle integrity and or the electrical conductive abnormalities (143).

The pathogenesis of myocardial fibrosis can be considered as a dichotomy of either presence at the macroscopic level (replacement or focal fibrosis) or the microscopic level (reactive or infiltrative interstitial fibrosis) (143). Extracellular remodeling is a fundamental biological process and differentiating effects from pathological processes versus ageing remains challenging (143). Myocyte death following an acute MI triggers activation of cardiac fibroblasts. The subsequent inflammatory response controls myocardial collagen turnover and subsequent extracellular fibre deposition, comparable mechanisms may be identified with other pathologies (143) (see Table 1-5). LGE identifying macroscopic fibrosis or scar has been noted in 9% to 13% of patients with unobstructed coronary arteries and cardiomyopathy possibly representing recanalization after MI (144, 145).

Biopsies of cardiac muscle may identify fibrosis, but this invasive technique has an associated risk of complications and samples may not be representative due to diversity in the range and anatomical location of fibrosis within the myocardium (143). CMR therefore offers an appealing non-invasive option both for identifying fibrosis and as a mechanism of permitting serial monitoring of fibrotic remodeling (143). LGE can be used for accurate quantification and visualization of focal fibrosis.

Table 1-5 Conditions associated with myocardial fibrosis

Condition	Characteristic features
Myocyte necrosis after Myocardial Infarction	Sub-endocardial and or transmural late gadolinium enhancement
Recurring toxic insults	Cellular apoptosis leading to inflammatory responses and localized formation of collagenous scars
Hypertrophic cardiomyopathy	Diffuse interstitial fibrosis in the regions of greatest hypertrophy
Infiltrative interstitial diseases e.g. Amyloidosis	Increasingly diffuse distributions, often sub-endocardial
Age related fibrotic remodeling	Disruption of this equilibrium may be seen with hypertension, obesity, diabetes mellitus, pressure and volume overload events.

1.9.4 LV remodeling in HF

Key terminology used in descriptions of HF is historical and based on the measurement of LV EF; the importance in HF is not only because of its prognostic importance (the lower the

EF the poorer the survival) but also as many clinical trials have traditionally selected patients based on their EF (usually based on radionuclide technique or echocardiography) (5).

EF is the stroke volume (which is the end diastolic volume (EDV) minus the end-systolic volume (ESV) divided by the EDV. In patients with reduced contraction and emptying of the LV (i.e. systolic dysfunction), stroke volume is maintained by an increase in EDV (because the LV dilates), i.e. the heart ejects a smaller fraction of a larger volume. The more severe the systolic dysfunction, the more the EF is reduced from normal and, generally, the greater the EDV and ESV (5). HF can also occur in the presence of a preserved EF (HFpEF) and is common in the elderly and reflects complex diastolic factors and sub-clinical systolic dysfunction but is not discussed further here.

LV remodeling can be considered as a maladaptive response to myocardial injury, which may affect LV systolic function and the patients' prognosis for survival. Following an ischemic insult, alterations to both ventricular geometry and function occur characterized by a complex series of histopathological changes in both injured and uninjured segments (131). At the cellular level it can largely be regarded as the response of cardiac myocytes to biochemical stress. Studies in animals have shown that following acute MI early structural changes occur, which include LV tissue loss with simultaneous LV cavity dilation; subsequently there is contraction of infarcted tissue with compensatory hypertrophy of non-infarcted tissue (131). In patients the maladaptive changes or pathological 'remodeling' which occurs in surviving myocytes and the extracellular matrix after MI leads to ventricular dilatation and impaired contractility, one measure of which is a reduced EF (5). Characteristically, untreated systolic dysfunction progressively worsens over time, with increasing enlargement of the LV and decline in EF, even though the patient may be symptomless initially. Two mechanisms are thought to account for this progression. The first is occurrence of further events leading to additional myocyte death (e.g. recurrent myocardial infarctions). The other is the systemic response induced by the decline in systolic function, particularly neurohumoral activation.

Two key neurohumoral systems activated in HF are the renin-angiotensin-aldosterone system and sympathetic nervous system. In addition to causing further myocardial injury, these systemic responses have detrimental effects on the blood vessels, kidneys, muscles, bone marrow, lungs, and liver, and create a pathophysiological 'viscous cycle', accounting for many of the clinical features of the HF syndrome, including myocardial electrical instability. Interruption of these two key processes is the basis of much of the effective treatment of HF (5). Highly effective heart failure therapies which have been shown to improve cardiac function in association with reverse molecular remodeling include neurohormonal blockade, mechanical unloading (LVAD) and cardiac resynchronization (146).

1.9.5 Management of HF and IHD

Therapeutic options include pharmacological approaches, coronary revascularization, and device therapy. Three classes of pharmacological agents (Angiotensin Converting Enzyme (ACE) Inhibitors/ Angiotensin II Receptor Blockers [ARBs], aldosterone antagonists, and Beta-adrenergic blockers) have substantially improved the treatment of those with heart failure and reduced ejection fraction in recent years (7).

Recent decades have seen two new causes of LV dysfunction recognized, namely, myocardial stunning and hibernation (6). Both imply the presence of viable myocardium, and evidence suggests they may contribute to LV dysfunction and heart failure in patients with CAD (6). Prior to the recognition of these concepts it was thought that impaired resting LV systolic function was an irreversible process (131). Myocardial stunning refers to the process whereby acute myocardial ischaemia rapidly impairs contractile function, but is not sufficient to produce myocardial necrosis. This dysfunction may persist for a prolonged period (hours to weeks), even after restoration of epicardial blood flow. The severity of changes and the duration of stunning relate to the degree of ischaemia and the underlying condition of the myocardium (131) but full functional recovery follows (6). Cumulative

stunning occurs in patients with CAD who have repeated episodes of demand ischaemia and could represent a substrate in the development of post ischaemic LV dysfunction (6).

Hibernating myocardium refers to resting LV dysfunction secondary to reduced coronary blood flow, which can be at least partially reversed either by myocardial revascularization and/ or by reducing myocardial oxygen demand (131). The term hibernation, often used synonymously for tissue viability, is a retrospective definition and is based on evidence of functional recovery following intervention such as revascularization (6). In contrast, the term viability, is a prospective definition, used to describe dysfunctional myocardium impaired by CAD with limited or absent scarring, which therefore has the potential for functional recovery but it does not imply evidence of functional recovery having occurred after interventions (6).

Various techniques to detect the presence of viable tissue have been developed to aid decisions on coronary revascularization (6) including LGE CMR in which animal studies have shown that independent of infarct age or wall motion, regions showing enhancement match regions of myocardial necrosis and irreversible myocardial injury whereas regions, which do not enhance post gadolinium based contrast injection, are viable (147, 148).

An assessment of viability is key in the prognosis of patients with CAD and LV dysfunction. Patients with viable myocardium who undergo revascularization have a better prognosis. While those with nonviable myocardium have worse outcome with revascularization, (both increased perioperative morbidity and mortality after revascularization) (131).

Further advanced heart failure therapies, include mechanical unloading using a Left Ventricular Assist Device (LVAD) and Cardiac Resynchronization Therapy (CRT) (146), which restores co-ordination between the left and right ventricles, improving mechanical efficiency and simultaneously improving contraction and relaxation and may be highly effective at improving cardiac performance (149). Each of these treatments has been shown

to improve cardiac function in association with reverse LV remodeling (6). Where such effects persist, reverse LV remodeling is associated with improved clinical outcomes (149). In recent years there has been general consensus between guidelines from ESC, European Heart Rhythm Association (EHRA), AHA/ACC/American College of Cardiology and Heart Rhythm for expanded indications for device therapy (150). ICDs are effective, relatively safe and superior to standard pharmacological therapies, but there are wide disparities in implantation rates both between countries and within countries due to cost implications (151). Recent studies based on longer term, follow up than the early trials suggest that the lifespan gain has been largely underestimated (152).

In conclusion, while a wide range of effective treatment options are now available for patients with HF, extending beyond pharmacological therapy, to include revascularization, CRT and ICD's, they have significant risks and costs associated and do not uniformly benefit patients. I therefore seek to establish if genetics can offer a novel method of risk stratification in heart failure.

2 Chapter 2: The role of Titin Truncating variants as a molecular mechanism underpinning left ventricular dysfunction in patients following myocardial infarction.

2.1 Introduction

MI is both an important cause of HF, and not-uncommon co-existing finding when patients are imaged using CMR. In the early period after MI (and other types of cardiac injury such as myocarditis, hypertension and valvular disease) ventricular remodeling occurs as a progressive process with resulting clinical effects, which may remain undetected for many years (153). Remodeling can be defined as

“molecular, cellular, and interstitial changes that are manifested clinically as changes in size, shape, and function of the heart after cardiac injury” (154)

Post MI remodeling which is disproportionate to the size of infarct is well recognised, but the reasons for such variable response to the infarct are largely unknown. One possibility is that in patients presenting with HF with small infarcts but substantial remodeling could be a dual pathological processes, perhaps due to an underlying genetic susceptibility to HF post-MI. In order to test this hypothesis we studied a cohort of patients with CMR evidence of MIs and examined the frequency of variation in genes more usually associated with DCM, a disorder that shares considerable phenotypic overlap.

Historically, the main terminology used to describe HF is based on measurement of LVEF. Major heart failure trials have largely enrolled patients with an EF $\leq 35\%$, often known as Reduced EF (HF-REF), or ‘systolic HF’, and only in these patients have effective therapies been demonstrated to date (5). In addition, LVEF is of prognostic importance in the post MI patient, as a risk factor for both sudden and non-sudden death (155). Furthermore LVEF is the cornerstone of decision making for prophylactic ICD implantation in this cohort of patients (155). The underlying basis of our study is to identify if genetic information can aid

the clinical management of patients with MI and LV dysfunction and therefore we have used LVEF as marker of disease severity.

To the best of our knowledge there have been no previous studies examining the role of Titin Truncating variants (*TTN*tv) or other key DCM genes in patients with MI and severe LV dysfunction. I focused on *TTN*tv as these variants represent the Mendelian genetic basis most closely associated with remodeling. The hypothesis being that in the cohort of patients with CMR evidence of a small MI and remodeling, the mutation burden of *TTN*tv may be higher in those with severely impaired LV function, than in those without LV dysfunction, or in the background population. It may be that the phenotype in these patients is not entirely attributable to ischaemia and that *TTN*tv should be specifically looked for in patients with disproportionate LV dysfunction for the size of infarction. The overall aim of this analysis is to establish if genetics can help us understand better LV remodeling process post-MI and to develop improved methods of risk stratification for patients and, potentially, family members.

2.1.1 Key questions

1. In subjects with CMR evidence of MI and severe left ventricular dysfunction (LVEF $\leq 30\%$) is there an increased mutation burden of *TTN* truncating variants (*TTN*tv)?
2. In subjects with CMR evidence of MI and severe left ventricular dysfunction (LVEF $\leq 30\%$) is there an increased mutation burden in key DCM genes other than *TTN*tv?
3. Does genetic information provide additive prognostic information in individuals with CMR evidence of MI?

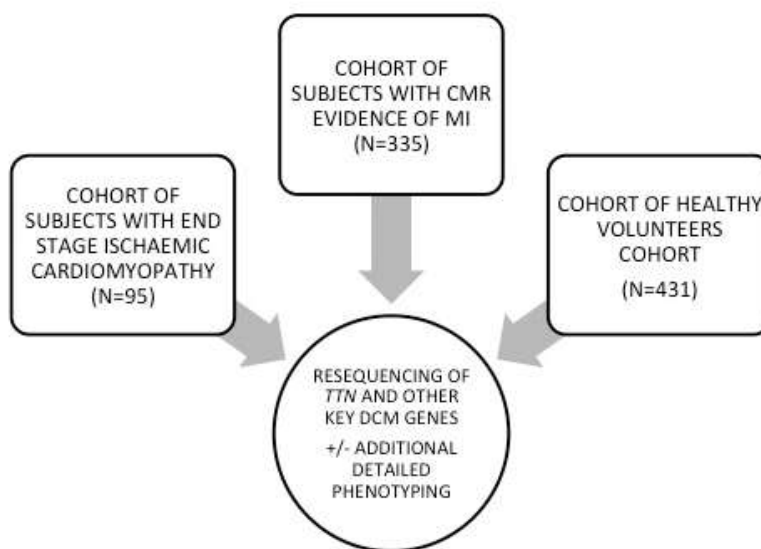
2.2 Methods

2.2.1 Study Setting and design

2.2.1.1 Study Cohorts

For the purposes of this project we consider three cohorts, see Figure 2-1. No subject within any cohort had a known familial relationship with any other subject. Studies were performed according to institutional guidelines and the approval of the local ethics committee.

Figure 2-1 Outline of study cohorts



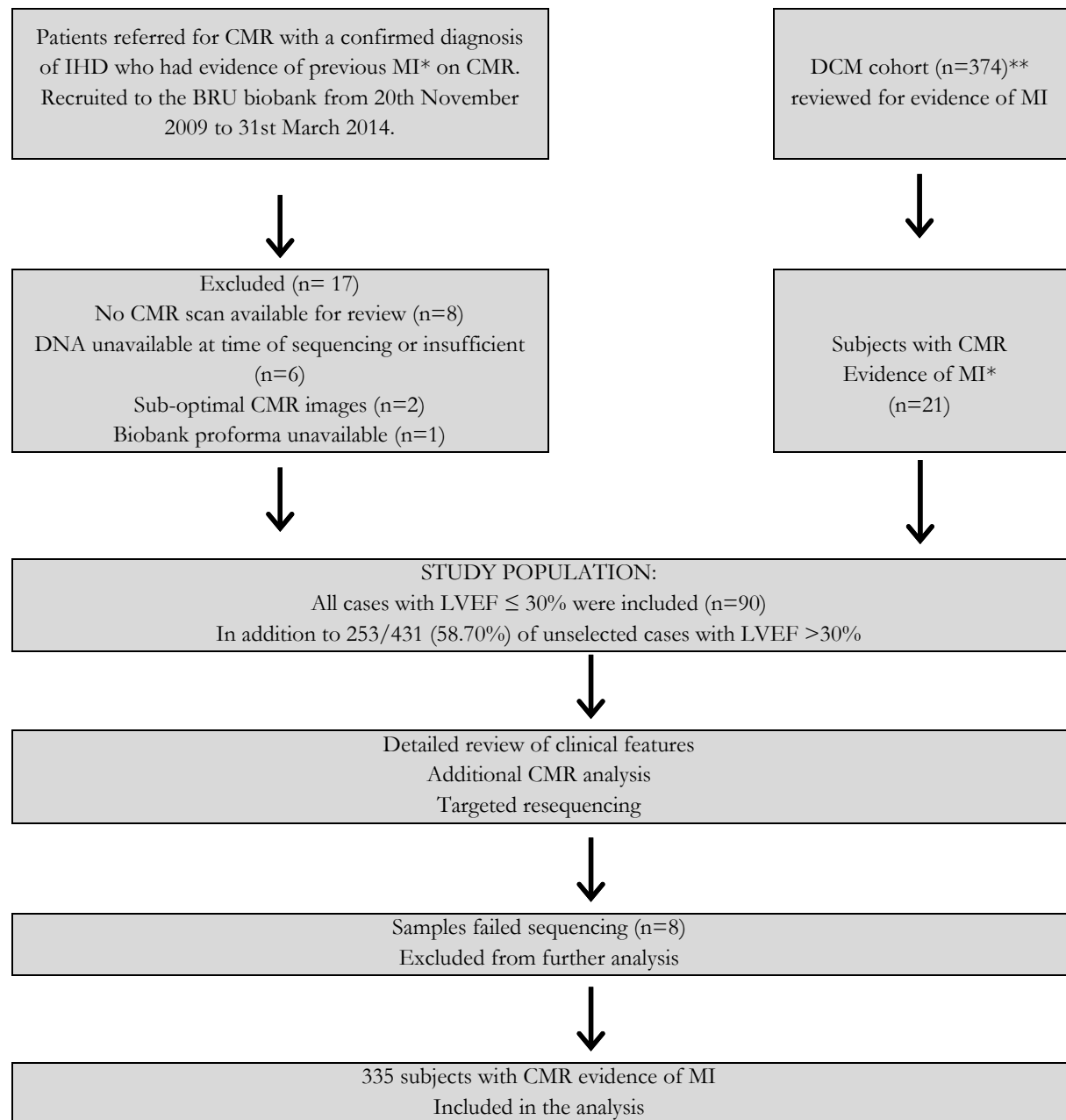
2.2.1.1 Cohort of subjects with CMR evidence of MI

The study population (see Figure 2-2) consists of three hundred and thirty-five patients with CMR evidence of MI, of predominantly European ancestry recruited through the National Institute of Health Research (NIHR)-funded Cardiovascular Biomedical Research Unit (CBRU) at the Royal Brompton and Harefield Hospitals NHS Foundation Trust (RBHT) and who underwent CMR from 24th April 2001 to 1st July 2013. All subjects were referred for CMR as part of their routine clinical management. Gadolinium-enhanced CMR characterized all patients in detail. All subjects were found to have CMR evidence of MI,

confirmed by two experts in CMR, blinded to genetic status. This cohort was subdivided into those with or without severe LV dysfunction.

- a. Subjects with CMR evidence of MI, with LVEF on CMR $\leq 30\%$.
- b. Subjects with CMR evidence of MI, with LVEF on CMR $>30\%$.

Figure 2-2 Study Flow showing inclusion criteria for cohort of subjects with CMR evidence of MI (n=335)



CMR, Cardiovascular Magnetic Resonance; LVEF, Left Ventricular Ejection fraction; MI, Myocardial Infarction. * MI is defined as partial or full thickness infarct affecting > 1 segment on CMR imaging, confirmed by two independent, experienced CMR readers, blinded to genetic status. ** Details of this DCM cohort have previously published(117).

2.2.1.3 Cohort of subjects with end-stage Ischaemic Cardiomyopathy

One hundred and one randomly selected patients, listed for cardiac transplantation and /or Left Ventricular Assist Device (LVAD) implantation between 1993 and 2011 at RBHT and prospectively enrolled in a tissue bank were studied. They all had a diagnostic label of ischaemic cardiomyopathy and represent cases of end-stage disease. Six cases outright failed DNA sequencing. Limited phenotype data is available on the remaining ninety-five cases included in this analysis. Eighty- seven (92%) of the cohort is male and seventy- six (80%) of the cohort self-reported Caucasian ethnicity.

2.2.1.3 Healthy volunteer cohort

Healthy adult volunteers (n=431) of predominantly European ancestry were prospectively recruited via advertisement at the MRC Clinical Sciences Centre, Imperial College London. Participants with previously documented cardiovascular disease, hypertension (HTN), diabetes, or hypercholesterolemia were excluded. The cohort of healthy volunteers were sequenced using the same methodology as cohorts with CMR evidence of MI and end-stage ischaemic cardiomyopathy, in order to provide comparable data on the background genetic variation in inherited cardiac disease genes.

2.2.2 Cardiac Magnetic Resonance (CMR) Imaging

Both MI subjects and healthy volunteers were characterized in detail by CMR. Healthy volunteers were excluded from further analysis if any cardiac abnormality was identified on CMR. Standard institutional MRI imaging protocols were followed and have been extensively detailed previously (117, 118).

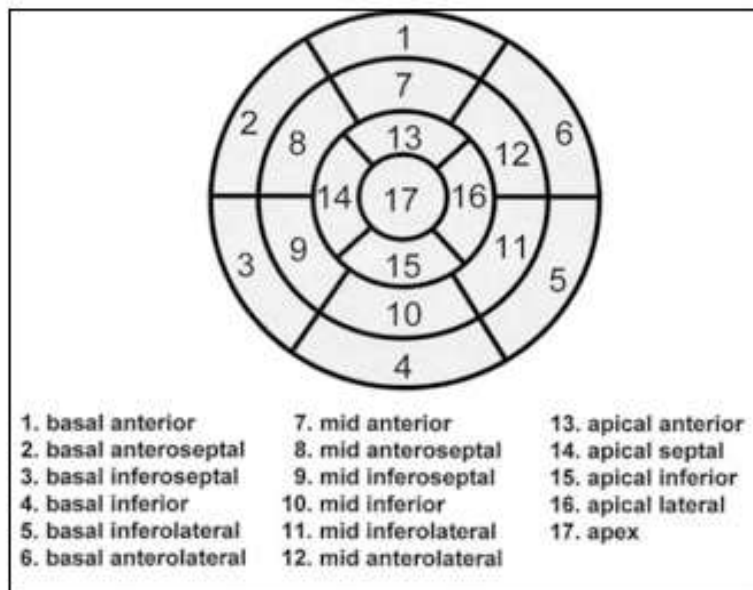
Briefly, in the MI Cohort, CMR was performed on Siemens Sonata 1.5-T or Avanto 1.5-T. Cine images were acquired using a steady-state free-precession sequence in standard 2-, 3- and 4-chamber long-axis views (TE [echo time]/TR [repetition time] 1.6/3.2 ms, flip angle 60°), with subsequent sequential 8-mm short-axis slices (2-mm gap) from the atrioventricular ring to the apex. Ventricular volumes and function were measured for both ventricles using standard techniques (156). 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 (CMR tools, Cardiovascular Imaging Solutions). MI subjects underwent gadolinium-enhanced CMR, unless contraindicated. Late gadolinium enhancement (LGE) images were acquired 10 min after intravenous gadolinium-DTPA (Schering AG; 0.1 mmol/kg) in identical short-axis planes using an inversion-recovery gradient echo sequence, as previously described (157). In the Healthy volunteer cohort CMR studies were performed on a 1.5T Philips Achieva system. 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. Volumes were measured and indexed as above for the MI cohort.

2.2.3 Additional CMR analysis

With MI subjects, a CMR-trained Clinical Fellow further analyzed the images, blinded to information on genetic status. The size of MI was quantified with segmental analysis, using the standard 17-segment cardiac model (see Figure 2-3). Both acute and chronic MI can be demonstrated on LGE CMR, with both the location and extent of infarction detectable. Infarct sizing has been shown to be highly reproducible, with minimal inter-observer and intra-observer variability (158). The infarcted myocardial segments were assigned to one the three major coronary arteries, as per standard methods (see

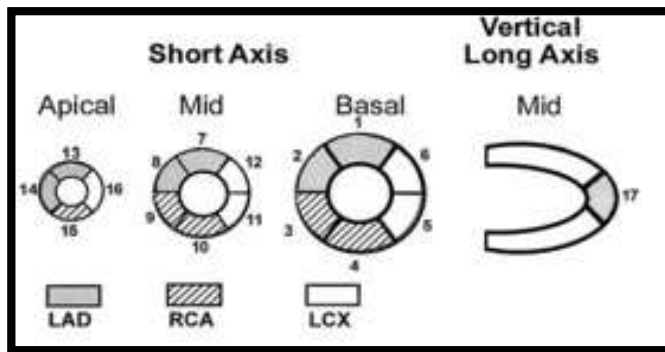
Figure 2-4). Ventricular wall motion abnormalities were defined as normal, hypokinetic, akinetic or dyskinetic. Additional CMR information recorded on each patient included infarct type (partial thickness or transmural infarct), infarct thickness (0-25%, 26-50%, 51-75%, 76-100%), wall thickness in most affected territory, degree of mitral regurgitation, evidence of infarcted infero-medial papillary muscle and perfusion abnormalities. Details of all CMR information recorded on each patient are shown in a copy of the specifically designed IHD pro-forma (see appendix Figure 4-2).

Figure 2-3 Standard 17 Segment Cardiac Model



Standard 17 segment cardiac model used to quantify the size of myocardial infarction and recommended nomenclature for tomographic imaging of the heart (159). Image reproduced with permission of the rights holder, Wolters Kluwer Health, Inc.

Figure 2-4 Coronary Artery Territories



Standard assignment of the 17 myocardial segments to the territories of the left anterior descending (LAD), right coronary artery (RCA), and the left circumflex coronary artery (LCX)(159), with the understanding that there will be anatomic variability. Image reproduced with permission of the rights holder, Wolters Kluwer Health, Inc.

2.2.3.1 Clinical Phenotype data on MI cohort

Detailed information on phenotype was collated and recorded alongside the CMR data on the electronic 'IHD proforma' (see appendix Figure 4-2). This included review of all data within the CBRU database, and where available, the electronic hospital records (Royal Brompton NHS Foundation Trust patients only). The majority of the data was gathered at the time of enrolment, but all information available at the time the proforma was completed was included. Data included reasons for referral for CMR, personal medical history, detailed family history with three generation pedigree, current medication, specific cardiac history, including prior revascularisation, the presence of significant coronary artery disease, history of arrhythmia (atrial fibrillation, ventricular tachycardia, non-sustained ventricular tachycardia and ventricular fibrillation) and implantable cardiac device therapy. Standardized clinical definitions were used wherever possible (see appendix Table 4-1).

2.2.4 Targeted Resequencing of *TTN*

The three cohorts (subjects with CMR evidence of MI, those with end stage ischaemic cardiomyopathy and healthy volunteers) were sequenced using the same methodology. Custom hybridization capture probes (SureSelect, Agilent) were designed to target genes implicated in a range of cardiovascular diseases including key DCM genes (see Table 1-3).

The evolution and optimization of this process have been extensively outlined previously (118). Briefly, the main target of interest was *TTN*, but in total ~200 genes were included (see Table 2-2 and appendix Table 4-6) DNA was extracted using standard automated approaches and quality and quantity was assessed by agarose gel electrophoresis and/or fluorometry (Qubit, Life Technologies). DNA libraries and target capture were performed according to the manufacturers protocols. 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. SOLiD 5500 paired-end sequencing (SOLiD 5500xl, Life Technologies) was used and the SOLiD Experiment Tracking System software utilized for run analysis reports.

2.2.5 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 v1.5-20 16 and Samtools v0.1.18. *TTN* variants called by any of these methods were taken forward for Sanger validation. Alignment metrics and coverage analysis were carried out using 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.

2.2.6 Variant annotation

Variants from NGS data were analysed using the established in-house customized pipeline as previously outlined (117). *TTN* variants were annotated relative to Locus Reference Genomic sequence (LRG) transcripts LRG_391_t1 and LRG_391_t2, in an inferred complete meta-transcript manually curated by the HAVANA group that incorporates all *TTN* exons other than a single alternative terminal exon unique to the shorter Novex-3 isoform, variants in which are reported relative to LRG_391_t2 and are excluded from this analysis. Variants were reported using standard Human Genome Variation Society nomenclature (160). The functional consequences of variants were predicted using the Ensembl Perl API (161) Variant Effect Predictor (47). Variants were classified as truncating if their consequence included one of the following sequence ontology terms: “transcript_ablation”, “splice_donor_variant”, “splice_acceptor_variants”, “stop_gained”, “stop_lost” or “frameshift_variant”. The in-house graphical user interface “Beehive” was utilized to identify the variation in *TTN* and additional key DCM genes in each cohort.

2.2.7 Sanger sequencing of *TTN* Truncating Variants

*TTN*trvs were the primary interest of this study and therefore variant confirmation of all *TTN*trvs called by NGS was undertaken using polymerase-chain-reaction amplification followed by dideoxy sequencing ‘Sanger Sequencing’. Primers were designed to amplify the region of *TTN* interest by PCR (see appendix Table 4-5). Amplification was performed using *Taq* DNA polymerase. PCR products were sequenced by capillary sequencing; using the BigDye Terminator Cycle Sequencing Kit and an ABI3500 Genetic Analyser (Applied Biosystems). Sequences were analysed using Sequencher (v5.3). Variants detected by NGS that did not validate by Sanger sequencing were discarded from further analysis (see Table 2-3).

2.2.8 Additional filters applied to other key DCM genes

In addition to *TTN*ts, the role of both *TTN* nonTVs and variants in other key DCM genes (*LMNA*, *MYH7*, *TNNT2*, *DSP*, *TCAP*, *TPM1* and *VCL*) were analysed. Sanger validation of these variants was not undertaken and instead more stringent filters (see Table 2-1) were applied to this data to reduce the number of variants detected, which were NGS sequencing errors. The sample filters chosen were based on earlier work within the group examining samples that had been analysed on the same NGS platform (SOLiD), which underwent Sanger sequencing, (personal communication R Walsh).

Table 2-1 Additional filters applied to variants other than *TTN*ts

Sample Filters	Variant filters
Coverage > 10x	• Global ExAC allele frequency <0.001
AB >0.2	• Allele frequency in individual ExAC population <0.001
AB>0.3 and coverage >30X	• Frequency across both our MI and HVOL cohort <3

AB, Allelic Balance; ExAC, Exome Aggregation Consortium; HVOL, Healthy Volunteer; MI, Myocardial Infarction.

2.2.9 Statistical Analysis

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. Odds Ratios are reported with 95% confidence intervals. All significance tests are two-tailed with a p value of less than 0.05 taken as significant. Multiple linear regression models were generated using known clinical covariates and optimized to minimize Akaike's Information Criterion (AIC). The relationships between clinical covariates, *TTN* genotype status and morphological parameters were evaluated by analysis of variance (ANOVA) between nested linear models.

2.3 Results

2.3.1 Quality of Sequencing

Callability refers to the percentage of a target sequence, sequenced with sufficient depth and quality that a heterozygous variant can be identified. It reflects the 'downstream usability' of

a data set and is used to confirm equivalent overall sequencing performance between our cohorts of subjects with CMR evidence of MI, healthy volunteers and those with end stage ischaemic cardiomyopathy, see Table 2-2, allowing unbiased comparison of mutation burdens (162). The main focus of the study was on *TTN* and Figure 2-5 shows the high quality of *TTN* sequencing across cohorts.

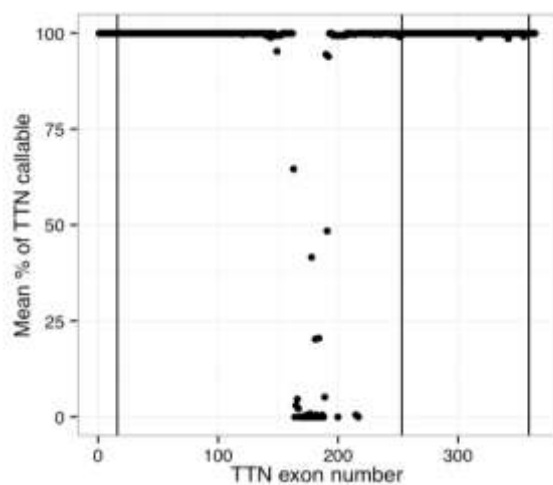
Table 2-2 Callability of key DCM genes across cohorts

	Cohort of subjects with CMR evidence of MI (n=335)				HVOL cohort (n=431)				End-stage ischaemic cardiomyopathy cohort (n=95)		
Gene	Max	Min	Mean		Max	Min	Mean		Max	Min	Mean
<i>DSP</i>	100.0	98.0	99.8		100	23.9	96.1		100.0	98.0	99.9
<i>LMNA</i>	100	50.9	96.8		100	29.2	96.2		100	85.4	97.9
<i>MYH7</i>	100	90.6	97.0		100	19.6	93.8		99.3	93.0	97.0
<i>TCAP</i>	100.0	21.7	98.6		100	28.4	96.9		100.0	89.7	99.0
<i>TNNT2</i>	100	96.0	100.0		100	23.4	95.0		100.0	99.7	100.0
<i>TTN</i>	99.5	97.9	99.1		100	27.5	96.9		99.4	92.4	99.0
<i>VCL</i>	100	95.1	99.7		100	23.7	95.3		100	96.1	99.6
<i>TPM1</i>	100	64.0	92.1		100	23.6	95.0		100.0	80.1	95.2

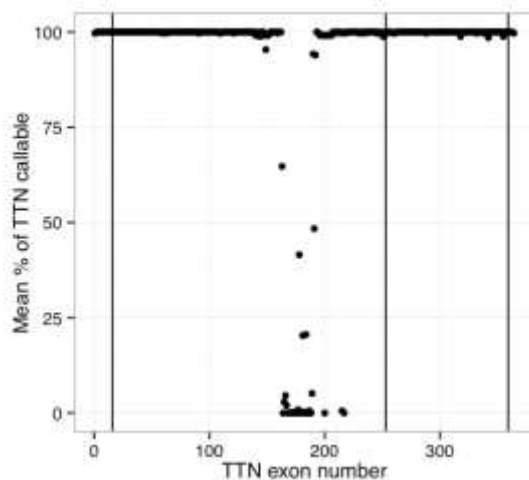
The maximum, minimum and mean percentage callability of key DCM genes is comparable across cohorts (MI subjects, HVOL cohort and severe end-stage ischaemic cardiomyopathy cases). Sequencing used SureSelect targeted capture and the SOLiD 5500 platform. DCM; Dilated Cardiomyopathy. HVOL; Healthy Volunteer. MI; Myocardial Infarction. Max; Maximum. Min; Minimum.

Figure 2-5 Quality of *TTN* sequencing across cohorts

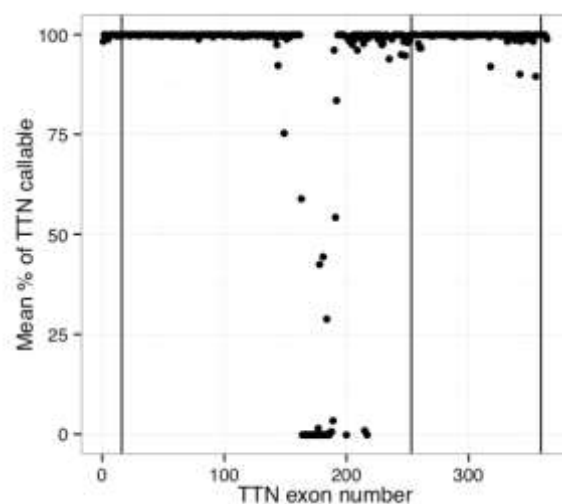
Cohort of subjects with CMR evidence of MI (n=335)



HVOL cohort (n=431)



End stage ischaemic cardiomyopathy cohort (n=95)



Scatterplots showing the mean percentage of each *TTN* exon callable across each cohort (Subjects with CMR evidence of MI, Healthy volunteers and end stage ischaemic cardiomyopathy cohort). Points represent the percentage of each exon sequenced with sufficient depth and quality to call variants as mean across samples. The vertical bars represent the domain boundaries between the boundaries Z disc, I Band, A band and M band. *TTN* exon numbering is according to LRG_391 and is not representative of exon size. Targeted resequencing using SureSelect targeted capture and SOLiD 5500 platform. MI; Myocardial Infarction. HVOL; Healthy Volunteer.

Table 2-3 *TTN*ts identified in each cohort that did not validate by Sanger sequencing

Cohort	BRU Code	Variant Type	Transcript variation	Protein variation	Chr. start	Coverage	Allelic Balance
Cohort with CMR evidence of MI	10CL03599	Frameshift	c.29006_29015delGAAGTGACAC	p.Gly9669ValfsTer6	179572279	NA	NA
	10DG00129	Nonsense	c.36646G>T	p.Glu12216Ter	179528047	18	0.22
	10DG00129	Nonsense	c.37231G>T	p.Glu12411Ter	179526540	12	0.17
	10DG00129	Nonsense	c.79645G>T	p.Glu26549Ter	179431214	41	0.21
	10PB02271	Frameshift	c.37770dupA	p. Pro12591ThrfsTer10	179522826	NA	NA
	10RB00013	Nonsense	c.105289G>T	p. Glu35097Ter	179396053	19	0.17
	10RP00728	Frameshift	c.37731delA	p.ALa12578ProfsTer369	179522865	NA	NA
	10TH03736*	Frameshift	c.37226dupT	p.Leu12410ProfsTer2	179526545	NA	NA
	10TM00255	Nonsense	c.85417G>T	p.Gly28473Ter	179425442	92	0.16
	12CH00267	Nonsense	c.1624G>T	p.Glu542Ter	179656837	17	0.17
	12GP00265	Nonsense	c.36964G>T	p.Glu12322Ter	179527140	4	0.50
Cohort with end-stage Ischaemic Cardiomyopathy	20LB01034	Frameshift	c.13606_13607insAACG	p.Ser4536LysfsTer5	179604354	NA	NA
	20RM01038	Frameshift	c.106542delTinsGT	p.Asp35514GlufsTer3	179393936	NA	NA
Cohort of Healthy Volunteers	14PM01944	Canonical splice variant	c.106532-1G>T		179393947	14	0.18
	14PS01275	Nonsense	c.6802G>T	p.Glu2268Ter	179639189	6	0.60
	14CS01449	Frameshift	c.63464delGinsAGCCC	p.Arg21155GlnfsTer13	179452670	NA	NA
	14CR01440	Frameshift	c.32588dupA	p.Val10865GlyfsTer7	179549443	NA	NA
	14AB01833	Canonical splice variant	c.37285+1G>C		179526485	5	0.50

Titin Truncating variants (*TTN*ts) identified in each of the cohorts (MI cohort, Severe, end-stage “Ischaemic Cardiomyopathy” cohort and healthy volunteer cohort), sequenced using the same methodology (SureSelect Targeted capture ICCv5 or ICCv6) and SOLiD, which did not validate by Sanger sequencing. Note that Frameshifts were called using a different programme to alternative variant categories, which did not provide comparable figures for coverage or reference read/variant read levels.

* The *TTN*tv c.37226dupT is located in exon 180. Primer design was repeatedly unsuccessful, with primers aligning to multiple (>1000) other areas of the genome. Coverage in this region was poor across all cohorts. This variant call was visually inspected using Integrative Genome Viewer and was consistent with a sequencing error.

2.3.2 Cohort with CMR evidence of MI: population and characteristics

A total of 335 patients with CMR evidence of MI are included in this analysis, and have undergone CMR and genomic sequencing as outlined previously.

A total of 9 *TTN*ts were identified by NGS and confirmed by Sanger sequencing. Therefore, the overall prevalence of *TTN*ts in this cohort is 2.7%. The characteristics of these *TTN*ts are outlined in Table 2-4 and CMR evidence of MI is shown in Figure 2-6.

Table 2-4 Characteristics of *TTN*ts (n=9) identified in MI cohort (n=335)

Subject ID	Variant Type	Transcript Variation	Protein Variation	Exon	Protein band	PSI* (%)
10DM03183	Nonsense	c. 7498C>T	p. Gln2500Ter	32	I band	100
10JC03016**	Canonical splice variant	c. 31426+1G>C	-	118	I band	74
10BM02061**	Canonical splice variant	c. 44816-1G>A	-	244	I band	100
10EF02449	Frameshift	c. 45344delT	p. Val15115AspfsTer62	246	I band	100
10BC00144**	Frameshift	c. 45756dupA	p. Tyr15253IlefsTer15	248	I band	100
10AF00538	Frameshift	c. 66010delG	p. Ala22004LeufsTer18	315	A band	100
10AO00220	Nonsense	c. 77421C>A	p. Tyr25807Ter	327	A band	100
10DM00077**	Frameshift	c. 81262_81269delCAGATGCT	p. Gln27088CysfsTer5	327	A band	100
10KW01906**	Canonical splice variant	c. 95415_95416+2delCAGT	-	344	A band	99

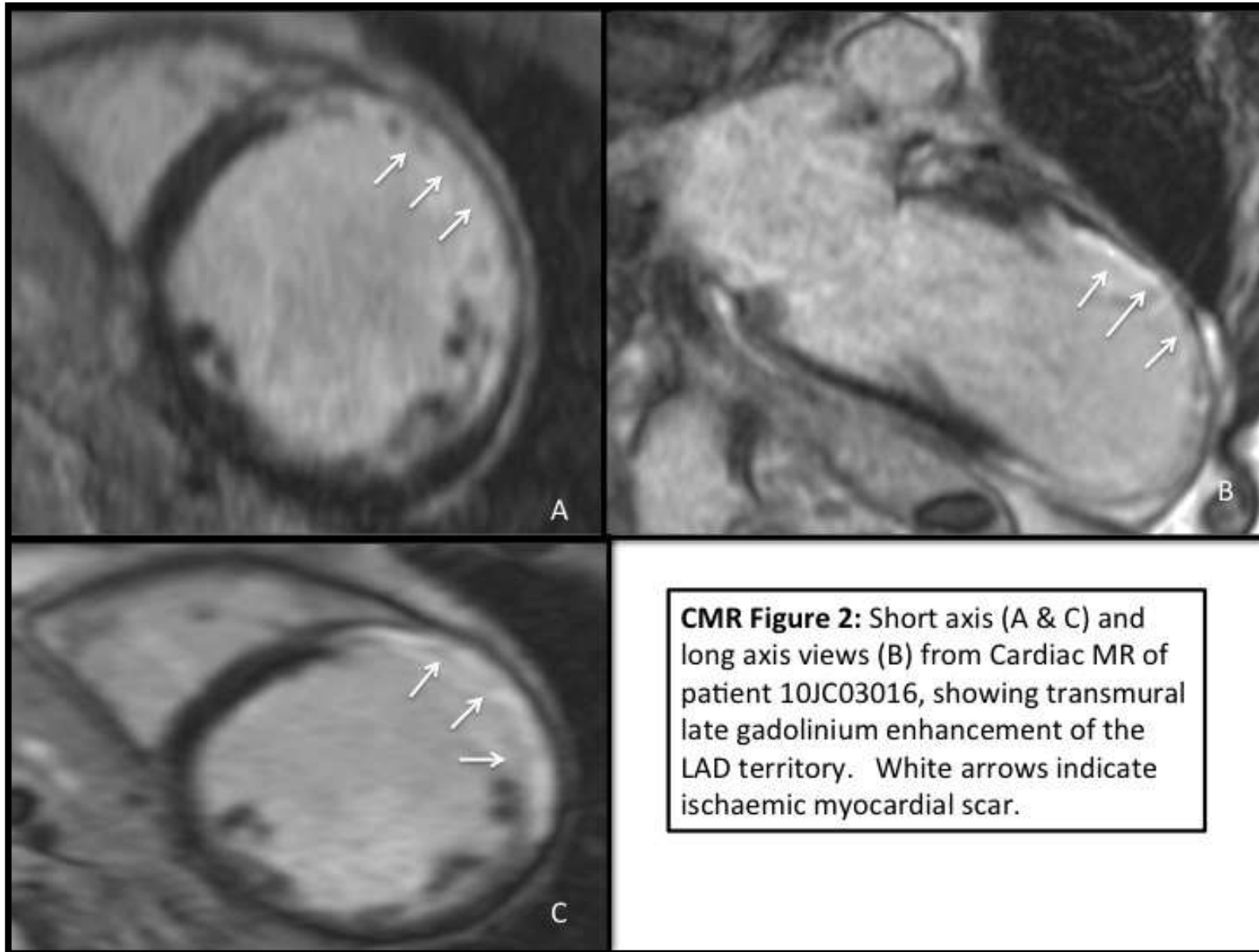
All subjects have Cardiac Magnetic Resonance evidence of Myocardial Infarction. *Represents the median Percentage Sliced In (PSI) value. For full explanation see methods.

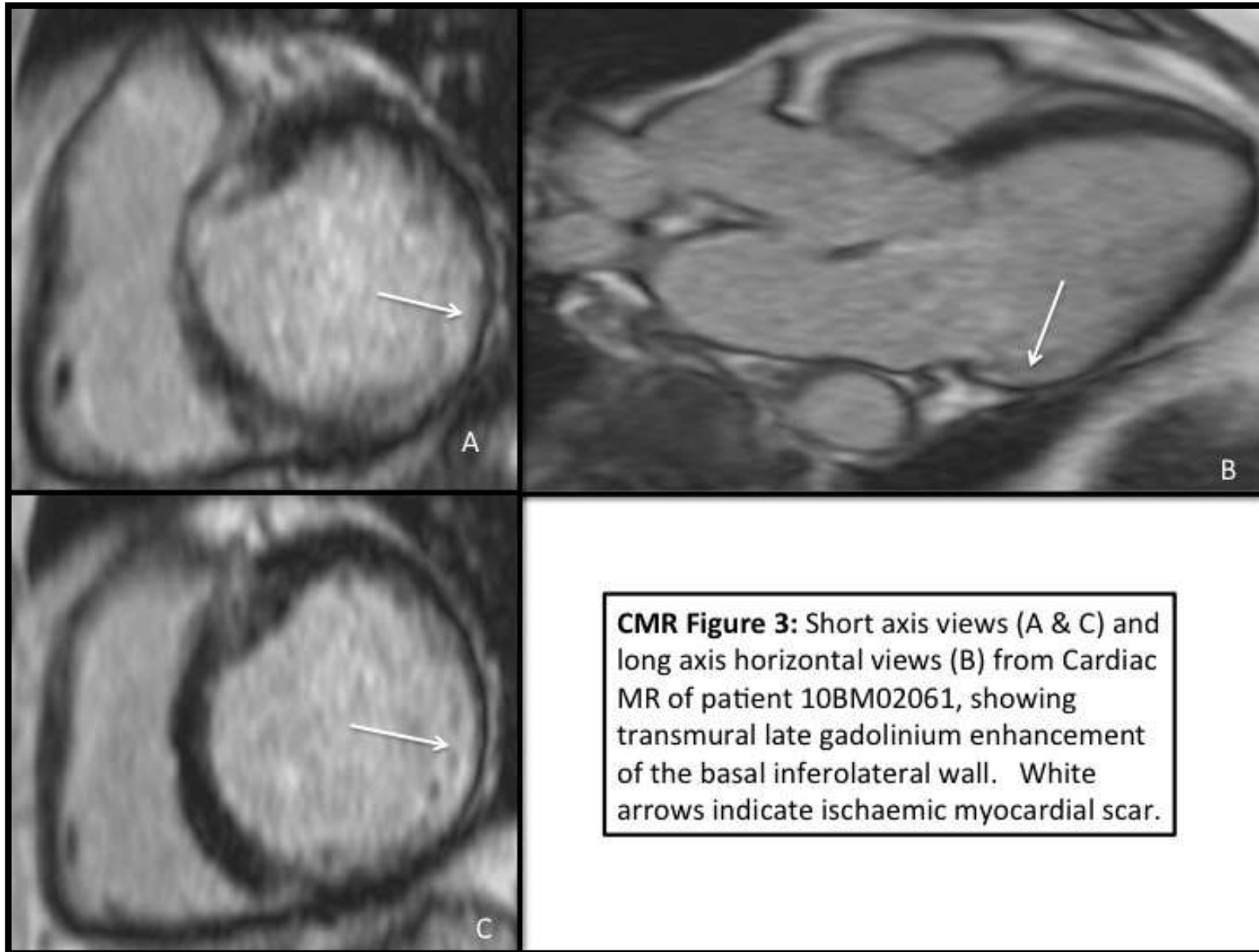
** Previously published(117). Transcript: ENST00000589042.1 Protein variation: ENSP00000467141.1

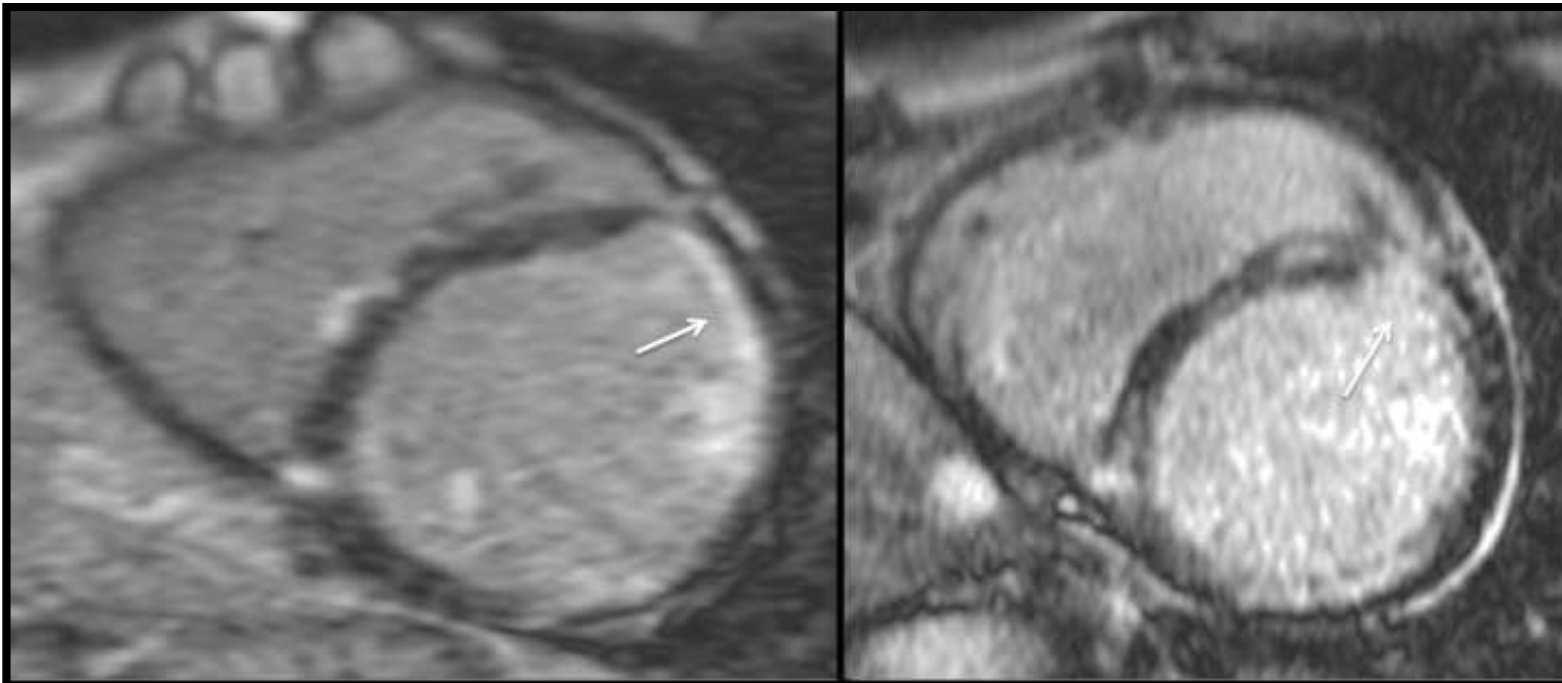
Figure 2-6 Cardiac MR images from cohort of subjects with CMR evidence of MI (n=335) with *TTN*vs (n=9)



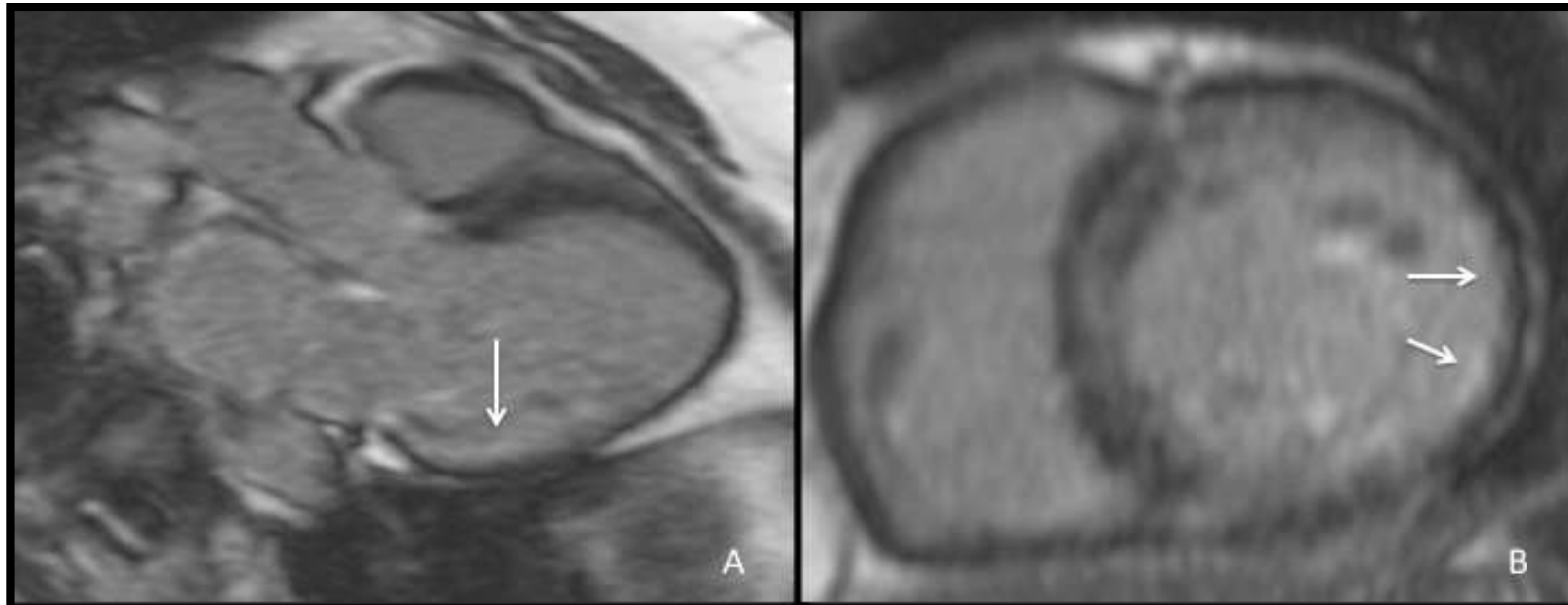
CMR Figure 1: Short axis view from Cardiac MR of patient 10DM03183, showing late gadolinium enhancement of the mid wall of the basal septum. White arrows indicate ischaemic myocardial scar.



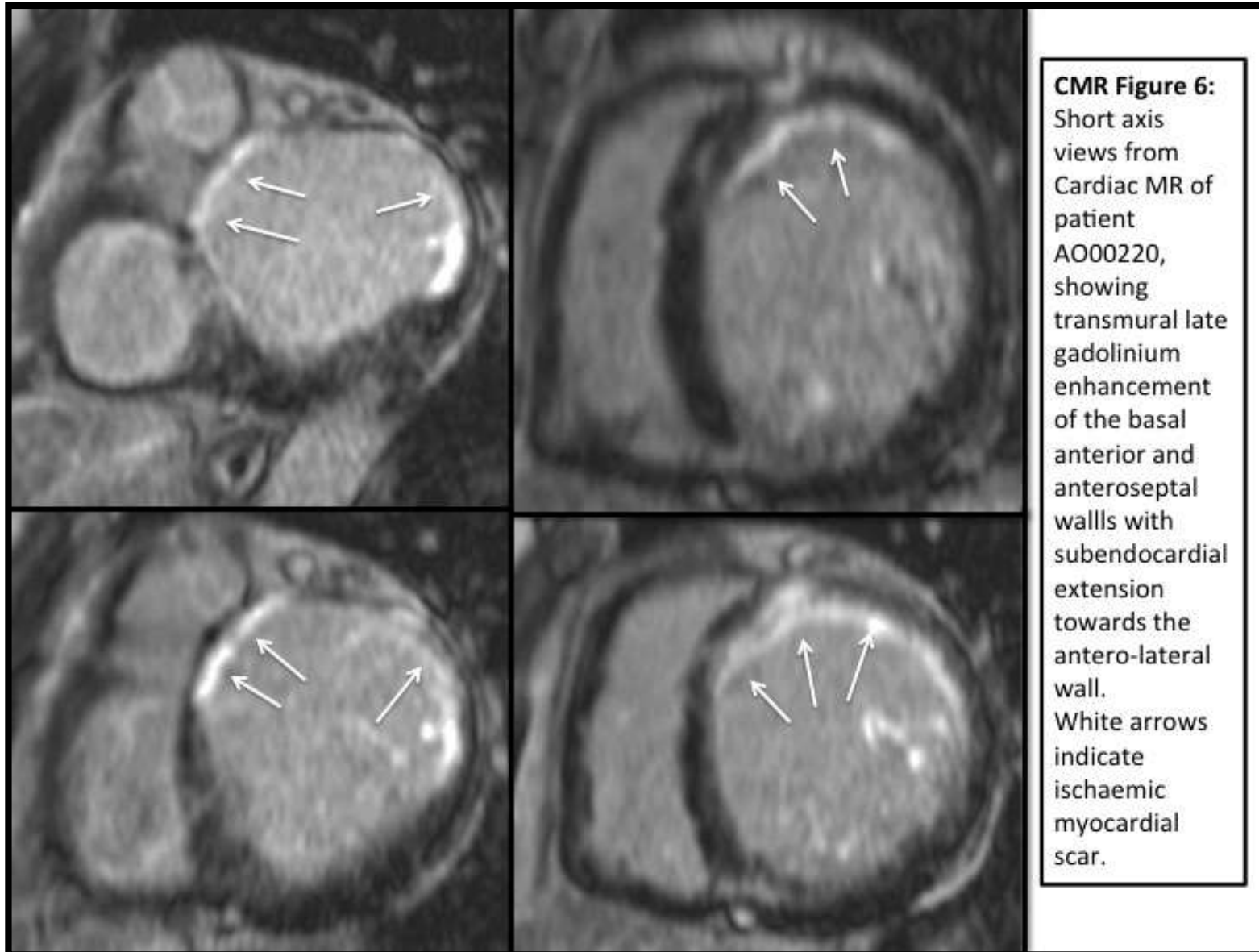


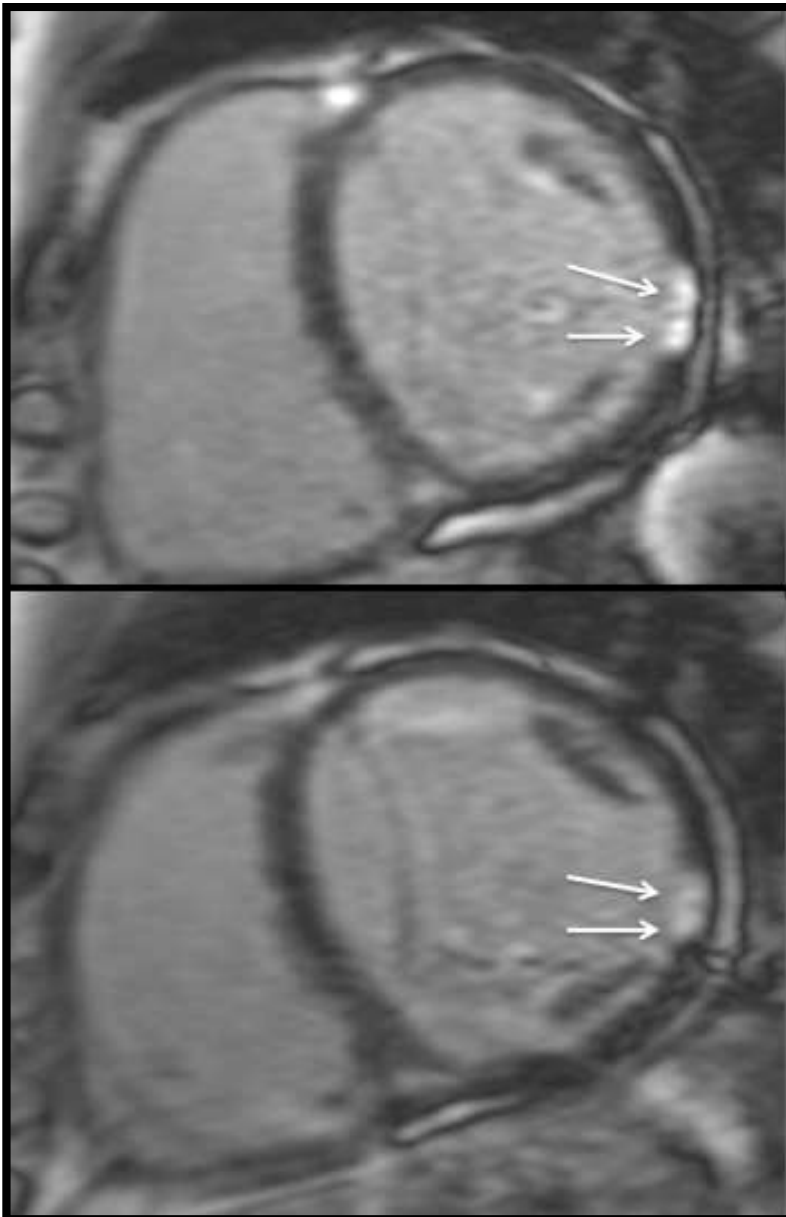


CMR Figure 4: Short axis views from Cardiac MR of patient 10EF02449, showing subendocardial late gadolinium enhancement of the anterior and anterolateral walls.. White arrows indicate ischaemic myocardial scar.

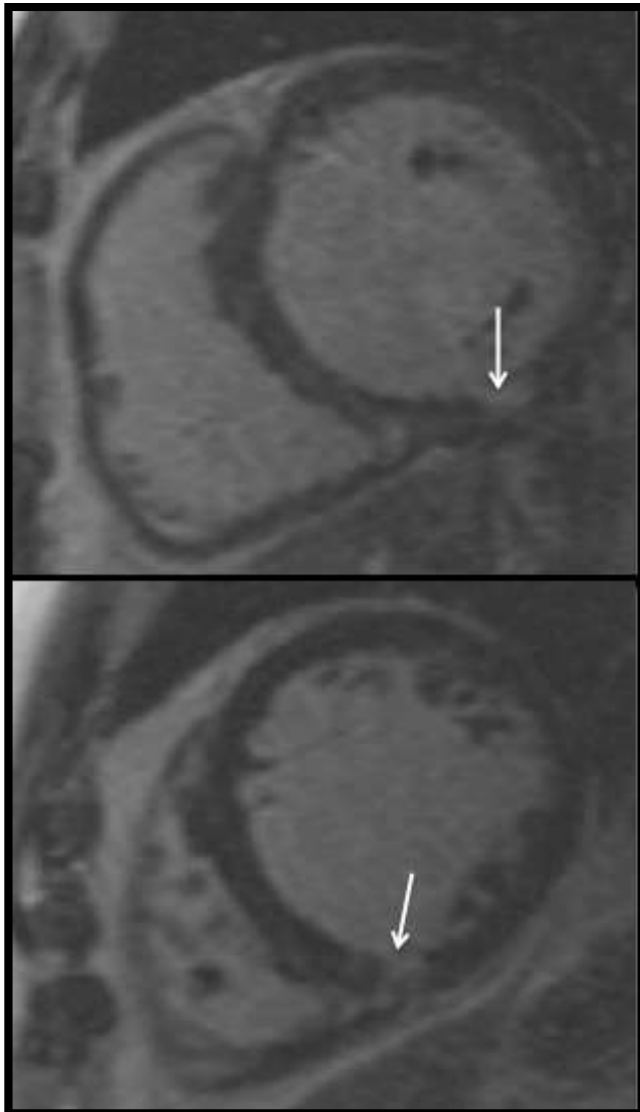


CMR Figure 5: Horizontal long axis (A) and short axis view (B) from Cardiac MR of patient 10AF00538, showing subendocardial late gadolinium enhancement of the inferolateral and lateral walls. White arrows indicate ischaemic myocardial scar.





CMR Figure 7: Short axis views from Cardiac MR of patient 10DM00077, showing transmurular late gadolinium enhancement of the inferolateral wall. White arrows indicate ischaemic myocardial scar.



CMR Figure 8: Short axis views from Cardiac MR of patient 10KW01906, showing transmurular late gadolinium enhancement of the basal inferior wall. White arrows indicate ischaemic myocardial scar.



CMR Figure 9: Short axis view from Cardiac MR of patient 10BC00144, showing partial thickness late gadolinium enhancement of the basal anteroseptal wall. White arrows indicate ischaemic myocardial scar.

Table 2-5 Baseline demographic and clinical characteristics of cohort of subjects with CMR evidence of MI (n=335), grouped by presence or absence of a *TTNtv*

		<i>TTNtv</i>- negative (n= 326)	<i>TTNtv</i>- positive (n= 9)	P value
Demographic data				
Male Sex		277/326 (85.0)	9/9 (100)	0.37
Age (at time of CMR)		67.87±10.48	62.89 ± 7.65	0.09
Ethnicity	Caucasian	268/326 (82.2)	7/9 (77.8)	0.66
	Other	58/ 326 (17.8)	2/9 (22.2)	
Hypertension		147/315(46.7)	2/9(22.2)	0.18
Diabetes		80/313(25.6)	1/8(12.5)	0.68
Smoking Status	Current Smoker	45/315 (14.9)	0/9	0.58
	Former smoker	167/315 (53.0)	6/9 (66.7)	
	Non Smoker	103/315 (32.7)	3/9 (33.3)	
Medication				
Beta blocker		237/316 (75.0)	8/9 (88.9)	0.46
ACE Inhibitor or ARB		262/318 (82.4)	8/9 (88.9)	1
Aldosterone antagonist		87/300 (29.0)	6/9 (66.7)	0.024
Antiplatelet		286/320 (89.4)	8/9(88.9)	1
Diuretic		171/310 (55.2)	6/9(66.7)	0.74
Statin		268/319 (84.0)	5/9 (55.6)	0.047
Coronary artery disease (CAD)				
Significant CAD or history of revascularization*		223/268 (83.2)	3/8 (37.5)	0.006
History of Arrhythmia				
Atrial Fibrillation		54/77 (70.1)	3 /4 (75.0)	1
Non Sustained VT		28/43 (65.1)	4/5 (80.0)	0.65
Sustained VT		1/19 (5.3)	0/1	1
Family history				
Family history of DCM		0/228	0/6	1
Family history of SCD		9/ 236 (3.8)	0/7	1
Family history of Premature CAD		52 /188 (27.7)	1/7 (14.3)	0.68

Baseline demographic and clinical characteristics of patients with Cardiac Magnetic Resonance (CMR) imaging evidence of Myocardial Infarction (MI) (n=335), grouped by presence or absence of a Titin Truncating Variant (*TTNtv*). Values are means +/- standard deviation or the number of patients and number of patients where data is available (with %) as appropriate. Groups were compared using Wilcoxon-Mann-Whitney (Wilcoxon Rank Sum) test for continuous variables, and Fisher's exact test for categorical variable. All tests are two-sided. Significant P Values (<0.05) are highlighted in bold. ACE I, Angiotensin Converting Enzyme Inhibitor; ARB, Angiotensin Receptor Blocker; DCM, Dilated Cardiomyopathy; SCD, Sudden Cardiac Death; VT, Ventricular Tachycardia. * For full definition see methods. Significant coronary artery disease is defined as >70% stenosis in any major coronary artery or >50% stenosis in the left main stem at the time of CMR. Revascularization refers to prior Coronary Artery Bypass Grafting (CAGB) or Percutaneous Coronary Intervention (PCI).

Table 2-6 CMR characteristics of cohort of subjects with CMR evidence of MI, grouped by the presence or absence of a *TTNtv*

CMR Parameters		<i>TTNtv</i> negative (n= 326)	<i>TTNtv</i> positive (n= 9)	P value
LV	EDVi	119.9± 44.6	129.9±30.6	0.27
	ESVi	75.3± 43.3	91.7± 34.6	0.13
	SVi	44.7±11.0	38.1±12.7	0.14
	EF	41.2 ± 14.7	31.2±13.4	0.05
RV	EDVi	76.7±22.9	87.6±33.8	0.26
	ESVi	33.2±18.9	52.4±32.6	0.09
	SVi	42.2±11.4	35.3±14.8	0.13
	EF	56.8±13.2	42.9±17.6	0.02
LVMi		86.4±25.9	93.0±18.1	0.23
Indication for CMR				
Assessment of ischemia, viability or fibrosis		181/311 (58.2)	3/7(42.9)	0.46
Diagnostic uncertainty		50/311 (16.1)	2 /7(28.6)	0.32
Prior to device insertion		21/311 (6.8)	1/7 (14.3)	0.40
Other		59/311 (19.0)	1/7 (14.3)	1
Infarct characteristics				
Number of infarcted segments		5.90±3.25	3.00±2.29	0.007
Infarct thickness				
0-25%		25 (7.7)	1 (14.3)	0.13
26-50%		40 (12.3)	1(14.3)	
51-75%		121(37.1)	6 (66.7)	
76-100%		140 (43.0)	1(14.3)	
Wall thickness in most affected territory		3.17±1.50	4.22 ± 1.86	0.04
Multi-territory infarct		102/326 (31.3)	1/9 (11.1)	0.29
Most affected territory				
LAD		139 (42.6)	4 (44.4)	0.36
LCx		54 (16.6)	3 (33.3)	
PDA		2 (0.6)	0	
RCA		113 (34.7)	2 (22.2)	

Continued on next page

Table 2-6 continued

	<i>TTN</i> tv-negative (N= 326)	<i>TTN</i> tv positive (N= 9)	P value
Wall motion of most affected territory			
Akinetic	221 (67.8)	3 (33.3)	0.028
Dyskinetic	26 (8.0)	0	
Hypokinetic	66 (20.3)	6 (66.7)	
Normal	13 (4.0)	0	
Mitral regurgitation			
None	160 (49.1)	5 (55.5)	0.821
Minimal	130 (39.9)	3 (33.3)	
Moderate	30 (9.2)	1 (11.1)	
Severe	6 (1.84)	0	
Infarcted infero-medial papillary muscles	80 (24.8)	3(33.3)	0.696
Inducible perfusion deficit	70/142 (49.3)	0/5	0.059
Main perfusion deficit territory			
LAD	40 (58.0)	0	1
LCx	7 (10. 2)	0	
PDA	1 (1.5)	0	
RCA	21 (30.4)	0	
Normal segments*	11.0 ±3.9	15.0±2.4	0.0019
Hibernating segments*	1.8 ±2.9	0±0	0.016

Cardiac Magnetic Resonance (CMR) characteristics of patients with Myocardial Infarction (MI), grouped by presence or absence of a Titin Truncating variant (*TTN*tv). Values are means +/- Standard Deviation or the number of patients and number of patients where data was available (with %) as appropriate. Groups were compared using Wilcoxon-Mann-Whitney (Wilcoxon Rank Sum) test for continuous variables, and Fisher's exact test for categorical variables. All tests are two-sided. Significant P Values (<0.05) are highlighted in bold. Measurements are indexed to body surface area where indicated (i).

LV, left ventricle; RV, right ventricle; EDVi/ESVi, indexed end-diastolic/systolic volume; EF, ejection fraction; LVMI, indexed LV mass; LAD, left anterior descending; LCx, left circumflex; PDA, posterior descending artery; RCA, right coronary artery. *The standard 17-segment model was used for analysis. The number of normal, hibernating and non-viable segments adds up to 17 in each case. Comments on hibernation were made only where there was sufficient information on contemporaneous coronary artery disease status). In all other cases the number of segments were divided in normal and non-viable.

Table 2-7 Characteristics of clinical presentation in patients with a *TTNtv* (n=9) from the cohort of subjects with CMR evidence of MI

Subject ID	10DM03183	10JC03016*	10BM02061*	10EF02449	10BC00144*	10AF00538	10AO00220	10DM00077*	10KW01906*
Indication for CMR	? Ischaemia	? Ischaemia	? Ischaemia	? Ischaemia/ Viability	Episode of NSVT ?ARVC	CPOE, minimal Trop minor CAD on coronary angiography, global LV Impairment on Echo, CMR to assess	Trop + CP and VT	? Myocarditis vs Chronic fibrosis	SOB, lateral T wave inversion, normal coronaries- Investigate further
Risk factors for CAD	+	+	+	+	+	+	+	+	+
History of revascularization or presence of significant CAD*	+	+	NA	+	-	-	-	-	-
Chest pain	-	NA	NA	-	-	+	+	-	NA
Presenting complaint	SOBOE, orthopnoea	Late presentation MI (2 years pre-CMR)	NA	SOBOE	Palpitations	CPOE	CP & presyncope	SOBOE	NA
Troponin	-	NA	NA	-	-	+	+	+	NA
ECG	Lateral T wave inversion	Anterolateral T wave inversion	Anterior q waves	<i>Post procedure only</i> - Infero-lateral T wave inversion and partial LBBB	NA	Non-specific	Non-specific	Partial RBBB. Anterior q waves	Lateral T wave inversion
Clinical evidence of other pathology e.g. myocarditis	-	-	-	-	-	-	-	-	Alcohol
Additional Information	VT, ICD post CMR	-	AF	VT, ICD post CMR	NSVT, ICD post CMR	-	NSVT, ICD post CMR	-	-

AF, Atrial Fibrillation; ARVC, Arrhythmogenic Right Ventricular Cardiomyopathy; Coronary Artery Disease; CP, Chest Pain; CPOE, Chest Pain On Exertion; CMR, Cardiac Magnetic Resonance; DCM, Dilated Cardiomyopathy; ICD, Implantable Cardioverter Defibrillator; NA, Not Available; NSVT, Non Sustained Ventricular Tachycardia; RBBB, Right Bundle Branch Block; SOBOE, Shortness of Breath On Exertion; VT, Ventricular Tachycardia. Significant coronary artery disease is defined as >70% stenosis in any major epicardial coronary artery or >50% stenosis in the left main stem at the time of CMR. Revascularization refers to with Coronary Artery Bypass Grafting (CAGB) or Percutaneous Coronary Intervention (PCI). * Previously published (117)

Table 2-8 CMR Characteristics of patients with a *TTNiv* (n=9) in the cohort of subjects with CMR evidence of MI

Subject ID	10DM03183	10JC03016*	10BM02061*	10EF02449	10BC00144*	10AF00538	10AO0020	10DM00077*	10KW01906*
LVEF	25%	26%	43%	21%	50%	54%	25%	18%	19%
Territory most affected by MI	LAD	LAD	LCx	LCx	LAD	LCx	LAD	RCA	RCA
Wall thickness in most affected territory	5 mm	2 mm	4mm	8mm	5mm	4mm	2mm	5mm	3mm
Multi-territory infarct	+	-	-	-	-	-	-	-	-
Wall motion of most affected territory	Akinetic	Akinetic	Hypokinetic	Hypokinetic	Normal	Hypokinetic	Akinetic	Hypokinetic	Hypokinetic
Infarct thickness	51-75%	76-100%	51-75%	26-50%	26-50%	0-25%	51-75%	51-75%	51-75%
Number of infarcted segments	4	8	1	2	2	4	4	2	1

LAD; Left Anterior descending; LCx; Left Circumflex; LVEF; Left Ventricular Ejection Fraction; RCA; Right Coronary Artery. * Previously published (117).

Table 2-9 Prevalence of *TTN*ts in subjects with CMR evidence of MI, grouped by level of LVEF compared to background population

	Cohort with CMR evidence of MI (n=335)	As compared to <i>TTN</i> tv burden in our HVOL Cohort (2.6%) *			As compared to the <i>TTN</i> tv burden in ExAC (1.13%)**		
	<i>TTN</i> tv +	Odds Ratio	95% CI	P value	Odds Ratio	95% CI	P Value
MI and LVEF ≤30% (n=86)	6 (6.98%)	3.73	0.91, 14.02	0.034	6.56	2.33, 15.00	0.0005
MI and LVEF >30% (n=249)	3 (1.21%)	0.74	0.12, 3.27	0.75	1.07	0.22, 3.17	0.76

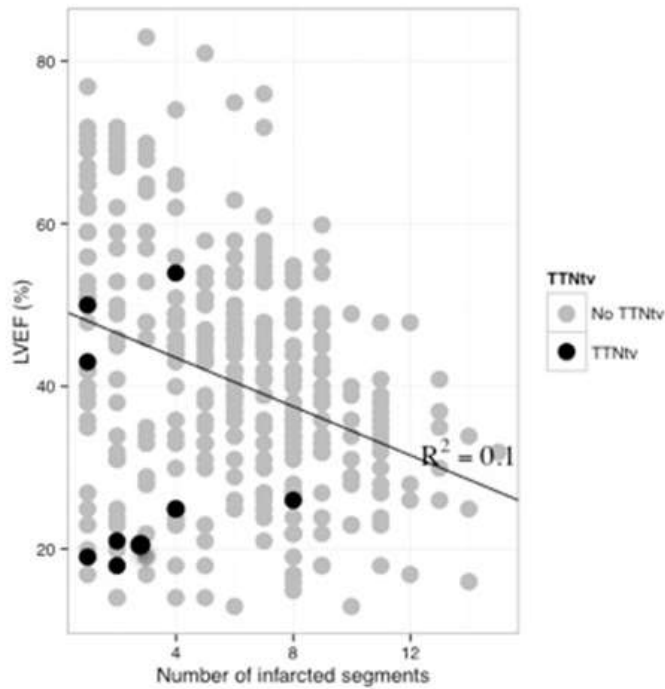
The prevalence of Titin Truncating Variants (*TTN*ts) in subjects with CMR evidence of MI, grouped by level of Left Ventricular Ejection Fraction (LVEF) compared to the background population rates.*Our Healthy Volunteers (HVOL) control cohort were simultaneously recruited and sequenced with the same methodology as MI subjects. 11 *TTN*ts were identified in 431 cases (2.6%).*The ExAC data represents all *TTN*tv (n=640) in the 363 meta-transcript exons of *TTN* with a PASS filter in ExAC, v0.3. For calculating the frequencies, the mean of the total allele count (n=113284) was used as the denominator rather than the total allele count in ExAC (n=121412). The mean of the total allele count varies throughout the gene/genome based on the coverage/quality of sequencing at each position (personal communication with R Walsh). The mean total allele count of 113284 is equivalent to 56642 controls.

Table 2-10 Mutation burden, variant type and distribution of *TTN*_{tv} in subjects with CMR evidence of Myocardial Infarct (n=335), grouped by level of Left Ventricular Ejection Fraction (LVEF)

	LVEF ≤30% (n=86)	LVEF >30% (n=249)	Odds Ratio	95% CI	P value
Total number of <i>TTN</i>_{tv}	6 (6.98%)	3 (1.21%)	6.11	1.27, 38.65	0.011
Variant Effect					
Nonsense	2 (2.33%)	0	*	*	0.065
Frameshift	2 (2.33%)	2 (0.80%)	2.93	0.21, 40.95	0.273
Canonical Splice variant	2 (2.33%)	1 (0.40%)	5.87	0.30, 348.92	0.163
Sarcomere Domain					
I-band	3 (3.49%)	2 (0.80%)	4.44	0.50, 53.97	0.109
A-band	3 (3.49%)	1 (0.40%)	8.89	0.70, 471.08	0.054
Exon Expression Level (PSI)					
High (≥90%)	5 (5.81%)	3 (1.20%)	5.03	0.96, 33.13	0.029

LVEF, Left Ventricular Ejection Fraction; PSI, Percentage Spliced In. Groups were compared using Fisher's exact test. All tests were two-sided. P values not corrected for multiple testing, as variables were not independent. Significant P Values (<0.05) are highlighted in bold. * Values not available as no subjects with LVEF>30% with a nonsense variant.

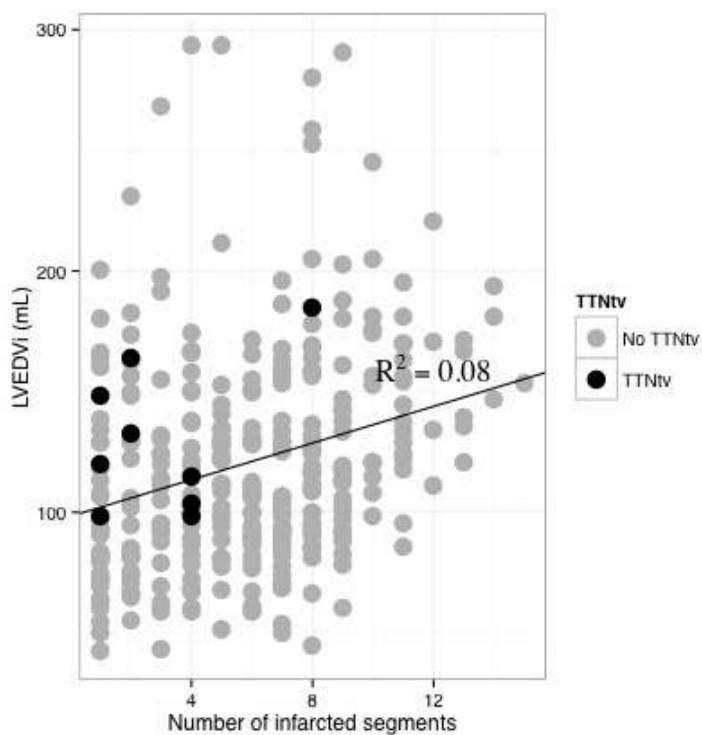
Figure 2-7 Scatterplot of subjects with CMR evidence of MI (n=335) showing LVEF and the number of infarcted segments



Scatterplot of subjects with CMR evidence of MI (n=335) showing Left Ventricular Ejection Fraction (LVEF) and the total number of infarcted segments in cases with or without a Titin Truncating Variant (TTNtv).

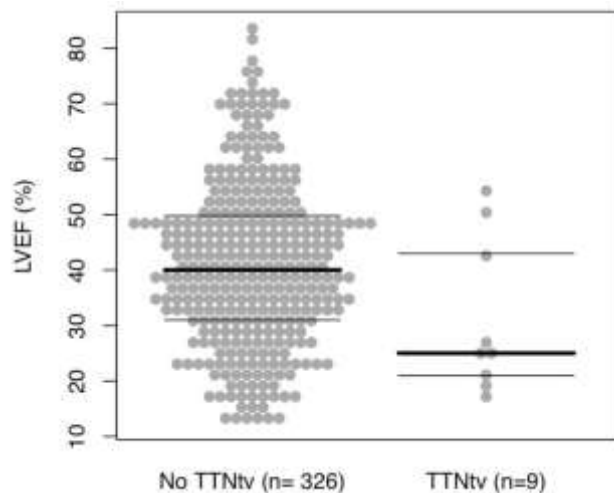
Linear regression line is shown.

Figure 2-8 Scatterplot of subjects with CMR evidence of MI (n=335) showing LVEDVi and the number of infarcted segments



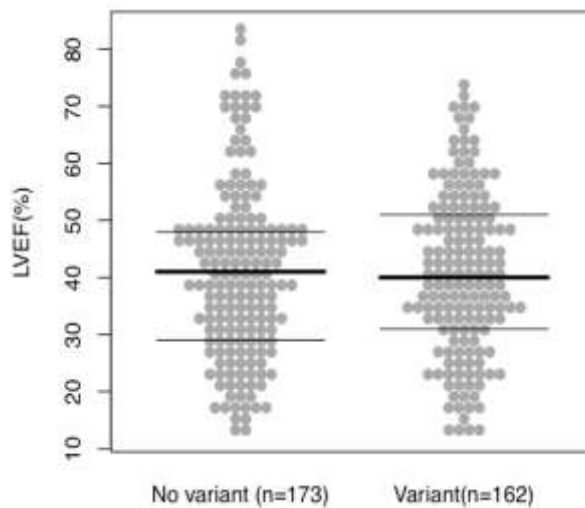
Scatterplot of subjects with CMR evidence of MI (n=335) showing Left Ventricular End Diastolic Volume indexed to Body Surface area (LVEDVi) and the total number of infarcted segments in cases with or without a Titin Truncating Variant (TTNtv). Linear regression line is shown.

Figure 2-9 Beeswarm plot of subjects with CMR evidence of MI (n=335) showing LVEF and presence or absence of a *TTNtv*



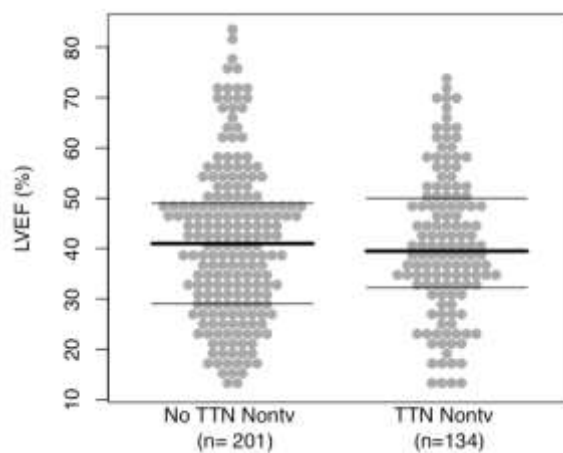
Beeswarm plot showing Left Ventricular Ejection Fraction (LVEF) and presence or absence of a Titin Truncating variant (*TTNtv*). Median values and interquartile ranges are overlaid.

Figure 2-10 Beeswarm plot of subjects with CMR evidence of MI (n=335) showing LVEF and presence or absence of variants in key DCM genes



Beeswarm plot showing LVEF in subjects with CMR evidence of MI grouped by the presence or absence of variants in key DCM genes other than *TTNtv*s. Median values and interquartile ranges are overlaid. DCM; Dilated Cardiomyopathy. LVEF; Left Ventricular Ejection Fraction. MI; Myocardial Infarction.

Figure 2-11 Beeswarm plot of subjects with CMR evidence of MI (n=335) showing LVEF grouped by the presence or absence of *TTN Non-tvs*



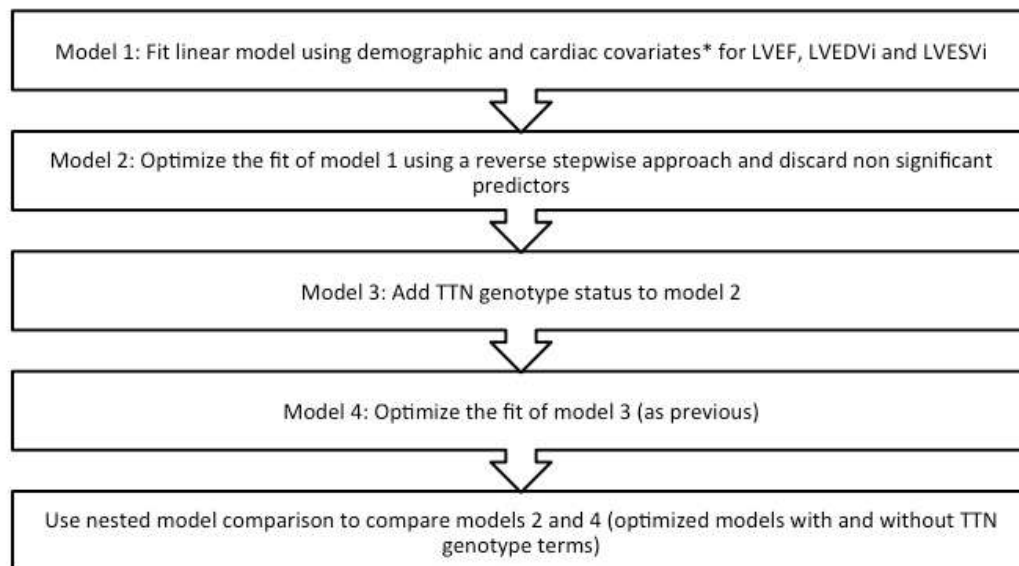
Beeswarm plot showing LVEF in subjects with CMR evidence of MI (n=335) grouped by the presence or absence of *TTN Non truncating variants (Non-tvs)*. Median values and interquartile ranges are overlaid. LVEF; Left Ventricular Ejection Fraction. MI; Myocardial Infarction.

2.3.3 Regression analysis

With the cohort of subjects with CMR evidence of MI, linear regression was used to further assess the relationship between *TTN* genotype and cardiac phenotypes. Multivariable models (see Figure 2-12) were generated for key morphological parameters (LVEF, LVEDVi and LVESVi) using all known demographic and clinical covariates only (termed Model 1) and then optimized using a reverse stepwise approach to minimize Akaike's Information Criterion (AIC). AIC is a goodness of fit measure, which is corrected for model complexity, with smaller values representing a better fit of data. This formal optimization step omitted numerous clinical covariates from the model (termed model 2). Model 3 represents the addition of genotype status (either the presence of a *TTN*tv, or the presence of a rare *TTN* missense variant) to the optimized model of clinical covariates (model 2). The formal optimization step was then repeated on model 3, to produce model 4. Analysis of Variance

(ANOVA) assessed the relationships between clinical covariates and the addition of *TTN* genotype between nested linear models (models 2 and 4). Results are outlined in Table 2-11, Table 2-13, Table 2-14 & Table 2-14.

Figure 2-12 Overview of the process of multivariable regression analysis



*Age, Sex, Ethnicity (Caucasian or other), total number of infarcted segments, infarct thickness, most affected territory (LAD, LCX, PDA, RCA), number of hibernating segments, infarct thickness, wall motion of most affected territory (normal, akinetic, dyskinetic, hypokinetic), mitral regurgitation (none, minimal, moderate, severe), multi-territory infarct, history of prior revascularization, and wall thickness of most affected territory (mm). Note that the presence of significant CAD is not in the model as the number of cases with missing data precluded its inclusion.

LVEF, Left Ventricular Ejection Fraction. LVEDVi, Left Ventricular End Diastolic Volume indexed to body surface area. LVESVi, Left Ventricular End Systolic Volume indexed to body surface area.

Table 2-11 Multiple regression model predicting LVEF

Multivariable regression model predicting LVEF by demographic, cardiac and genetic covariates.
For full details see text. Multiple R^2 is 0. 36, adjusted R^2 is 0. 34

Model terms		Coefficient	Std.Err	95 % CI	p
(Intercept)		69.3	3.4	62.7, 75.9	<0.00001
Presence of a <i>TTN</i> truncating variant		-17.6	4.1	-25.8, -9.5	<0.0001
Wall motion of most affected territory	Normal	0.0	-	-	<0.00001
	Hypokinetic	-13.7	3.6	-20.8, -6.7	
	Akinetic	-18.7	3.6	-25.7, -11.7	
	Dyskinetic	-27.4	4.2	-35.7, -19.1	
Mitral regurgitation	None	0.0	-	-	<0.0001
	Minimal	-4.2	1.4	-7.0, -1.4	
	Moderate	-9.9	2.4	-14.6, -5.2	
	Severe	-11.1	5.1	-21.1, -1.2	
Number of hibernating segments (per segment)		-1.0	0.2	-1.5, -0.6	<0.0001
Number of infarcted segments (per segment)		-0.8	0.3	-1.3-, 0.3	0.001
History of prior revascularization		3.2	1.4	0.5, 5.9	0.02
Non-Caucasian Ethnicity		3.7	1.7	0.3, 7.1	0.03

CI, Confidence Intervals; LVEF, Left Ventricular Ejection Fraction; Std.Error, Standard error.

The model was formally assessed to check that assumptions of linearity, randomness and homoscedasticity have been met (see appendix Figure 4-3). The model appears accurate and generalizable to the population.

Table 2-12 Multiple regression model predicting LVEDVi

Multivariable regression model predicting LVEDVi by demographic, cardiac and genetic covariates. For full details see text. Multiple R^2 is 0.34, adjusted R^2 is 0.31

Model terms		Coefficient	Std.Err	95 % CI	P
(Intercept)		87.9	18.5	51.6,124.3	<0.00001
Presence of a <i>TTN</i> variant		22.8	12.7	-2.2, 47.8	0.074*
Mitral regurgitation	None	0.0	-	-	-
	Minimal	11.7	4.5	3.0, 20.5	<0.0000000001
	Moderate	44.6	7.4	30.0,59.2	
	Severe	78.5	15.5	48.0,109.1	
Wall motion of most affected territory	Normal	0.0	-	-	-
	Hypokinetic	22.0	10.9	0.5, 43.4	<0.0001
	Akinetic	28.7	10.9	7.2, 50.3	
	Dyskinetic	58.7	12.9	33.3,84.1	
Total number of infarcted segments (per segment)		2.4	0.8	0.9, 3.9	0.002
Number of hibernating segments (per segment)		2.3	0.7	0.8, 3.7	0.002
Age		-0.6	0.2	-1.0, -0.2	0.005
Male Sex		11.8	5.9	0.3, 23.4	0.045
No history of prior revascularization		10.2	4.3	1.7, 18.7	0.02
Non-Caucasian Ethnicity		-12.4	5.3	-22.9, -1.9	0.02

CI, Confidence Intervals; LVEDVi, Left Ventricular End Diastolic Volume indexed to body surface area; Std.Error, Standard error. * Note the optimized model does not include *TTN* Truncating variants but this data is presented here for information, as it is the main covariate of interest.

Table 2-13 Multiple regression model predicting LVESVi

Multivariable regression model predicting LVESVi by demographic, cardiac and genetic covariates. For full details see text. Multiple R^2 is 0.35, adjusted R^2 is 0.33

Model terms		Coefficient	Std.Error	95% CI	P
(Intercept)		44.1	16.8	11.1, 77.1	0.009
Presence of a <i>TTN</i> variant		32.3	12.3	8.1, 56.4	0.009
Mitral regurgitation	None	0.0	-	-	<0.0000000001
	Minimal	10.5	4.3	2.1, 18.9	
	Moderate	39.0	7.1	25.1, 52.9	
	Severe	70.8	15.0	41.3, 100.3	
Wall motion of most affected territory	Normal	0.0	-	-	<0.0000001
	Hypokinetic	21.7	10.5	1.0, 42.4	
	Akinetic	29.7	10.5	9.0, 50.4	
	Dyskinetic	61.1	12.4	36.6, 85.5	
Number of hibernating segments (per segment)		2.9	0.7	1.5, 4.3	<0.0001
Total number of infarcted segments (per segment)		2.3	0.7	0.8, 3.7	0.002
No history of prior revascularization		10.6	4.1	2.5, 18.7	0.01
Age		-0.5	0.2	-0.8, -0.1	0.02

CI, Confidence Intervals; LVESVi, Left Ventricular End Systolic Volume indexed to body surface area; Std.Error, Standard error.

Table 2-14 Multivariable linear modeling of the relationship between *TTN*_{tv} genotype and phenotype for LVEF, LVEDVi and LVESVi in subjects with MI

Phenotype	Model terms	R ²	P value for model comparison
LVEF	Age, sex, ethnicity, total number of infarcted segments, wall motion of most affected territory, hibernating segments, history of prior revascularization and mitral regurgitation ¹	0.31	0.00005
	Age, sex, ethnicity, total number of infarcted segments, wall motion of most affected territory, hibernating segments, history of prior revascularization, mitral regurgitation and presence of a <i>TTN</i>_{tv} ²	0.35	
LVESVi*	Age, sex, ethnicity, total number of infarcted segments, wall motion of most affected territory, hibernating segments, history of prior revascularization and mitral regurgitation ¹	0.33	0.0107
	Age, sex, ethnicity, total number of infarcted segments, wall motion of most affected territory, hibernating segments, history of prior revascularization, mitral regurgitation and presence of a <i>TTN</i>_{tv} ²	0.34	
LVEDVi*	Age, sex, ethnicity, total number of infarcted segments, wall motion of most affected territory, hibernating segments, history of prior revascularization and mitral regurgitation ¹	0.31	0.0731
	Age, sex, ethnicity, total number of infarcted segments, wall motion of most affected territory, hibernating segments, history of prior revascularization, mitral regurgitation and presence of a <i>TTN</i>_{tv} ²	0.31	

Multivariable linear regression model of the relationship between Titin (*TTN*) genotype and phenotype for Left Ventricular Ejection Fraction (LVEF), Left Ventricular End Diastolic Volume (LVEDV) and Left Ventricular End Systolic Volume (LVESV) in subjects with Myocardial Infarction (MI). * Values are indexed to body surface area.

For each of LVEF, LVEDVi and LVESVi two models are compared. First, LVEF, LVEDVi and LVESVi are modeled as a function of demographic and clinical covariates, and the models optimized (see methods). Retained predictors are listed in column three. Next *TTN* genotype is added to the model. A *TTN*_{tv} may be present or absent and a rare *TTN* missense variant may be present or absent. The model is re-optimized again and significant predictors retained. The two models are compared using ANOVA. The p-values for comparison of models after addition of *TTN* status are provided.

¹ Covariates entered into the model but omitted following optimization were infarct thickness, location of most affected territory, the wall thickness of most affected territory and presence of a multi-territory infarct. Note that the presence of significant CAD was not included in the original model due to the high level of missing data (n=125).

² Presence of a rare *TTN* Missense variant omitted after optimization.

2.3.4 Healthy Volunteer Cohort

A total of 431 healthy volunteers (HVOLs) were sequenced for variation in *TTN* and additional key DCM genes, using the same methodology as disease cohorts. All HVOLs have undergone CMR as detailed previously (118) and have no current cardiac phenotype. A total of 11 *TTN* variants were identified by NGS and confirmed by Sanger sequencing. Therefore, the prevalence of *TTN* variants in this cohort is 2.6%. The characteristics of these *TTN* variants are outlined in Table 2-15.

Table 2-15 *TTN* variants in the Healthy Volunteer cohort (n=431)

BRU Code	Variant Type	Transcript Variation	Protein Variation	Exon	Protein Band	PSI*
14RH013531	Canonical splice variant	c. 11254+2T>C		46	I Band	0.07
14AO013721	Canonical splice variant	c. 11254+2T>C		46	I Band	0.07
14SS01830	Nonsense	c. 10852C>T	p. Gln3618Ter	46	I Band	0.07
14ZN01340	Frameshift	c. 17508dupA	p. Gly5837ArgfsTer9	61	I Band	0.47
14JD01896	Nonsense	c. 21142C>T	p. Arg7048Ter	74	I Band	0.36
14EC01433	Canonical splice variant	c. 30803-2A>G		115	I Band	0.84
14AH015392	Canonical splice variant	c. 32554+1G>C		131	I Band	0.17
14MO01427	Frameshift	c. 45391delA	p. Ile15131TyrfsTer46	247	I Band	1.00
14SM01546	Nonsense	c. 67348C>T	p. Gln22450Ter	319	A Band	1.00
14JM01448	Frameshift	c. 67159delA	p. Ile22387Ter	319	A Band	1.00
14JC01930	Nonsense	c. 98506C>T	p. Arg32836Ter	353	A Band	1.00

Sanger sequencing validated all *TTN* variants included in this analysis. Targeted re-sequencing used the same methodology as the Myocardial Infarct cohort and severe end-stage ischaemic cardiomyopathy cohort (SureSelect Targeted capture and sequencing on the SOLiD 5500 platform). *Represents the median Percentage Sliced In (PSI) value. For full explanation see methods. ENST00000589042. ENSP00000467141

2.3.5 End stage ischaemic cardiomyopathy cohort

A total of 95 cases of end stage ischaemic cardiomyopathy were sequenced for variation in *TTN* and additional key DCM genes, using the same methodology as the cohort of subjects with CMR evidence of MI and Healthy volunteers. A total of three *TTN* variants were identified by NGS and confirmed by Sanger sequencing. Two of the *TTN* variants were identified in a single case. Therefore the prevalence of *TTN* variants in this cohort is 2.1% (see Table 2-16). The characteristics of these *TTN* variants are outlined in Table 2-17.

Table 2-16 Prevalence of *TTN*tv_s in subjects with end stage ischaemic cardiomyopathy compared to background population

Study cohort (n=95)	As compared to <i>TTN</i> tv burden in our HVOL Cohort (2.6%) *			As compared to the <i>TTN</i> tv burden in ExAC (1.13%)**		
<i>Subjects with a TTN</i> tv +	Odds Ratio	95% CI	P value	Odds Ratio	95% CI	P Value
2 (2.1%)	1.01	0.10, 4.99	1	1.88	0.22, 7.02	0.29

The prevalence of Titin Truncating Variants (*TTN*tv_s) in subjects with severe end stage ischaemic cardiomyopathy compared to the background population rates.*Our Healthy Volunteers (HVOL) control cohort were simultaneously recruited and sequenced with the same methodology as MI subjects. 11 *TTN*tv_s were identified in 431 cases (2.6%).**The ExAC data represents all *TTN*tv (n=640) in the 363 meta-transcript exons of *TTN* with a PASS filter in ExAC, v0.3. For calculating the frequencies, the mean of the total allele count (n=113284) was used as the denominator rather than the total allele count in ExAC (n=121412). The mean of the total allele count varies throughout the gene/genome based on the coverage/quality of sequencing at each position (personal communication with R Walsh). The mean total allele count of 113284 is equivalent to 56642 controls.

Table 2-17 *TTN*tv_s in the end-stage Ischaemic Cardiomyopathy cohort (n=95)

BRU Code	Variant Type	Transcript Variation	Protein Variation	Exon	Protein Band	PSI**
20KP00984*	Frameshift	c. 11183dupG	p. Leu3729ThrfsTer9	46	I Band	7
20KP00984*	Frameshift	c. 17823delA	p. Ile5941MetfsTer8	62	I Band	39
20GC00999	Frameshift	c. 83504_83505delCT	p. Ser27835LeufsTer10	327	A Band	100

Sequenced using the same methodology as the MI cohort. SureSelect Targeted capture and sequencing on the SOLiD 5500 platform. All validated by Sanger sequencing. * Two *TTN*tv_s identified in the same case. **Represents the median Percentage Sliced In (PSI) value. For full explanation see methods. ENST00000589042. ENSP00000467141

2.4 Discussion

The primary aim of this study was to identify if *TTN*tv_s, which are common in the general population, and are the major genetic cause of DCM, play an important role in LV dysfunction in the post MI patient with impaired ventricular function. The data show an increased burden of *TTN*tv_s in subjects with CMR evidence of MI and severe LV dysfunction as compared to both subjects with CMR evidence of MI, without severe LV dysfunction and background population rates. There is no evidence from this data that other key DCM genes or *TTN* nonTV_s play a role in LV dysfunction in the post MI patient.

2.4.1 Cohort of subjects with CMR evidence of MI

The first hypothesis was that in subjects with CMR evidence of MI and severe LV dysfunction (LVEF ≤30%) there an increased mutation burden of *TTN*tv_s. This cohort of subjects with CMR evidence of MI (see Figure 2-6), consists of 335 largely male, largely

Caucasian subjects (see Table 2-5 and Table 2-7 for full demographic data) with a range of LV function. Analyses show that the prevalence of *TTN*tv is significantly enriched in those subjects with severe LV dysfunction (LVEF $\leq 30\%$) compared to both subjects without severe LV dysfunction (LV $> 30\%$) (6.98% vs 1.21, $p=0.01$) and background population rates (see Table 2-9). There was a significant difference exon expression level of *TTN*tv associated with or without severe LV dysfunction (see Table 2-10). The size of infarct was associated with a lower LVEF (see Figure 2-7), although the majority of subjects with *TTN*tv had small infarct sizes with severely reduced LVEF (see Figure 2-6, Figure 2-7 and Table 2-7). In those subjects with *TTN*tv there was no evidence they had hypokinetic but non -dilated cardiomyopathy (see Figure 2-8).

Considered from an alternative viewpoint, I note that patients with a *TTN*tv had lower LVEFs than those without ($31.2 \pm 13.9\%$ vs $41.2 \pm 14.7\%$; $p=0.026$). An LVEF $< 30\%$ is used to guide device therapy in post-MI patients, in this group *TTN*tv were significantly enriched as compared to patients with an MI and higher LVEFs (6.98% vs 1.21, $p=0.01$). Whether these findings could have prognostic importance or influence clinical management will require further investigation.

Multiple regression models predicting LVEF, LVEDVi and LVESVi were undertaken, taking into account cardiac, non-cardiac and genetic covariates (see Figure 2-12, Table 2-13 and Table 2-14). The model predicting LVEF explains $\sim 37\%$ of the variance in LVEF post MI, with the presence of a *TTN*tv predicting an absolute reduction in LVEF of $\sim 17\%$. The presence of a *TTN*tv was associated with a significant difference in LVESVi and a trend towards significance for LVEDVi.

The second hypothesis was that in subjects with CMR evidence of MI and severe LV dysfunction (LVEF $\leq 30\%$) there is an increased mutation burden in key DCM genes (other than *TTN*tv). In this cohort there was no evidence of an enrichment of *TTN* nonTVs or other key DCM gene variants in those cases with severe LV dysfunction as compared to those

without severe LV dysfunction (see Figure 2-10 and Figure 2-11). *TTN* missense variants are common in the population, but currently their role in disease remains poorly understood. Due to the small study size, in each analysis (*TTN*_{tv}, other key DCM genes and *TTN* non TV's) I have considered only heterozygous variants and have not explored the possible role of homozygous or compound heterozygous variants as important causes or modifiers of LV dysfunction in the post MI patient. It may be that future studies elucidate which *TTN* missense variants are associated with important phenotypic effects.

The third hypothesis was that genetic information provides additive prognostic information in individuals with CMR evidence of MI. Outcome data on this cohort has been collected and analysis of this data is awaited. In this cohort of 335 patients with CMR evidence of MI, there were nine subjects with *TTN*_{tv}s. Five of these nine subjects (see Table 2-7) met diagnostic criteria for DCM (117) but equally had infarcts that had been discounted as important based on infarct size and/or patent epicardial coronary arteries (that does not exclude occlusive MI). All five cases had clear evidence of MI on CMR and supporting evidence of IHD (see Figure 2-6 and Table 2-7). Anecdotally, the findings of severe LV dysfunction in association with a small MI, which is insufficient to explain the degree of LV dysfunction is a not an uncommon finding. These patients are often labelled as a possible “dual pathology”. If outcome data from this cohort identifies a difference for those subjects with severe LV dysfunction and CMR evidence of MI with or without a *TTN*_{tv} then this could influence the clinical management of these patients and potentially their family members.

2.4.2 Severe end stage ischaemic cardiomyopathy cohort

There were three *TTN*_{tv}s identified in a total of two subjects within the severe end stage ischaemic cardiomyopathy cohort giving a prevalence of 2.1%, which was not significantly increased over the background population rate (comparison with ExAC p=0.29).

This is the only occurrence of two *TTN*tv's in a single individual in any of cohorts included here. These variants were Sanger validated so cannot be attributed to sequencing errors. I note the PSI in both of these variants was reduced and so these are non-constitutively expressed exons. While one variant (c.17823delA; (p.Ile5941MetfsTer8)) has a PSI of 39% and is novel, the second variant (c.11183dupG; (p.Leu3729ThrfsTer9)) has a PSI of 7% and is seen in ExAC with a frequency of 0.0001. The frequency of this second variant in the population and the associated low PSI would mean it is less likely to be associated with a significant phenotypic effect. It is not possible to confirm whether the two variants found in the single subject are in *cis* or *trans* from current sequencing data. We have limited phenotype data on these samples and so it is possible that this subject had evidence of skeletal muscle weakness but it would seem unlikely that a cardiac phenotype in such a patient would have been labelled as 'Ischaemic Cardiomyopathy'.

This cohort was not examined in detail to identify if other key DCM genes or *TTN non TV's* were enriched as the small cohort size (n=95) meant that statistically significant findings were unlikely.

There are a number of potential reasons beyond simple chance why this relatively small study group did not have an increased prevalence of *TTN*tv's. There is little additional phenotype data on this cohort but it may differ in terms of age, sex, ethnicity and family history compared to our MI cohort of subjects with CMR evidence of MI. For example, where individuals are undergoing cardiac transplant assessment then perhaps the severity of this disease may encourage family members to present for clinical screening. If additional family members were found to have any evidence of cardiomyopathy it is likely to have lead to a review of diagnostic label in the proband, such as DCM rather than Ischaemic Cardiomyopathy. Conceivably the end stage ischaemic cardiomyopathy cohort could represent subjects who effectively have a reduced chance of having a genetic basis for their disease, as those with any potential familial involvement have been selected against.

Analysis of the outcome data on the cohort of subjects with CMR evidence of MI may suggest additional causes of potential selection bias in the end stage, ischaemic cardiomyopathy cohort. Aspects to consider include whether subjects with *TTN*ts may have more severe LV dysfunction but actually not otherwise reach the criteria for assessment for heart transplantation, which is based on multiple factors (163). Additionally, earlier studies within our group have demonstrated an increased incidence of sustained VT in *TTN*tv positive subjects with DCM (117), so perhaps in an ischaemic cardiomyopathy population *TTN*ts could be associated with an increased risk of sudden death, meaning fewer *TTN*tv positive patients live long enough to undergo assessment for cardiac transplantation.

2.4.3 Study design limitations

In addition to the relatively small study size (n=335), the study design format must be considered in the interpretation of these results. In this single centre study, the MI cohort consisted of patients who had all been referred for a CMR, which is not a routine baseline investigation for all individuals with IHD in our institution but is the only technique that will detect small MIs. A number of referrals will have been made due to a query over viability or diagnostic uncertainty (Table 2-8), which may bias the IHD population examined here. Furthermore, CMR was not undertaken specifically to identify or exclude MI and protocols were not optimized to rule out evidence of MI. Kim *et al* (129) reported a multicenter, double-blinded, randomized trial evaluating the efficacy of infarct detection by LGE in 282 patients with acute MI (defined as <17 days post MI) and an equal number of patients with chronic MI (defined as 17 days to 6 months post MI). From the lowest dose of gadolinium contrast (0.05 mmol/kg of gadoversetamide) used the highest (0.3 mmol/kg of gadoversetamide) the sensitivity of CMR for detecting MI increased for acute MI from 84 to 99% and from 83 to 94 percent for chronic MI (133). It is feasible that MI, particularly

chronic MI will have remained undetected in a number of patients referred for CMR in our institution, who will therefore not have met inclusion criteria for the study.

2.4.4 Limitations of phenotyping

Baseline phenotype data on the MI cohort including medical therapy, family history and the presence of significant CAD (see appendix Figure 4-2) were recorded at the time of recruitment to the Biomedical Research Unit, which was usually at the time of CMR, but also at later time points such as at routine clinical review. A number of factors relating to the routine clinical management of patients over this time period could reduce the power of this study to detect genotype-phenotype correlation. The first factor is that typical medication prescribing rates may have changed as guidelines and knowledge on the beneficial effects of pharmacological therapy on remodeling post MI improved (153, 154) enhanced. Secondly, access to coronary angiography to establish the presence of significant coronary artery disease, which is potentially amenable to revascularisation, may have improved. Lastly, device therapies such as ICD's have become increasingly commonly used. Such devices may have a beneficial effect on remodeling (153) but their presence is likely to preclude a CMR.

2.4.5 Limitations of NGS sequence data

All cohorts in this study were sequenced concurrently, using the same methodology allowing unbiased comparisons despite any potential unique biases or errors of the sequencing platform. The highly modular nature of *TTN*, which involves repeated sequences with high homology and it's enormous size meant that such a study would not have been feasible prior to the advent of NGS technology (115). NGS also allowed multiple additional genes to be simultaneously sequenced for inclusion in the study. The callability for each of the DCM genes of interest was equivalent across each cohort (see Table 2-2). The quality of *TTN* sequencing across cohorts was high (see Figure 2-5) although regions within the I-band of

Titin (exons 150-220; 6315bp; 2105aa) representing ~5% of the gene (118) exhibit poor callability. The exons in this region are small and contain repetitive sequencing which affects capture and mapping, as has been seen previously (118). Although *TTNtv*'s may have been missed in this region, the coverage varied little between cohorts allowing valid comparisons to be made. Future studies of *TTN* which incorporate longer read lengths will address the issues of alignment and mapping which are evident with mapping short reads against such a repetitive sequence (115) and identify if *TTNtvs* are a frequent finding in this region.

A total of 41 *TTNtvs* were identified by NGS across the three cohorts in this study, of which 23 validated by Sanger sequencing (56%). Those not validated by Sanger sequencing, almost exclusively had low coverage (see Table 2-3). The exception to this was sample 10TM00255 in which a nonsense variant (c.85417G>T; p.(Gly28473*)) was identified at a coverage level of 92X. NGS sequencing errors would be expected to be low at this level of coverage. However, this variant had an allelic balance of 0.16. The allelic balance refers to the number of alternate alleles divided by the total number of alleles sequenced (alternate and reference alleles). A heterozygous variant would be expected to have an allelic balance of ~0.5. The high level of coverage and low allelic balance raised the possibility that the nonsense *TTNtv* could be a mosaic variant, as has been identified in other studies employing NGS (164). The sample was further sequenced on an alternative NGS platform (MiSeq). No evidence of the *TTNtv* was identified and therefore no evidence that this was a mosaic variant.

2.5 Conclusion

This study identifies a novel role for truncating variants in *TTN*, a DCM gene, in post-MI systolic dysfunction. Intriguingly, the effect size of the *TTNtv* is equal or greater than many of the infarct covariates (e.g. number of infarcted segments) used to guide re-vascularisation therapy. Based on these findings it will be important to further explore the relationship

between TTNtv and post-infarct remodeling and function, and to investigate whether genetic stratification of the post-MI patient can inform treatment strategies.

2.6 Outline of further work

Outcome data on the MI cohort has been captured by patient questionnaire with verification and additional information obtained where possible by hospital and GP records. The primary endpoint of interest is all-cause death. The secondary end point is the composite of all cause death or major adverse cardiac event (cardiac arrest/ transplant/ CRT/ICD/LVAD).

Analysis is on going to determine if genetic information can provide additive prognostic information in individuals with CMR evidence of MI.

3 Chapter 3: Identification of the genetic basis of an ultra-rare cardiac phenotype: The application of whole exome sequencing in Histiocytoid Cardiomyopathy

3.1 Ultra-rare cardiac phenotypes

Ultra-rare disorders are considered to be those with a prevalence one-thousandth of rare disorders, so $<1:2,000,000$. Although by definition these disorders are uncommon, together they form a considerable disease burden (165). Recently, whole exome sequencing has proven effective in delineating the genetic aetiology of many such phenotypes (166, 167). In addition to the diagnostic benefit for the individual and their family, there are potential opportunities to identify underlying molecular pathways, which could suggest new therapeutic avenues. The benefits are not limited to rare or ultra rare diseases and similar approaches to those described here could equally be applied to more common phenotypes for which the genetic basis in an individual remained unsolved.

3.1.1 Introduction to Histiocytoid Cardiomyopathy

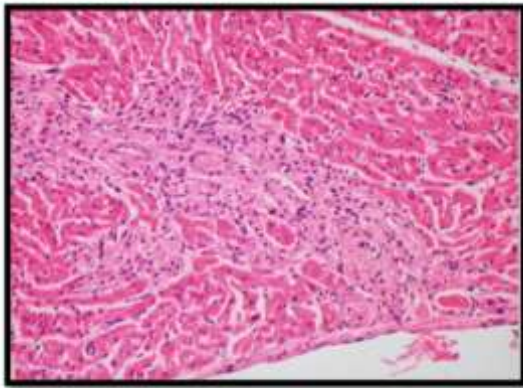
Histiocytoid Cardiomyopathy (CM) (OMIM#500000) is an ultra-rare disorder, characterized by the presence of histiocytoid cells within the myocardium (see Figure 3-1), sub-endocardial nodules (see Figure 3-2) and increased heart weight. It was first described in 1962, as ‘arachnoidosis of the myocardium (a contribution to the problem of rhabdomyoma of the heart)’ (168). Over 150 cases have since been published, under numerous synonyms including infantile cardiomyopathy with histiocytoid change (169), oncocytic cardiomyopathy (170), myocardial or conduction system hamartoma (171), infantile xanthomatous cardiomyopathy (172, 173), focal lipid cardiomyopathy (174), isolated cardiac lipidosis (175), Purkinje cell hyperplasia (176), foamy myocardial transformation of infancy (177), peculiar focal myocardial degeneration (178) and idiopathic infantile cardiomyopathy (179). It is unclear whether all of these diagnostic labels truly represent the same

underlying pathology, Malhotra *et al* (180) suggest that cases reported as congenital glycogenic tumours of the heart, familial cardiac lipidosis, and foamy myocardial transformation in a child with abnormal mitochondrial respiratory chain function may represent disorders other than Histiocytoid CM.

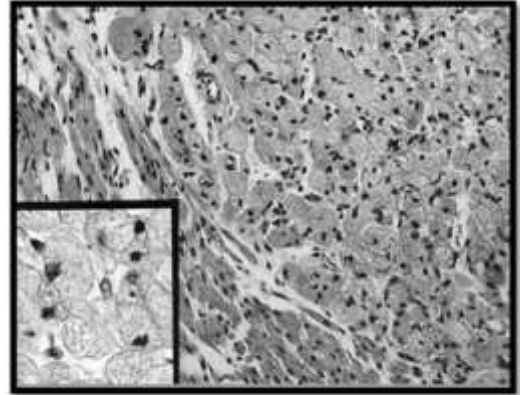
All regions of the heart including the accessory atrioventricular conduction pathways may involve foci of abnormal myocytes (180). Histologically these foci consist of ill-defined islands of large polygonal histiocytoid cells (altered cardiac myocytes) with a centrally located dark nucleus and light finely granular eosinophilic or foamy cytoplasm without cross-striations (181). The abnormal cells may exhibit striking mitochondrial hyperplasia (78). The original designation of the term “Histiocytoid” stems from the morphological resemble of the abnormal cells to lipid-laden histiocytes (a type of immune cell). Later investigations revealed these cells to contain myofibrils interposed with Z lines, indicating these were myocardial cells (178, 182). The structural features of histiocytoid cells detected in the myocardium are similar to those of oncocytic cells, also described in association with Histiocytoid CM in a wide range of other organs (180). Oncocytes are epithelial cells with abundant, granular, eosinophilic cytoplasm due to presence of numerous large mitochondria of varied sizes (183).

Figure 3-1 Histology of Histiocytoid Cardiomyopathy

A



B



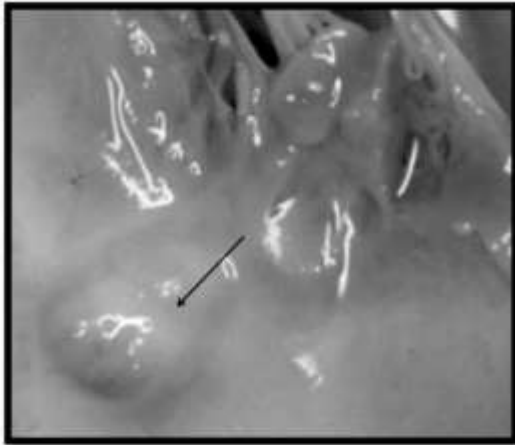
A: Histology from a child who died suddenly. Tetralogy of Fallot was evident at post-mortem, but the myocardium was otherwise unremarkable. The histological hallmarks are collections of pale cells with very finely vacuolated cytoplasm resembling histiocytes among normal myocytes. Ultra-structurally these cells contain large number of mitochondria (184). Image reproduced with permission of the rights holder, Dr Michael Ashworth..

B: Histiocytoid cells within the myocardium showing large round myocytes with smooth borders, granular and eosinophilic cytoplasm, and small irregular nuclei (185). Image reproduced with permission of the rights holder, BMJ Publishing Group Ltd.

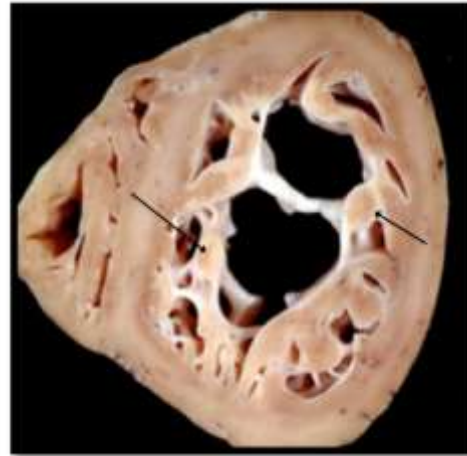
On macroscopic examination the heart may appear enlarged but otherwise unremarkable or discrete yellow-tan nodules may be present (186). In a series of 53 cases Malhotra *et al* (180) noted the characteristic morphological findings of sub-endocardial, epicardial, or valvular discrete nodules in 16/53 cases. These nodules may mimic a mass lesion in the sub-endocardial and sub-epicardial regions (181). Nodules (see Figure 3-2), may be scattered throughout the myocardium or localised to the valves or sub-endocardium; their size may range from a few millimetre to over one centimeter (184).

Figure 3-2 Macroscopic findings in Histiocytoid Cardiomyopathy

A



B



A Sub-endocardial nodule (0.5 cm) on the atrial aspect of the tricuspid valve (185). Image reproduced with permission of the rights holder BMJ Publishing Group Ltd.

B Explanted heart; showing coarse trabeculations and deep recesses consistent with Left Ventricular Non Compaction. There are numerous well-delineated pale yellow nodules within the sub-endocardium (187). Image reproduced with permission of the rights holder John Wiley and Sons.

3.1.2 Natural history and prevalence of Histiocytoid CM

Throughout the literature, a very strong female bias (4:1) is described (188), although males with Histiocytoid CM are rarely reported (189). Reports are predominantly in Caucasians (80%) (190). Most cases (90%) present in female children under two years of age (188) with severe cardiac dysrhythmia, cardiac arrest simulating sudden infant death syndrome or with DCM leading to fulminant cardiac failure (188). In the largest case series of 53 patients, sudden death was the presenting feature in 11 cases (180). Malignant arrhythmia in an older female child (aged 11 years) with only mild features of cardiomyopathy has been reported, with the diagnosis made based on histology at post-mortem (191). Initial presentations with more ‘benign’ arrhythmias such as Wolf Parkinson White (WPW) syndrome are also widely recognised (185, 192).

In the largest paediatric autopsy series to date of Sudden Unexplained Cardiac Death (SUCD) (n=103) attributable to undiagnosed structural heart disease (morphologically normal hearts

were excluded), Ilina *et al* (193) identified two cases of Histiocytoid CM (1.9%). Although rare in comparison to the more common causes of myocarditis (35.9%), hypoplastic left heart syndrome (18.4%) and DCM (16.5%), multisystem cases in this study were excluded, which may have resulted in under ascertainment of Histiocytoid CM cases (see Table 3-1). The true prevalence of Histiocytoid CM is unknown and likely to be underestimated for several reasons. Some cases may be labelled as “cot death” (194) or Sudden Infant Death Syndrome (SIDs) (195) where the characteristic histological features may either not be evident or may not be recognised, particularly likely in health systems where detailed examination of the heart by a paediatric pathologist with a specialist interest in cardiac disorders is uncommon. Kearney *et al* (196) noted that in many series of tumours in children, myocardial hamartoma (aka Histiocytoid CM) were not reported; in their study, most cases did not present as a distinct mass with increased vascular supply or a discrete nodule and they comment that unless the tumour is specifically examined microscopically, its gross appearance is very similar to fibrosis and thus cases may be overlooked. In addition, the numerous synonyms may lead to failure to appreciate individuals affected with the same disorder. Finally, in the vast majority of cases the diagnosis is made based on characteristic findings at post-mortem or post cardiac transplantation examination of the native heart. These patients may represent just the “tip of the iceberg” of the most severely affected patients.

3.1.3 Multisystem involvement in Histiocytoid CM

Cardiac and extra-cardiac anomalies have frequently been noted (180). Associated cardiac abnormalities include LVNC (187, 197), endocardial fibroelastosis (198, 199), Teratology of Fallot (184, 190), atrial septal defects (196, 200) and hypoplastic left heart (190, 201). The high prevalence of anomalies involving the nervous system and eyes (see Table 3-1) raised

the possibility of a mitochondrial aetiology (78). The presence of multiple organ involvement is an important diagnostic clue in mitochondrial disorders (82).

Table 3-1 Extra-cardiac anomalies reported in association with Histiocytoid CM

Cranial/ Brain	Reference	Ophthalmic	Reference
Microcephaly	(202)	Microphthalmia	(170, 192, 196)
Agenesis corpus callosum	(170)	Corneal opacities	(192, 196)
Hydrocephalus	(192)	Oblong pupils	(196)
		Aphakia	(170)
Other		Megalocornea	(170)
Cleft lip or palate	(200)	Cataract	(170)
Laryngeal web	(196)	Bilateral retinal hypoplasia	(78)
Peripheral hypotonia	(200)		
Oncocytic cells in the adrenal or thyroid glands	(203)		
Diffuse steatosis of the liver	(78)		
Meckel's diverticulum	(196)		

Note this is not an exhaustive list, but is intended to illustrate that on detailed examination (often post-mortem) Histiocytoid CM is frequently a multi-system disorder.

3.1.4 Familial cases

A familial tendency in Histiocytoid CM was first reported over 40 years age (177, 202) with reports of a brother and sister with a form of infantile cardiomyopathy characterised by accumulation of mitochondria in the cytoplasm of myocardial fibres. Onset was at birth and four weeks of age respectively, with death was from congestive heart failure at 19 days and four months. Both siblings had microcephaly. Severe mitochondrial changes were noted in the myocardial fibrils in addition to the accumulation of neutral fat. Suarez *et al* (177) reported five cases termed “Foamy myocardial transformation of infancy”, including two pairs of affected sibs (one pair of female sibs and one pair of male sibs) and comment that while autosomal recessive inheritance would fit with the cases they describe it wouldn't explain the excess of affected females. Subsequent reports have suggested a familial tendency in ~5% of cases overall (195), though this estimate may have a downward bias due to the incidence of recurrent miscarriage in some families (204).

3.1.5 Clues to the molecular basis of Histiocytoid CM

Despite the distinct pathological features, the underlying molecular basis of Histiocytoid CM remained unclear at the outset of this study. Hypotheses relating to underlying aetiology ranged from a developmental anomaly of the atria-ventricular conduction system, multifocal tumours of Purkinje Cells, developmental arrest of cardiomyocytes with aggregations of haematoma-like cardiomyocytes resembling oncocytes (186, 205), myocardial hamartoma (180) or prenatal (post-viral) myocardial injury (180). It was proposed that Histiocytoid CM is a mitochondrial cardiomyopathy based on the cytoplasmic appearance of affected cells, which is due to an extensive accumulation of mitochondria. Cells are morphologically abnormal and associated with decreased mitochondrial enzyme activity (78, 180, 186, 205). Mitochondrial cardiomyopathies are associated with deficiencies in the function of mitochondrial enzyme systems, responsible for oxidative phosphorylation and composed of five enzyme complexes of the Mitochondrial Respiratory Chain (MRC) (180). Mutations in either mitochondrial or nuclear DNA may lead to mitochondrial disorders, which are highly variable and may affect single or multiple organs. Early support for a mitochondrial hypothesis included reports of a biochemical deficiency of reducible cytochrome b (respiratory complex III) in the myocardium, which was noted in one patient over thirty years ago (180, 206). Subsequently Andreu *et al* (207) revisited this case and applied molecular studies, reporting a missense mutation in the mitochondrial cytochrome b gene and therefore the first molecular defect in association with Histiocytoid CM, in a case of a multisystem disorder. Although the cardiac features in this case were thought similar to other cases of Histiocytoid CM, the patient presented with failure to thrive and did not manifest arrhythmias (180) leading to suggestions that this may have been another disorder.

Later, Vallance *et al* (78) reported a sporadic case of a female infant with the A8344G mtDNA mutation, detected in both liver and cardiac muscle tissue, she had presented with failure to thrive and sudden unexpected death at 11 months of age. Post mortem examination

showed Histiocytoid CM, diffuse steatosis of the liver and bilateral retinal hypoplasia. The cardiac myocytes showed severe mitochondrial hyperplasia on electron microscopy, with dispersion of the sarcomeres. Special stains of the frozen heart muscle revealed an absence of complex IV (cytochrome c oxidase) in many of the myocytes. The A8344G mtDNA mutation is most commonly associated with a clinical presentation of the triad of Myoclonic Epilepsy, Myopathy, and Ragged Red Fibers (MERRF). Both the mtDNA mutations reported by Andreu (207) and Vallance (78) seem to be sporadic and have not been replicated in additional cases (195). Further studies by Shehata *et al* (195) used whole genome expression analysis in 12 Histiocytoid CM cases and 12 age matched controls and identified two significantly down regulated gene sets, the first at 1q21.3 and the second at 2q12.1, no clear candidate genes suggested.

Recently, Cataldo *et al* (208) reported a single female infant with LVNC, WPW and Histiocytoid CM in whom they identified compound mutations in *CACNA2D1* and *RANGRF* mutations. Mutations in *CACNA2D1* have previously been reported in Brugada syndrome (209), Sudden Unexplained Death (SUD) (210), epilepsy and intellectual disability(211). Variants in *RANGRF* have been linked with Brugada syndrome, though the role is uncertain (212).

The strongest clues to the molecular aetiology appeared to be the marked female predominance, leading to suggestions that Histiocytoid CM is an X-linked genetic disorder (213) with prenatal lethality in males (195) and the large excess of sporadic cases of Histiocytoid CM which suggests that *de novo* mutations may be the underlying cause of the disorder, as has been seen in other rare Mendelian diseases (214). The causative gene may have a role in mitochondrial function (215).

3.1.6 Aim

The aim of this study was to utilize the power of WES to delineate the underlying molecular basis of disease in a subject with an ultra-rare cardiac phenotype, Histiocytoid CM. To provide diagnostic genetic information for the family and identify if there were any insights from associated molecular pathways, which could have implications for other cardiac diseases.

3.1.7 Hypothesis

From the background literature on Histiocytoid CM, where the majority of cases are sporadic, a *de novo* dominant disease is most likely. An X-linked disorder could explain the excess of female cases.

3.2 Methods

3.2.1 Ethics approval

An infant with a diagnosis of Histiocytoid CM and both unaffected parents were recruited through the NIHR Cardiovascular Biomedical Research Unit at the Royal Brompton and Harefield NHS Foundation Trust. Four additional unrelated probands were recruited via UK regional genetics services. Studies were performed according to institutional guidelines, with ethical approval.

3.2.2 Whole Exome Sequencing approaches

WES was undertaken using DNA extracted from EDTA venous blood samples. We would not expect a target such as Histiocytoid CM, which is ultra-rare and has very specific characteristics, to exhibit marked genetic heterogeneity.

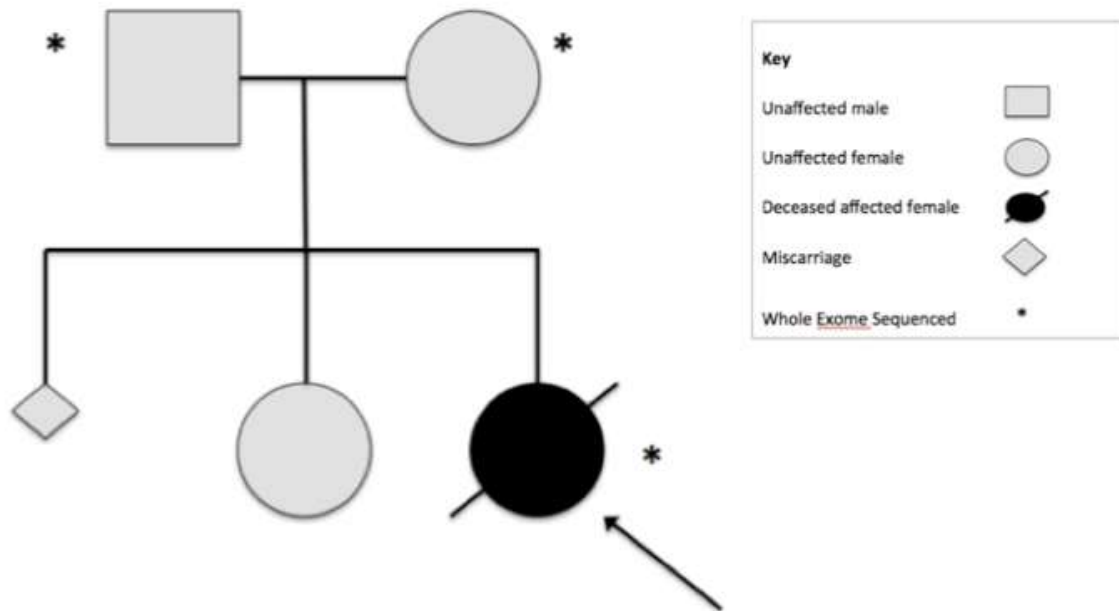
Two independent approaches were taken. The first was a family-based approach, involving a “Trio” where there was an affected child (Subject 1, referred to as the proband) and both unaffected parents (see Figure 3-3). An excellent candidate gene (*NDUFB11*) was identified

(see results and discussion) and Sanger sequencing was undertaken in four additional unrelated cases to replicate these findings. When no further variants in our candidate gene were found we undertook a second approach to WES using an “overlap” approach of multiple unrelated affected singleton cases.

3.2.2.1 Case details of subject 1 (affected child from family ‘trio’): a female infant with Histiocytoid Cardiomyopathy (CM) and multisystem features

The proband was the second child of non-consanguineous parents (maternal age 36 years, paternal age 39 years). See Figure 3-3. There was a maternal history of an early miscarriage, at around six weeks gestation. Both parents are in good general health. There was a maternal history of “fainting episodes” in adolescence, which spontaneously resolved with age. No other family history of note. The antenatal history was of early pregnancy bleeding for first six weeks. The ultrasound scan at 18/40 was normal. Subsequently, an antenatal diagnosis was made of alternating bradycardia, tachycardia and Supraventricular Tachycardia (SVT), which lead to an increase in monitoring during pregnancy. A minor pericardial effusion was noted at one point, thought secondary to episodes of SVT in utero. The antenatal clinical suspicion was of Long QT syndrome and parents had preliminary investigation with ECG’s, which were normal. A female infant was born at 38/40 weeks gestation via an emergency caesarean section, for foetal heart rate of 160bpm. Birth weight was 3.69kg (91st centile). Apgar scores were five at one minute and nine at five minutes. No obvious dysmorphic features were reported and specifically there were no skin lesions. Initial chest x-ray showed cardiomegaly.

Figure 3-3 Pedigree of family used in 'trio' approach to whole exome sequencing



The infant was transferred to our institution for further investigation and management at 12 hours of life. She developed episodes of SVT requiring adenosine without haemodynamic compromise, treated with propranolol and flecainide. Initial echocardiogram was thought to be minimally abnormal with some right atrial hypertrophy, a prominent left atrial appendage and some pericardial effusion. Her left ventricle was a little hypertrophied and quite hyper-trabeculated with some systolic impairment. It was expected that the morphological changes might regress once the arrhythmias were adequately treated. Feeding difficulties were evident at an early stage and nasogastric feeding was required from six weeks. At five months of age she underwent a Nissen fundoplication and Percutaneous Endoscopic Gastrostomy (PEG) insertion for pharyngeal stage dysphagia. Formal nerve conduction studies at that time consistent with congenital mild to moderate bulbar palsy. Subsequently she required a PEG-Jejunostomy at 13 months of age.

Further echocardiogram at six months of age showed a hypertrophied and hyper-trabeculated left ventricle with preserved systolic function and features of non-compaction. Until this age her main issues related to the feeding difficulties and recurrent chest infections due to

aspiration. At eight months of age she was admitted with seizures and was treated with antibiotics and antivirals. She developed broad complex tachycardia associated with cardiovascular collapse, requiring inotropic support and Cardiopulmonary Resuscitation (CPR) and was admitted to the Paediatric Intensive Care Unit (PICU) for cooling and inotropic support. In PICU she had further runs of Non Sustained Ventricular Tachycardia (NSVT) despite treatment with propranolol, amiodarone and magnesium. Torsade *de* Pointes cardiac arrest occurred requiring chest compressions and three external cardiac defibrillation shocks to cardiovert. Emergency implantation of an implantable cardiac defibrillator and permanent pacemaker with right atrial and ventricular pacing leads was undertaken and she was listed for cardiac transplantation. The intractable arrhythmias continued despite increasing medication and therefore she proceeded to undergo thoracic sympathectomy in order to interrupt the major source of norepinephrine release to the heart and therefore decrease the propensity for arrhythmia. She remained in hospital throughout this period and formal ophthalmology review was undertaken at eleven months of age, though this will have been limited by her clinical state. A divergent squint was noted and fundi were reported as normal. Extensive genetic testing was undertaken during this period, see Table 3-2.

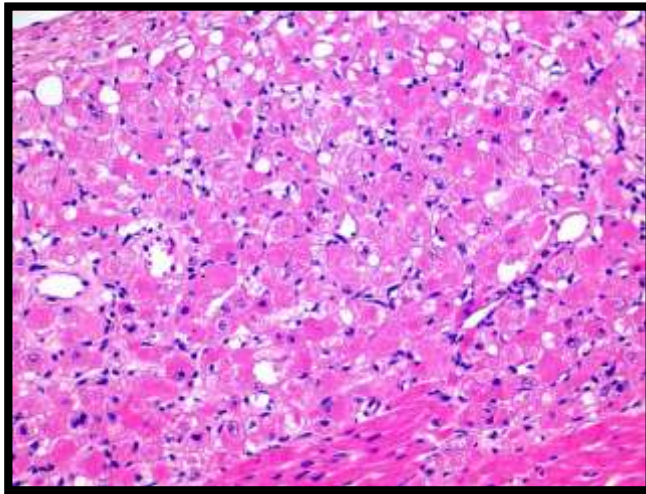
Table 3-2 Initial genetic investigations undertaken in the proband

Test	Gene (s)	Phenotype	Result
Array -CGH analysis*	Array Karyotype	? Copy number variation	No imbalance detected.
Direct sequencing	LMNA	DCM, arrhythmia	No mutation detected
Direct sequencing	LDB3	Myofibrillar Myopathy	No mutation detected
Direct sequencing	CAV3	Limb girdle muscular dystrophy type 1C	No mutation detected
Direct sequencing	FHL1	Four-and –a half-LIM domain 1 gene	No mutation detected
Direct sequencing	BAG3, DES	Cardiomyopathy-Rippling Muscle disease	No mutation detected
Direct sequencing	RYR2, CASQ2	CPVT	No mutation detected
Cardiomyopathy NGS panel	MYBPC3	HCM	No mutation detected
	MYH7	HCM	No mutation detected
	TNNT2	HCM	No mutation detected
	TNNI3	HCM	No mutation detected
	MYL2	HCM	No mutation detected
	MYL3	HCM	No mutation detected
	TPM1	HCM	No mutation detected
	ACTC1	HCM	No mutation detected
	CSRP3	HCM	No mutation detected
	PLN	HCM	No mutation detected
	FHL1	HCM	No mutation detected
	PRKAG2	HCM	No mutation detected
	LAMP2	HCM	No mutation detected
	GLA	HCM	No mutation detected
In-house research NGS panel of ~160 ICC genes			No likely pathogenic variants detected **

Initial genetic investigations undertaken in the proband (subject 1: affected child with Histiocytoid Cardiomyopathy (CM)), which failed to identify the molecular basis of disease. * Using an oligonucleotide array with 60,000 probes across the genome. ** Variant of Uncertain Significance in PLEC nsSNP c.10186G>T, p.V3396L seen at frequency of 0.00036 in ESPe) and in one individual recruited to the Cardiovascular Biomedical Research Unit at the Royal Brompton & Harefield NHS Trust, with a family history of cardiac disease but where cardiac screening found no abnormality.

By 13 months of age, echocardiogram showed moderate Left Ventricular Hypertrophy (LVH), mainly of the left ventricular free wall. At 14 months of age she underwent orthotopic cardiac transplantation. The indication was intractable arrhythmia. Examination of the native heart, was consistent with Histiocytoid CM, see Figure 3-4. Post-transplant she required Extracorporeal Membrane Oxygenation (ECMO) support with delayed chest closure. There had been post-operative acute ischaemic injury to the right leg following acute femoral artery thrombosis requiring fasciotomy and skin grafting.

Figure 3-4 Histology of native heart from proband (Subject 1) with Histiocytoid CM



Longitudinal section of myocardium from infant with Histiocytoid Cardiomyopathy. A few normal myocytes are present in the lower part of the field. The remainder of the cells are histiocytoid myocytes
Image reproduced with permission from Rea *et al* (216)

At 17 months of age (three months post transplant) echocardiography showed good biventricular systolic function and endomyocardial biopsy showed no rejection. From a neurodevelopment point she appeared to have some developmental delay, although this was difficult to fully assess in view of multiple co-morbidities. She sat independently at ten months; by 18 months of age, she was not walking and had minimal speech, but good comprehension.

At 18 months of age she developed acute allograft reaction (a combination of acute and accelerated chronic vascular rejection) and died. Post mortem examination of the brainstem did not demonstrate any abnormality despite the history of bulbar palsy. There was evidence of residual histiocytoid change in the small amount of native heart left after transplant and there was focal histiocytoid change in the thyroid, lungs and choroid plexus of the brain. Family screening involved clinical screening of both parents with CMR and ECG, which were within normal limits. The elder sister (see Figure 3-3) had a normal echo and ECG at age four years.

3.2.3 Whole Exome Sequencing using a ‘Trio’ approach

WES was undertaken after the death of the subject 1 (once the diagnosis of Histiocytoid CM was made on pathology). This was performed on 3000ng of genomic DNA from the ‘trio’

consisting of affected child and both unaffected parents (paternity as reported), looking for genes with rare homozygous or compound heterozygous, or *de novo* variants. Libraries were constructed for all samples and prepared according to the manufacture's protocols (Agilent SureSelect Human All Exon v4 +UTR kit (71Mb)) and (100-bp) paired-end sequence reads were generated on the Illumina HiSeq 2500 platform. Alignment of sequence reads to human reference genome (hg19) was done using BWA v0.7.5, and variants were called using the GATK v2.8-1 software package. Only variants which passed standard GATK filters, had a genotype quality score of >20, had depth of coverage of >20x and an allelic balance of >20 were included in the analysis. Summary figures for coverage in each sample in the 'trio' are shown in Table 3-3.

Table 3-3 Summary figures of coverage for WES in each sample of the family 'trio'

Sample	Total reads	Bases x10*	Bases x20*	Bases x30*	% Callable
10CM02571 Affected child	119067512	97.8	93.4	87.6	99.4
10DM02524 Unaffected father	91775652	91.9	80.5	70.0	97.8
10NM02525 Unaffected mother	103827148	98.7	96.5	90.2	99.1

The 'trio' consists of child affected with Histiocytoid Cardiomyopathy and both unaffected parents. Figures are for the coverage of the protein-coding target. * refers to the proportion of the target with a depth of coverage of at least 10/20/30 reads. The % callable indicates the percentage of the target sequenced with sufficient depth and quality to call variants. Sequencing was undertaken using SureSelect Human All Exon V4+UTR design (71Mb) and the Illumina HiSeq 2500 platform. WES; Whole Exome Sequencing.

3.2.4 Whole Exome Sequencing using an 'overlap' approach

WES was performed on 50ng of genomic DNA from five affected unrelated individuals, including the affected individual used in the trio approach. In this approach I was looking for genes with rare protein altering variants in more than one individual. Libraries were constructed for all samples and prepared according to the manufacture's protocols (Nextera Rapid Capture Exome (37Mb)) and (150-bp) paired-end sequence reads were generated on the Illumina NextSeq 500 platform. Bioinformatic analysis and variant prioritization were as outlined previously. Summary figures of coverage for each sample are shown in Table 3-4.

Table 3-4 Summary figures of coverage for WES using an ‘overlap’ approach

Sample	Total reads	Bases x10*	Bases x20*	Bases x30*	% Callable
10CM02571*	240236446	98.9	98.0	96.8	98.8
20EA03620	235792968	99.0	98.2	97.3	98.9
20EM03618	199892386	98.8	97.9	96.8	98.7
20PO03585	251739384	98.9	98.1	97.1	98.8
20TH03619	183239678	98.5	97.1	95.4	98.4

WES; Whole Exome Sequencing. The overlap approach involved five unrelated subjects with Histiocytoid Cardiomyopathy. * refers to the proportion of the target with a depth of coverage of at least 10/20/30 reads. The % callable indicates the percentage of the target sequenced with sufficient depth and quality to call variants. Sequencing was undertaken using Nextera 37Mb exomes (Protein coding Target region=37Mb: +/- 2 bp of coding exons) and the NextSeq 500 platform.

Nextera Rapid Capture Exome (37Mb) used in the ‘overlap’ approach, contains 11

mitochondrial DNA regions (which were not covered by the SureSelect Human All Exon

V4+UTR design (71Mb) kit used in the ‘trio’ approach). These were annotated and filtered

as outlined previously. The percentage callable for each mtDNA region in each sample is

shown in Table 3-5.

Table 3-5 Percentage callability for mtDNA regions included in WES

Gene ID	Start	End	Sample ID				
			10CM02571*	20EA03620	20EM03618	20PO03585	20TH03619
MT-ND1	3306	4262	100	100	100	100	100
MT-ND2	4469	5511	100	100	91.4	100	100
MT-CO1	5903	7445	100	100	99.9	100	100
MT-CO2	7585	8266	100	100	100	100	100
MT-ATP6	8365	9204	100	100	100	100	100
MT-CO3	9206	9990	100	100	100	100	100
MT-ND3	10058	10404	100	100	100	100	100
MT-ND4	10469	12137	100	100	100	100	100
MT-ND5	12336	14145	100	100	100	100	100
MT-ND6	14148	14673	100	100	100	100	100
MT-CYB	14746	15887	100	100	100	100	100

Whole Exome Sequencing (WES) used the Nextera Rapid Capture Exome (37Mb) in each of the five unrelated cases of Histiocytoid CM used in the ‘overlap’ approach. Sample 10CM02571* is the affected child from the family trio and was discarded from final overlap analysis as causative gene had been identified in the time lapse from WES to analysis.

3.2.5 Mapping mitochondrial DNA (mtDNA) variants

WES data generated in the overlap approach of multiple unaffected individuals included

coverage of eleven mtDNA regions (see Table 3-5). The population frequency of mtDNA

variants is not currently available through ExAC, but is available via MITOMAP

(<http://www.mitomap.org/bin/view.pl/Main/SearchAllele>) a human mitochondrial genome database containing a GenBank derived from 30589 sequences with greater size than 15.4kbp. These sequences may not all be of equal quality and some are from individuals with disease.

Our WES data was mapped to HG19, which is the UCSC variant of the GRCh37 genome assembly. It includes the NC_001807 mitochondria sequence rather than the Revised Cambridge Reference Sequence (rCRS), which is the mitochondrial reference sequence, used by MITOMAP. The rCRS is included in the GRCh38 assembly and hg38 Genome Browser (ChrM) (RefSeq accession number NC_012920). In order to identify mtDNA variants of interest, variants were prioritized as per methods outlined below within our data and then the LiftOver tool, (accessed via <https://genome.ucsc.edu/cgi-bin/hgLiftOver> on 8th November 2015) was used to for conversion of each variant of potential interest from HG19 to hg38. MITOMAP was then searched for each variant to identify the population frequency.

3.2.6 Variant prioritization

In each WES approach, xBrowse software (<https://atgu.mgh.harvard.edu/xbrowse>) was used to prioritize variants of interest by filtering for protein altering variants occurring at varying frequencies in each cohort according to various plausible inheritance patterns.

3.2.7 Sanger Sequencing of candidate genes

For analysis of *NDUFB11* (the strongest candidate gene), primers were designed to amplify the coding exons and intron-exon boundaries of *NDUFB11* by PCR, so that additional samples could be examined for any evidence of sequence variation in *NDUFB1*.

Amplification was performed using *Taq* DNA polymerase. PCR products were sequenced by capillary sequencing; using the BigDye Terminator Cycle Sequencing Kit and an ABI3500 Genetic Analyser (Applied Biosystems). Sequences were analysed using Sequencher (v5.3). The coding sequence of all samples was fully covered.

For analysis of the *de novo* nonsense *FAM135* mutation detected in the affected child by WES (see results), primers were designed to amplify the only the coding region surrounding the variant c.474C>G (p. Tyr158Ter) and Sanger sequencing carried out in the child and both parents as outlined above. For full details of primers see appendix Table 4-5.

3.2.8 Identifying variants in candidate genes in publically available WES data

During the course of this study Shehata *et al* (204) published data reporting *NDUFB11* mutations in two cases of Histiocytoid CM in which they had undertaken WES; their data from three further cases is publically available and was searched for any potential candidate variants we identified.

3.3 Results

3.3.1 Whole Exome Sequencing using a ‘Trio’ approach

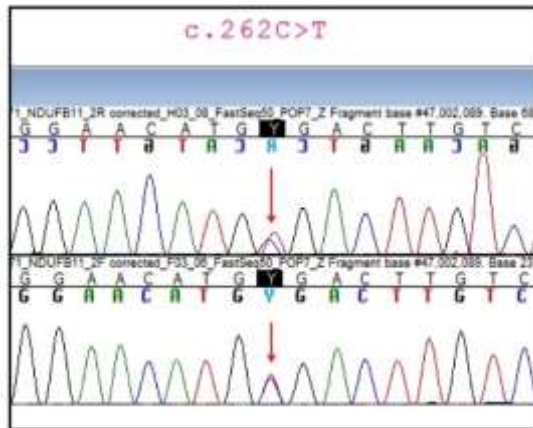
3.3.1.1 Considering a *de novo* dominant model of inheritance

We applied a *de novo* dominant model of inheritance, filtering for protein altering variants (nonsense, frameshift, essential splice site, missense and in-frame) which are present in the heterozygous state in the affected child, and where the parents are homozygous for the reference sequence. This identified 5 variants (see Table 3-6). Each of these variants was manually inspected using Integrative Genome Viewer (IGV) in both the child and both parents. Two variants were consistent either with sequencing errors or were present in a parent and therefore not *de novo* (see appendix Table 4-11).

The remaining three variants appeared genuine *de novo* variants. We then filtered for those present at a frequency of less than 0.001 in ALL of the Exome Aggregation Consortium (ExAC v0.3), 1000 Genomes and ATGU in-house controls. The frequency filter was set at this level to reflect the rarity of the disease and the severe, early onset presentation. The

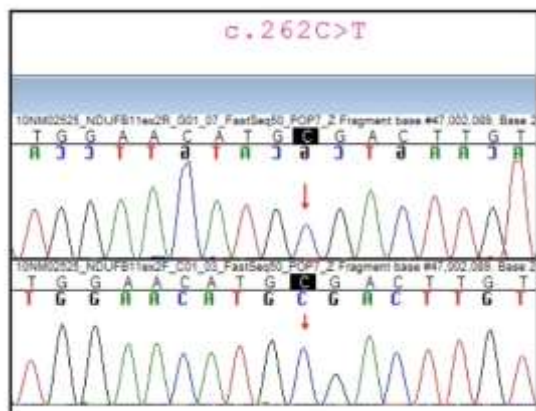
number of candidate variants in the affected child was reduced to two (See Table 3-7). These variants were prioritised as candidate genes and taken forward for validation by Sanger sequencing (see Figure 3-5, Figure 3-6, Figure 3-7, Figure 3-8, Figure 3-9 and Figure 3-10).

Figure 3-5 Sequence electropherogram of affected child showing the *NDUFB11* mutation



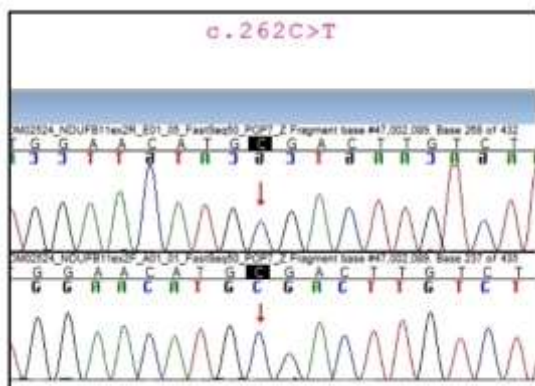
Sequence electropherogram from genomic DNA of child affected by Histiocytoid CM (subject 1) from the family ‘trio’ showing the *NDUFB11* mutation identified by WES and confirmed by Sanger sequencing. The image shows part of *NDUFB11* exon 2. Forward and reverse reads shown. The red arrow points to the double peak in the electropherogram showing heterozygosity for the nonsense mutation in *NDUFB11* c.262C>T. The coverage at this base was 62X, with 27 reference reads, 33 alternate reads. CM; Cardiomyopathy; WES; Whole Exome Sequencing.

Figure 3-6 Sequence electropherogram from unaffected mother showing no evidence of *NDUFB11* mutation



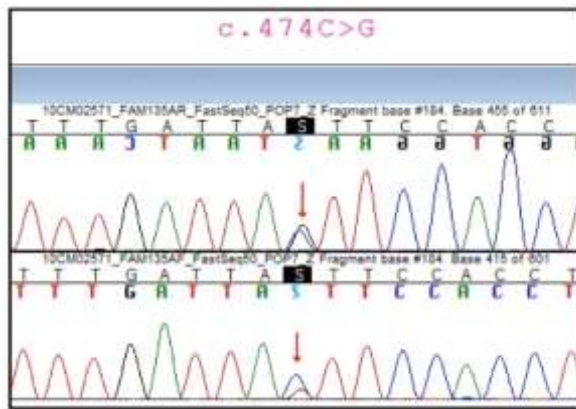
Sequence electropherogram from genomic DNA of the unaffected mother of subject 1 from the 'trio' of WES showing no evidence of *NDUFB11* sequence-level mutation which had been identified in the affected child. The image shows part of *NDUFB11* exon 2. Forward and reverse reads shown. The red arrow points to the base where the nonsense mutation in *NDUFB11* c.262C>T was detected in the affected daughter. The coverage at this base on WES was 50X, with no evidence of the mutation. WES; Whole Exome Sequencing.

Figure 3-7 Sequence electropherogram from the unaffected father showing no evidence of *NDUFB11* mutation



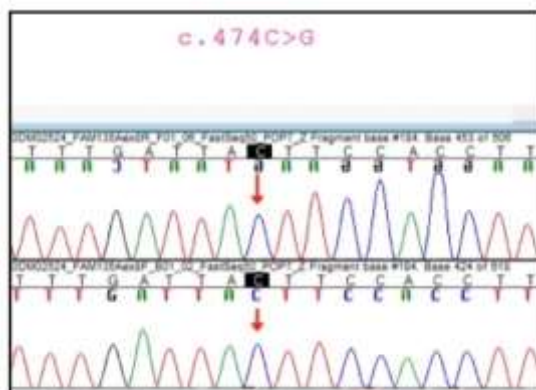
Sequence electropherogram from genomic DNA of the unaffected father of subject 1 from the 'trio' of WES showing no evidence of *NDUFB11* mutation which was identified in the affected child. The image shows part of *NDUFB11* exon 2. Forward and reverse reads shown. The red arrow points to the base where the nonsense mutation in *NDUFB11* c.262C>T was detected in the affected daughter. The coverage at this base on WES was 26X, with no evidence of the mutation. WES; Whole Exome Sequencing.

Figure 3-8 Sequence electropherogram from the affected child showing the *FAM135A* mutation



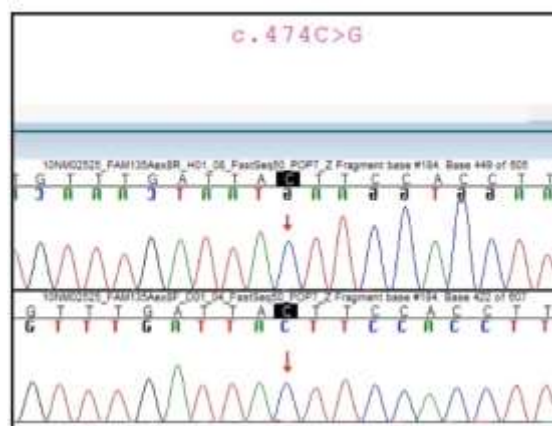
Sequence electropherogram from genomic DNA from the affected child (subject 1) from the family ‘trio’, showing the *FAM135A* sequence-level mutation identified by whole exome sequencing and confirmed by Sanger validation. The image shows part of *FAM135A* exon 8. Forward and reverse reads shown. The red arrow points to the double peak in the electropherogram showing heterozygosity for the nonsense mutation in *FAM135A* c.474C>G; p.Tyr158Ter. The coverage at this base was 78X, with 32 reference reads, 46 alternate reads.

Figure 3-9 Sequence electropherogram from unaffected mother showing no evidence of *FAM135A* mutation



Sequence electropherogram from genomic DNA of the unaffected mother of subject 1 from the ‘trio’ of WES showing no evidence of *FAM135A* sequence-level mutation. The image shows part of *FAM135A* exon 8. Forward and reverse reads shown. The red arrow points to the base where the nonsense mutation in *FAM135A* c.474C>G; p.Tyr158Ter was detected in the affected child. The coverage at this base on WES was 75X, with no evidence of the mutation.

Figure 3-10 Sequence electropherogram from unaffected father showing no evidence of *FAM135A* mutation



Sequence electropherogram from genomic DNA from the unaffected father child subject 1 from the ‘trio’ of WES showing no evidence of *FAM135A* sequence-level mutation. The image shows part of *FAM135A* exon 8. Forward and reverse reads shown. The red arrow points to the base in the electropherogram where the nonsense mutation in *FAM135A* c.474C>G; p.Tyr158Ter was detected in the affected child (subject 1). The coverage at this base on WES was 57X, with no evidence of the mutation.

We noted that of the two protein-altering variants consistent with a *de novo* dominant (monoallelic) model of inheritance (see Table 3-7) the nonsense mutation in *NDUFB11* located on the X chromosome provided a potential explanation for the excess of cases in females (with presumed early male lethality) and fits with the hypothesized role in mitochondrial function, while the second gene (*FAM135A*) is not expressed in the heart.

Table 3-6 Putative *de novo* protein altering variants identified in the affected child

Genes	Chromosome	Genomic Position	Reference	Alternate	Category of variation	ATGU in-House Controls	1000 Genomes	ExAC	Confirmed <i>de novo</i> variant
CDK11A	1	1635061	C	A	missense	0	0.002	0.0004	No
HLA-DQA1	6	32609312	A	C	missense	0.4	0.46	0.10	No
FAM135A	6	71186967	C	G	nonsense	0	0	0	Yes
CD163L1	12	7521563	T	TGA	frameshift	0	0	0.01	No
NDUFB11	X	47002089	G	A	nonsense	0	0	0	Yes

All *de novo* protein altering variants identified in the affected child and absent in the unaffected parents. Protein altering variants (nonsense, frameshift, essential splice site, missense and in-frame). These variants appeared consistent with a *de novo* dominant (monoallelic) model of inheritance, but on manual inspection, two were either present in a parent or consistent with a sequencing error (see text for details).

Table 3-7 Candidate genes for Histiocytoid Cardiomyopathy

HGNC symbol	<i>NDUFB11</i>	<i>FAM135A</i>
Name	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 11	Family with sequence similarity 135, member A
Functional class	stop_gained	stop_gained
Transcript	ENST00000276062.8: c.262C>T	ENST00000457062.2: c.474C>G
Protein	ENSP00000276062.8: p.Arg88Ter	ENSP00000409201.2: p.Tyr115Ter
Genomic location	Chromosome X: 47002089	Chromosome 6: 71186967
Highly expressed in the heart	Yes	No

Candidate genes identified through WES using a ‘trio’ approach. These variants are consistent with a *de novo* dominant (monoallelic) model of inheritance, present at a frequency of less than 0.001 in ALL of the Exome Aggregation Consortium (ExAC), 1000 Genomes and ATGU in-house controls and there was no evidence on manual inspection that these variants were either present in a parent, or consistent with a sequencing error (see text for details). Data on gene expression in the heart from <http://www.gtexportal.org/> accessed on 28th July 2016).

At this point *NDUFB11* was felt to be a very strong candidate gene for Histiocytoid CM.

Sanger Sequencing was undertaken of the entire coding sequence of *NDUFB11* in the four additional, unrelated singleton affected cases with a diagnosis of Histiocytoid CM and did not identify any further mutations in *NDUFB11*.

3.3.2 Validation of *NDUFB11* as a new disease gene

In addition to attempting to replicate our findings of *NDUFB11* as the causative gene in additional cases of Histiocytoid CM, I further explored the role of *NDUFB11* in other phenotypes.

The Deciphering Developmental Disorders (DDD) study is large multicenter project, combining recruitment of patients from all 24 UK and Ireland Regional Genetics services, with the research and bioinformatics expertise of the Wellcome Trust Sanger Institute.

Recruitment of patients began in April 2011 and was primarily focused on children with

12,000 individuals with severe and extreme developmental phenotypes. The overall aim is to facilitate translation of genomic sequencing technologies into the NHS (Firth et al., 2011, Wright et al., 2015). Data from this study is available via the DECIPHER database (Databases of genomic variation and Phenotype in Humans using Ensembl Resources). Hypothesizing that variation in *NDUFB11* (+/- HCCS, +/- COX7B: see discussion) would be identifiable in children with related developmental phenotypes such as LVNC cardiomyopathy or eye malformations (see discussion), the database was accessed on 11th September 2015 (<https://decipher.sanger.ac.uk/>) and searched for variants in any of these genes. At that point data from 4295 trios was available. No sequence variants were identified in any of these three genes.

Hypothesizing that mutations in *NDUFB11* could be a cause of Sudden Infant Death Syndrome, we interrogated sequence data from *NDUFB11* on 427 cases of Sudden Infant Death Syndrome and identified a single novel *NDUFB11* missense variant was identified in a hemizygous male. This variant (*NDUFB11*: NM_001135998: exon2: c. G283A: p. Val95Ile) is found with an allele frequency of 0.00006848 in ExAC, including 3 hemizygous males and 3 heterozygous females, with no homozygous females. This variant is predicted to be benign by both PolyPhen and SIFT.

An identical histological condition is recognised in animals, including Savannah kittens (217), when it is generally known as congenital Purkinje fiber dysplasia and where it has been seen in association with LVNC, as it has been in children (218). Slides from archive material were made available but unfortunately we were unable to extract adequate good quality DNA to sequence *NDUFB11*.

3.3.3 Whole Exome Sequencing using an ‘overlap’ approach

Whilst trying to identify additional cases of Histiocytoid CM in which to search for further variants in our candidate gene *NDUFB11* further WES was undertaken using the ‘overlap’

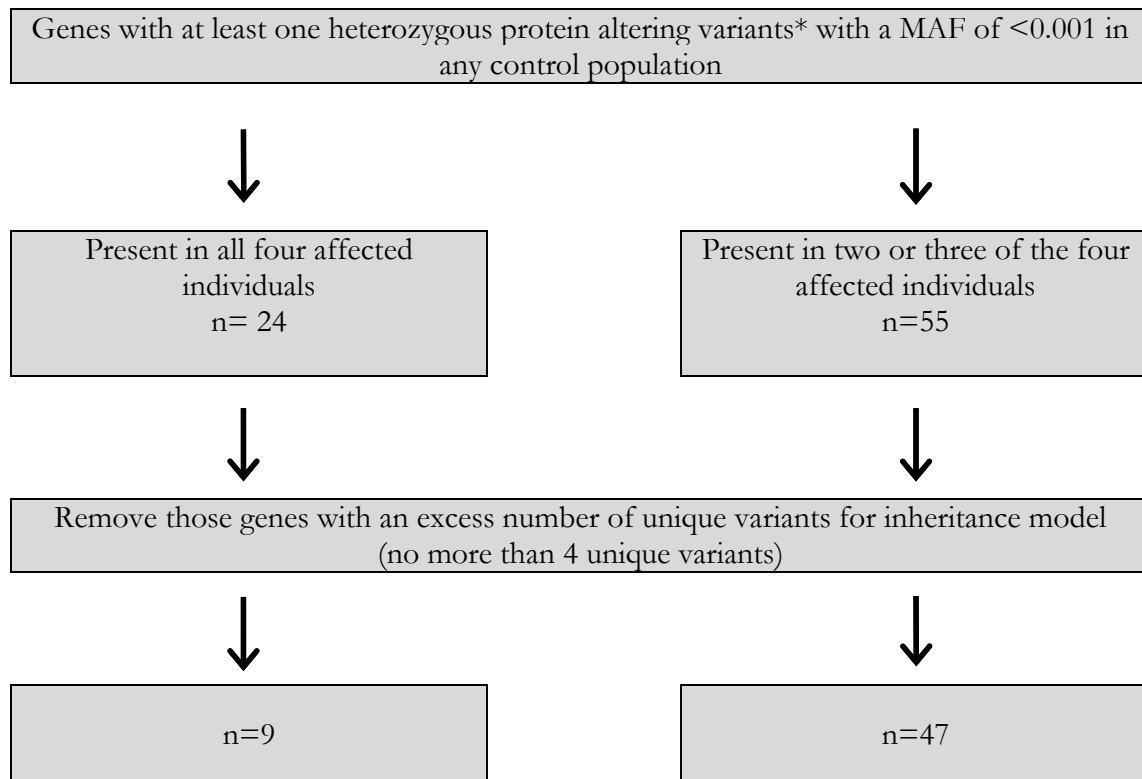
approach of five unrelated affected singleton cases in order to identify further candidate genes (this included subject 1 from the ‘trio’ approach). Simultaneous to this *NDUFB11* was published as the causative gene in Histiocytoid CM (204) and Microphthalmia and Linear Skin defects syndrome (MLS syndrome) (219). I therefore excluded subject 1(10CM02571: the affected child from the family trio, with the causal nonsense mutation in *NDUFB11* (c.262C>T) from further analysis. This left four unrelated cases for analysis using an overlap approach. As outlined earlier all variants passed standard GATK filters, had a genotype score of >20 and allelic balance of >20. There was no filtering on level of coverage at this stage, as this option was unavailable within the current version of xBrowse used to analyze an overlap cohort. In the WES analysis using the ‘trio’ the aim was to detect potentially pathogenic variants. In the overlap analysis the focus is directed at identifying genes, which have rare damaging variants in multiple individuals. The analysis was divided into identification of candidate genes from either nuclear encoded genes or mtDNA genes.

3.3.3.1 Identifying candidate nuclear encoded genes

While I hypothesized that *de novo* mutations would be the most likely cause of disease, in the absence of parental data on the four overlap cases I was unable to filter on this basis and so we started with a dominant model (of which *de novo* variants are a subset), accepting that the study would be underpowered for a genetically heterogeneous disorder. I initially filtered for genes (see Figure 3-11) which had one or more protein altering variants (nonsense, frameshift, essential splice site, missense and in-frame) in the heterozygous state in all four of the four affected individuals (both unique and different variants in the same gene were considered, which were present at a frequency of <0.001 in any control populations (ExAC (or any ExAC sub-population), 1000 Genomes (or any 1000 genomes sub-population) and xBrowse in-house reference samples (n=375 exomes)). It was noted that many of these genes had a very large number of unique variants, which we considered implausible and likely due

to sequencing pipeline artifacts. Genes with an excess number of unique variants for a monoallelic inheritance model were then excluded (no more than four unique variants per gene).

Figure 3-11 Filtering strategy to identify candidate genes for Histiocytoid CM consistent with a mono-allelic cause of disease

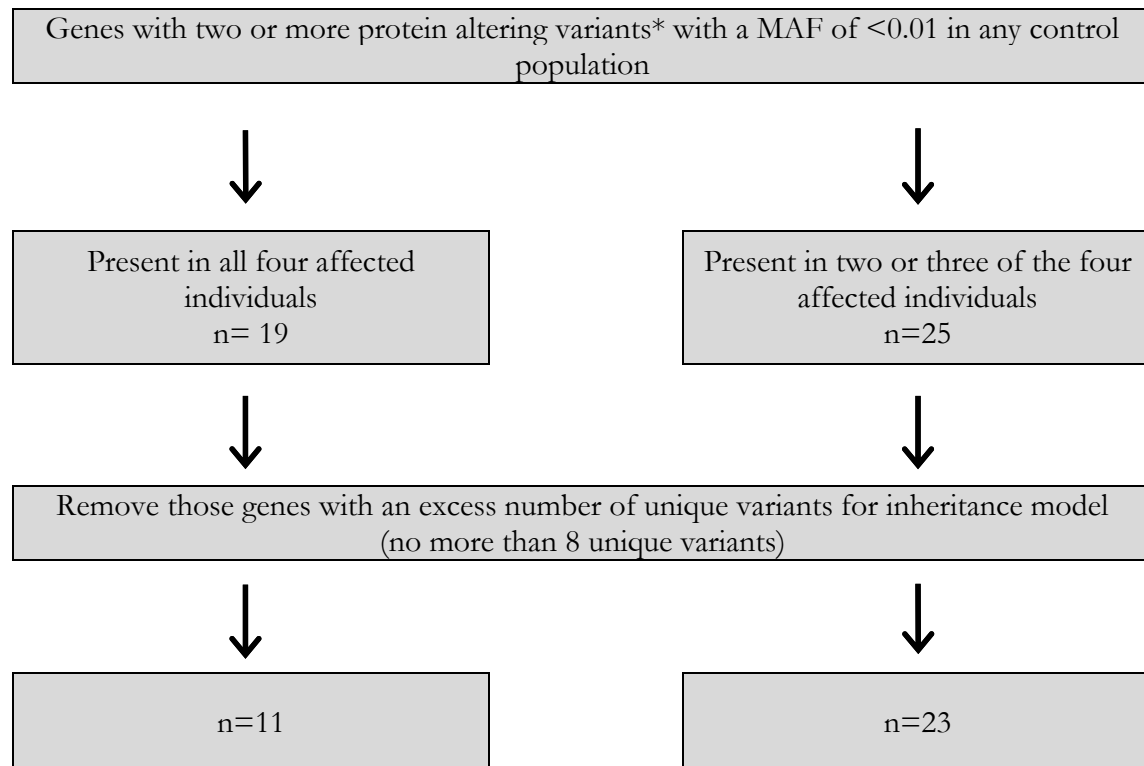


The first filtering strategy to identify candidate genes for Histiocytoid CM using WES of an ‘overlap’ approach involved a cohort of four unrelated affected individuals with Histiocytoid CM. Candidate genes must have at least one heterozygous variant in all four unrelated cases and therefore potentially be consistent with a mono-allelic cause of disease. For comparison, also shown is the number of additional genes which would have been considered at each stage of filtering if we had set more relaxed criteria of requiring only two or three of the cases in the cohort to have hits in each gene.

*Nonsense, Frameshift, essential splice site, missense, and inframe variants. Includes all SNPs and InDels. Controls populations filter: Minor Allele Frequency (MAF) of < 0.001 in the global ExAC population, any of the six ExAC sub-populations, in the global 1000 genomes population or any the five 1000 genomes sub-populations. CM; Cardiomyopathy. WES; Whole Exome Sequencing.

This process was then repeated, using a more relaxed population frequency filter of a MAF of <0.01 in any control population, but requiring at least two variants in each candidate gene in each affected individual, in order to identify genes with a potential bi-allelic pattern of inheritance (see Figure 3-12). Genes with an excess number of unique variants (no more than 8 allowed) were then discarded. As expected, some genes were identified by both methods of filtering; duplicate genes were counted only once, leaving a final list of 14 genes. Each gene was then examined further, with consideration of which genes had exactly one rare variant per person and therefore would be truly consistent with mono-allelic inheritance, which genes contained only two variants per person and therefore could be consistent with bi-allelic inheritance, and finally which genes contained multiple rare variants per individual and therefore were suspicious of sequencing errors. The genes identified through these filtering strategies are shown in Table 3-8.

Figure 3-12 Filtering strategy to identify candidate genes for Histiocytoid CM consistent with a bi-allelic cause of disease



The second filtering strategy to identify candidate genes for Histiocytoid CM using WES with an ‘overlap’ approach involving a cohort of four unrelated affected individuals with Histiocytoid CM.

Candidate genes must have at least two variants in all four unrelated cases and therefore potentially be consistent with a bi-allelic cause of disease. A higher threshold of MAF was used to filter for variants considered in the bi-allelic model to reflect a conservative background carrier rate in the population. For comparison, also shown is the number of additional genes which would have been considered at each stage of filtering if we had set more relaxed criteria of requiring only two or three of the cases in the cohort to have hits in each gene. *Nonsense, Frameshift, essential splice site, missense, and inframe variants. Includes all SNPs and InDels. The 1000 genomes filters are based on Phase 3 release (5/2/13). The ExAC filters are based on frequency of variants amongst all individuals in ExAC. CM; Cardiomyopathy. WES; Whole Exome Sequencing.

Table 3-8 Genes (n=15) with rare protein altering variants identified in all four unrelated individuals affected by Histiocytoid CM

Gene	Chromosome	Number of Unique Variants	No. of PTV's	No. of non-TV's	N subjects with >1 variant allele	Only one rare Variant per Person	Only two rare Variants per Person	Total number of rare variants in the cohort	RVIS score*	HI score**	Probability of being loss of Function Intolerant (pLI)***	Missense Z Score rank (%)****
<i>ANKRD20A4</i>	9	1	0	1	0	Yes	No	4	NA	89.8	0.32	95.7
<i>ANKRD36</i>	2	4	4	0	4	No	No	15	NA	94.5	0.0	47.5
<i>CTDSP2</i>	12	7	0	7	4	No	No	28	-0.23 (36.9%)	13.6	0.40	75.9
<i>FRG1</i>	4	2	0	2	4	No	Yes	8	-0.25 (35.4%)	51.5	0.0	50.1
<i>FRG1B</i>	20	5	1	4	4	No	No	18	NA	81.3	0.0	70.6
<i>GGT1</i>	22	8	1	7	4	No	No	29	-0.09 (46.9%)	64.4	0.68	94.8
<i>IGHJ6</i>	14	2	2	0	4	No	Yes	12	NA	NA	NA	NA
<i>KMT2C</i>	7	2	0	2	4	No	No	12	-2.52 (0.9%)	53.6	1	71.5
<i>PCLO</i>	7	3	0	3	0	No	No	6	-0.18 (40.6%)	48.8	1	30.7
<i>POTED</i>	21	7	0	7	4	No	No	13	NA	94.9	0.74	94.6
<i>PRSS3</i>	9	6	2	4	4	No	No	22	-0.14 (43.3%)	81.4	0.0	46.3
<i>SETD8</i>	12	5	0	2	4	No	No	20	0.1 (61.5%)	32.0	0.95	84.6
<i>TEKT4</i>	2	4	1	3	4	No	No	15	1.27 (93.6%)	88.0	0.0	4.5
<i>TTC28</i>	22	4	1	3	4	No	No	10	0.89 (89.3%)	18.7	NA	NA

Fifteen genes in which rare protein altering variants were identified in all four of the unrelated individuals affected by Histiocytoid CM. For details of how these genes were filtered see text. Genes were identified as potentially consistent with either mono-allelic inheritance (variants present at a MAF of < 0.001 in any control population) or bi-allelic model (variants present at a MAF of < 0.01 in any control population) or both. Where genes were potentially consistent with both mono-allelic or bi-allelic inheritance then the higher number of unique variants from the bi-allelic model (because of the lower MAF used) is noted here.* Control population (ExAC & 1000 genomes, Global population frequency and any sub-population). Protein Truncating Variants (PTVs) include nonsense, frameshift and essential splice site variants. Non-Truncating variants (non-TVs) refer to missense and in-frame variants

* Accessed via <http://genie-intolerance.org/index.jsp> (22nd October 2015) ** Accessed via <https://decipher.sanger.ac.uk/browser> (22nd October 2015). *** Accessed via <http://exac.broadinstitute.org/> (18th November 2015). **** Missense Z scores for all genes were ranked and converted to percentiles- the higher the percentile the more constrained (intolerant of variation). CM; Cardiomyopathy. MAF; Minor Allele Frequency. NA; Not Available.

From further evaluation of each of the genes (see Table 3-8) there was only one gene (*ANKRD20A4*) with a single rare variant per person in the cohort of four cases of Histiocytoid CM and therefore truly consistent with mono-allelic inheritance and two genes (*FRG1*, *IGHJ6*) with exactly two rare variants per person in the cohort and therefore consistent with bi-allelic inheritance. Details of the specific variants identified in each of these genes are outlined in Table 3-9, Table 3-10, Table 3-11. Each of the variants was manually reviewed by visual inspection on IGV and they were all consistent with sequencing or alignment errors.

Table 3-9 Summary of variants identified in ANKRD20A4

ANKRD20A4																
Variant	CaCategory	In-silico predictions				Population frequency				RS ID	Sample ID				IGV	Comment
		Polyphen	SIFT	Mut Taster	FATHMM	1kg	1kg_popmax	ExAC	ExAC_popmax		20PO03585	20EA03620	20EM03618	20TH03619		
c.2088A>T p.Gln696His	missense	-	+	-	-	0	0	0	0	rs75696372# rs78465713#	*	*	*	*	Reads mapped to multiple other regions on the genome. Eight additional rare variants were present in multiple samples which had been other removed by earlier quality filters. Likely sequencing or alignment error.	ANKRD20A4 has no known disease associations. The rarity and severity of Histiocytoid CM mean it is unlikely that a single variant, which was identified in all four cases and which has been identified multiple times in dbSNP, without any clear disease association is the cause.

Summary of any variants identified in *ANKRD20A4* (chromosome 9: ankyrin repeat domain 20 family member A4), which was the only gene identified by WES using an overlap approach in 4 individuals with Histiocytoid CM, which was truly consistent with a mono-allelic inheritance model. For In silico predictions + refers to a prediction of probably damaging by Polyphen, damaging by SIFT or FATHMM or disease causing by Mutation Taster while – refers to benign by Polyphen, tolerated by SIFT or FATHMM. # Validated by multiple, independent submissions to the refSNP cluster. Population frequency (1kg; Global 1000 genomes population: 1kg_popmax; 1000 genomes any sub-population: ExAC; Global ExAC population frequency: ExAC_popmax;any ExAC sub-population). * Heterozygous for variant.

Table 3-10 Summary of variants identified in **FRG1**

		FRG1														
Variant	Category	In-silico Predictions				Population frequency				RS ID	Sample ID				IGV	Comment
		Polyphen	SIFT	Mut Taster	FATHMM	1kg	1kg_popmax	ExAC	ExAC_popmax		20PO03585	20EA03620	20EM03618	20TH03619	Reads mapped to multiple other regions on the genome. In addition to all four of these cases having two identical variants, the additional case of Histiocytoid CM (10CM02571- Subject 1 from the trio approach) also underwent WES using this pipeline and both of these variants were also identified in that sample. This suggests these variants are likely to be pipeline specific sequencing artifacts.	
c.568A>G p.Lys190Glu	Missense	-	-	+	-	0	0	0	0	rs184307882	*	*	*	*		Deletions of FRG1 are associated with FacioScapuloHumeral Muscular Dystrophy.
c.604G>A p.Val202Ile	Missense	-	-	+	-	0	0	0	0	rs6846627 #	*	*	*	*		

Summary of any variants identified in *FRG1* (chromosome 4: FSHD region gene), which was one of only two genes identified by WES using an overlap approach in 4 individuals with Histiocytoid CM, which was consistent with a bi-allelic inheritance model. For In silico predictions + refers to a prediction of probably damaging by Polyphen, damaging by SIFT or FATHMM or disease causing by Mutation Taster while – refers to benign by Polyphen, tolerated by SIFT or FATHMM. # Validated by multiple, independent submissions to the refSNP cluster. Population frequency (1kg; Global 1000 genomes population: 1kg_popmax; 1000 genomes any sub-population: ExAC; Global ExAC population frequency: ExAC_popmax; any ExAC sub-population). * Heterozygous for variant.

Table 3-11 Summary of variants identified in IGHJ6

IGHJ6																
Variant	Category	In silico Predictions				Population frequency				RS ID	Sample ID				IGV	Comment
		Polyphen	SIFT	Mut Taster	FATHMM	1kg	1kg_popmax	ExAC	ExAC_popmax		20PO03585	20EA03620	20EM03618	20TH03619		
c.18_19delNN p.Met7GlyfsTer?	Frameshift	N/A				0	0	0	0	N/A	**	**	*	*	Although called as two separate variants (chr14:106329449 ATG> A & chr14:106329453 A >ACC), these appear as consistent with a single sequencing error, called twice in all samples, either in the heterozygous or homozygous form. The same variants were also present (in heterozygous form) in sample 10CM02571, which was sequenced on the same pipeline, and therefore these are likely pipeline specific sequencing artifacts.	Immunoglobulin Heavy Joining 6 (IGHJ6) has no known clear disease associations.
c.16delNinsGGT p.Tyr6GlyfsTer?	Frameshift	N/A				0	0	0	0	N/A	**	**	*	*		

Summary of any variants identified in IGHJ6 (chromosome 14: Immunoglobulin Heavy Joining 6), which was one of only two genes identified by WES using an overlap approach in 4 individuals with Histiocytoid CM, which was consistent with a bi-allelic inheritance model. Population frequency (1kg; Global 1000 genomes population: 1kg_popmax; 1000 genomes any sub-population: ExAC; Global ExAC population frequency: ExAC_popmax; any ExAC sub-population). * Heterozygous for variant, ** Homozygous for variant. CM; Cardiomyopathy. NA; Not available. WES; Whole Exome Sequencing.

3.3.4 Applying a Mitochondrial DNA model of inheritance

From the background literature on Histiocytoid CM there was a high *a priori* chance that mtDNA variants could be implicated. We considered any mitochondrial region with hits in two or more of the cases from the cohort of four, all coding variants were included and there was no filtering for an excess number of unique variants. Population data on these genes is not included in ExAC and therefore MITOMAP was searched for each individual variant identified (<http://www.mitomap.org/MITOMAP>). The amino acid change and its effects (synonymous or non-synonymous change) were also noted from MITOMAP. Non-synonymous candidate variants identified in the mtDNA regions are outlined in (see appendix Table 4-12). MT-ND1 had two different non-synonymous variants identified in two cases (one variant in each case), of which one variant was present at a low population frequency and is considered a strong candidate gene (see Table 3-12).

Table 3-12 Summary of candidate variant identified in MT-ND1

Gene	Chromosome	Variant	Category	Population Frequency	Sample ID			
					20PO03585	20EA03620	20EM03618	20TH03619
ND1	M	m. 3308T>G	Nonsense	4 previous reports in MITOMAP	*			

MT-ND1 was the only candidate mitochondrial DNA gene identified by WES using an overlap approach in 4 individuals with Histiocytoid CM, considering a mitochondrial inheritance model * Review on IGV indicates the coverage at this base x623 with 98% of reads called as alternate (G) and 2% as reference (T) This variant is absent from all other samples sequenced on the same pipeline. Validation with Sanger sequencing is currently underway.

3.4 Discussion

3.4.1 Merits of different Whole Exome Sequencing approaches

Despite extensive genetic, metabolic and mitochondrial investigations during the routine clinical care of subject 1 (affected child in the family ‘trio’) the cause of Histiocytoid CM was unknown. In order to identify the underlying molecular basis we applied WES, which has proved a powerful tool for the identification of disease causing variants in rare,

Mendelian disease. In particular, WES allows the identification of newly arising '*de novo*' events (55), which we hypothesized, was a likely cause of this ultra-rare and typically sporadic disorder. In each generation, 70 to 175 *de novo* point mutations are expected of which up to three are anticipated to cause protein-coding changes (101, 102). Consistent with this, when analysing the data from the family 'trio'; two rare *de novo* protein coding variants were identified see Table 3-7. Subject 1 (10CM02571) underwent WES twice during the course of this study, once to be analysed as part of the family 'trio' and once to be analysed as part of an 'overlap' cohort of unrelated singleton cases. There were some differences in the techniques used for WES, which affect direct comparisons, but they are broadly comparable. Without the benefit of parental data the number of candidate genes in subject 1 increased from 2 (*de novo* dominant) to 95 (inherited dominant) variants; illustrating the substantial power of using trios rather than singletons in WES. Similarly, the DDD study found that the addition of WES data from both parents of an affected child resulted in a ten-fold reduction in the number of potential casual variants that required evaluation as compared to the sequence data of just the affected child alone (97). In comparison, where one or both parents (allowing for both dominant or recessive disease) are similarly affected then 'trio sequencing' offers only a three times or 1.5 times reduction respectively (97).

Where parental samples are unavailable then an 'overlap approach' of multiple unrelated individuals has proved effective in elucidating the underlying molecular cause of several rare, severe developmental disorders including Kabuki syndrome (166) and Bohring–Optiz syndrome (214). We used the ExAC database and xBrowse reference samples (n=375 exomes) for filtering out variants previously seen in population cohorts; substantially reducing the number of candidate variants but a considerable number of variants remained. A limitation of the way this data has been generated and analysed is that we do not have normal controls sequenced using the same pipeline, which would be beneficial for filtering to

identify likely sequencing errors generated through our specific NGS and bioinformatic pipelines. Of the two candidate genes identified through the overlap approach, which would be consistent with bi-allelic inheritance (*FRG1* and *IGHJ6*); both appeared to the sequencing errors on visual inspection of sequencing reads. A large set of internally generated exome sequences allows for the exclusion of systematic artefacts specific to the peculiarities of a production pipeline (89, 96) and this will be incorporated into future WES projects.

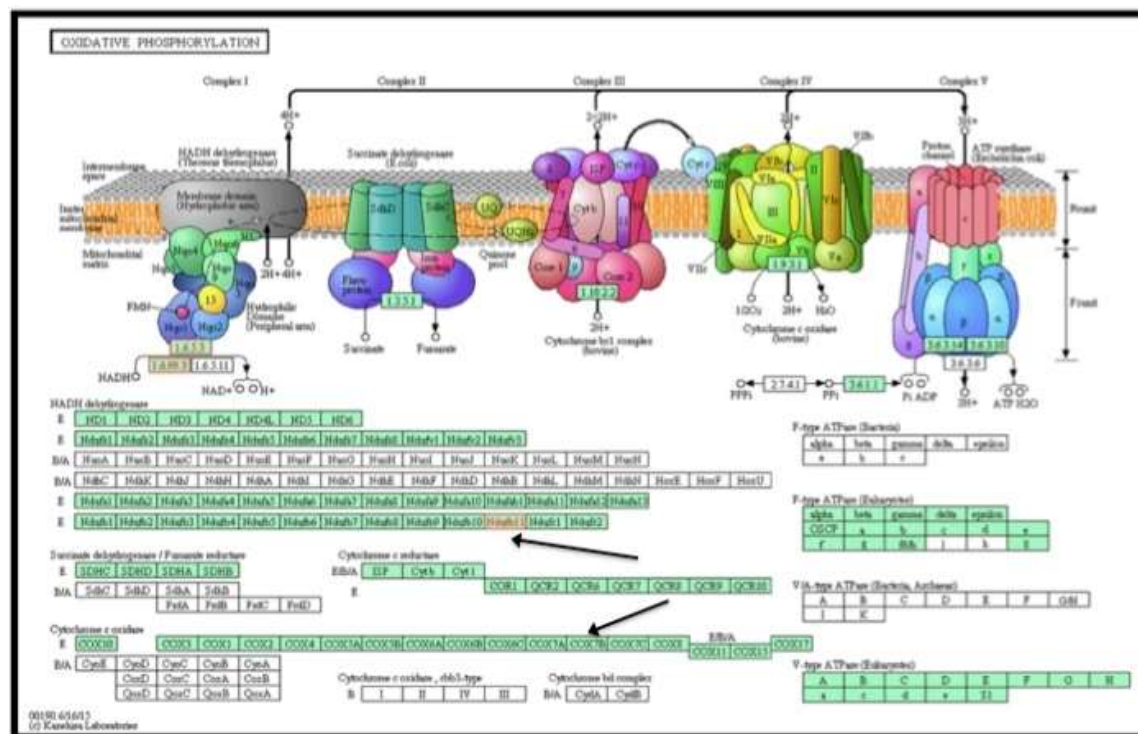
3.4.2 WES using a ‘trio’ approach: Identification of *NDUFB11* as the causative gene of Histiocytoid CM

Of the two *de novo* variants in subject 1 (affected child in the family ‘trio’) *NDUFB11* (NADH dehydrogenase (ubiquinone) 1 beta sub-complex, 11, 17.3kDa), which is a protein-coding gene with the alternative gene names CI-ESSS, ESSS, NP17.3, Np15, P17.3 was felt to be the strongest candidate gene, due to its position on the X chromosome, high expression in the heart and its role in mitochondrial function, see Table 3-7. *NDUFB11* is a component of mitochondrial complex I (see Figure 3-13) within the Mitochondrial Respiratory Chain (MRC), which catalyses the first step in the electron transport chain, the transfer of two electrons from NADH to ubiquinone, coupled to the translocation of four protons across the membrane (Carroll et al., 2002).

Further support for the pathogenic nature of *NDUFB11* variant came from the knowledge that truncating variants in *NDUFB11* are uncommon; within the ExAC database there are no truncating variants (nonsense, frameshift or essential splice site) variants in *NDUFB11*, and no homozygous missense variants (<http://exac.broadinstitute.org/> accessed on 26th October 2015). Sanger sequencing of the coding sequence of *NDUFB11* in four additional unrelated cases of Histiocytoid CM failed to identify any additional *NDUFB11* variants. While we were exploring other avenues to replicate and or validate our findings, two independent groups reported *NDUFB11* as the causative gene in both Histiocytoid CM (204) and Microphthalmia Linear Skin Defects syndrome (MLS syndrome)(219). The associated

mutations (and the mutation identified in Subject 1 in our study), including reported diagnosis is shown in Figure 3-14.

Figure 3-13 The KEGG Pathway of Oxidative Phosphorylation (Mitochondrial Respiratory Chain)



The KEGG pathway is a molecular interaction network diagram represented in terms of KEGG Orthology (KO) groups.

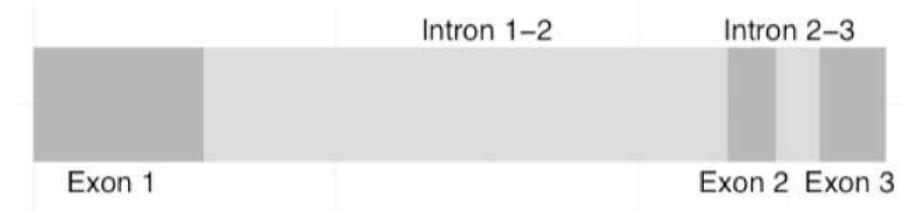
(<http://www.genome.jp/kegg->

<http://www.ebi.ac.uk/Tools/kegg/pathway/hsa00190+54539>). Accessed on 14th September 2015.

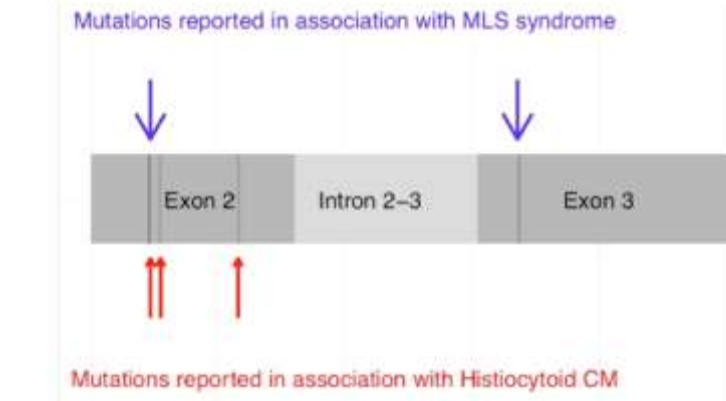
NDUFB11 plays a key role in Complex I. COX7B plays a key role in complex IV. Arrows highlight both.

Figure 3-14 Illustration of *NDUFB11* showing reported truncating mutations and associated phenotypes

A Illustration of *NDUFB11* showing each of the three exons (dark gray) and two introns (light grey).



B *NDUFB11* zoomed in on exons 2 and 3 to show the reported truncating mutations.



C Table of all reported truncating variants in *NDUFB11* and associated phenotypes.

Reported diagnosis	Histiocytoid CM (case presented here)	MLS syndrome	MLS syndrome	Histiocytoid CM	Histiocytoid CM
Protein Variant	p. Arg88Ter	p. Arg88Ter	p. Arg134Serfs*3	p. Trp85Ter	p. Tyr108Ter
Genomic Coordinate	47002089	47002089	47001806	47002097	47002027

3.4.3 Evidence for pathological consequence of truncating variants in *NDUFB11*

Functional studies confirming the pathological consequences of truncating mutations in *NDUFB11* include shRNA-mediated *NDUFB11* knockdown in HeLa cells, to demonstrate that *NDUFB11* is essential for MRC complex 1 assembly and activity as well as cell growth and survival (219) and morpholino-mediated knockdown of *ndufb11* in zebrafish embryos; demonstrating defective cardiac tissue with evidence of cardiomegaly, looping defects and

arrhythmia(204). In the mouse, *Ndufb11* is expressed throughout embryonic development and in the adult mouse it is expressed in multiple organs, notably, the brain, heart, kidney and skeletal muscle. Interestingly, in embryonic development the transcript is enriched in the anlagen of the facial structures, the heart, eyes, brain and spinal cord (220) which may explain many of the observed disease traits which are features of *NDUFB11* mutations (see Figure 3-15 and Table 3-13).

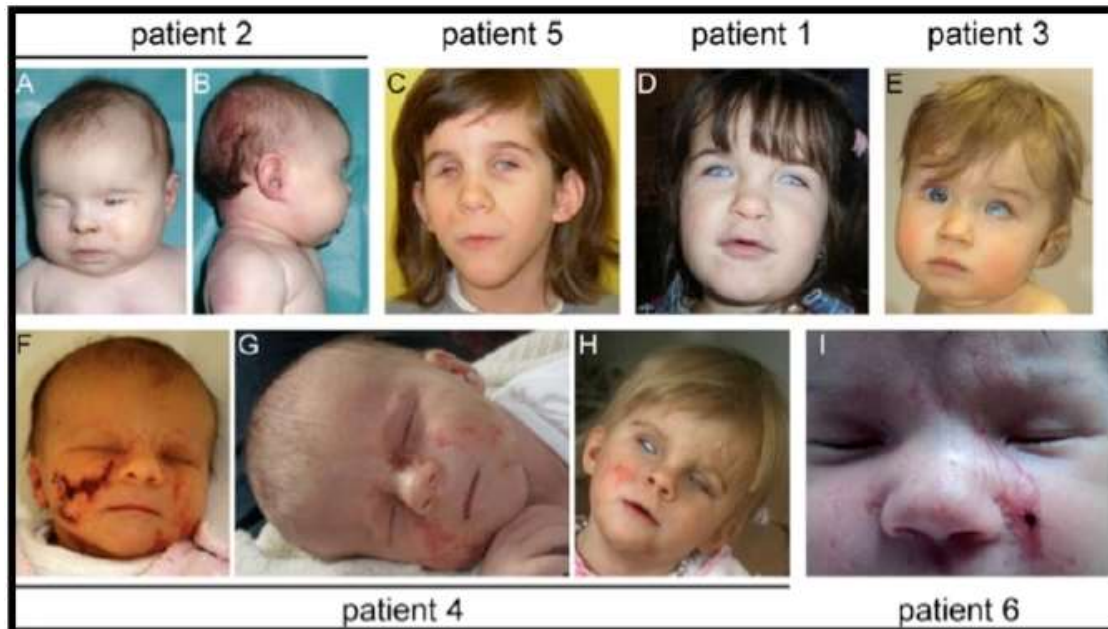
3.4.3.1 Microphthalmia with Linear Skin Defects (MLS) syndrome and Histiocytoid Cardiomyopathy (CM) are allelic (genetically related) disorders

First described in 1988 (221), the Microphthalmia with Linear Skin Defects (MLS) Syndrome (OMIN 309801) is also known as MIDAS (Microphthalmia, Dermal Aplasia, Sclerocornea) syndrome or Gazali-Temple syndrome (222-224). It is a rare, X-linked dominant disorder with male lethality *in utero* (219). It is often classified as either a neurocutaneous or neurodevelopmental disorder. In affected females, it is characterized by unilateral or bilateral microphthalmia and linear skin defects which have with a ‘scalded’ appearance (222). The linear skin defects are classically limited to the face and neck and occur along the lines of Blaschko, which correspond to cell migration pathways evident during embryonic and fetal skin development (see Figure 3-15).

Typically the skin lesions heal with age leaving hyper pigmented lesions (63). Happle *et al* (225) suggested the mnemonic MIDAS (Microphthalmia, Dermal Aplasia and Sclerocornea) and proposed that it was distinct from Goltz focal dermal hypoplasia (another X-linked multiple system disorder with skin involvement) because of involvement only of the face and neck and as MIDAS patients lack the dermal aplasia and fat herniation through the lesions and the skeletal abnormalities seen in Goltz syndrome. The eponym has fallen

out of favour as dermal aplasia has not been demonstrated on histology of the skin defects in MLS patients (226) and so we use the term MLS in preference.

Figure 3-15 Typical skin lesions and eye features of patients with MLS syndrome



Photographs showing the typical skin lesions and eye features of patients with MLS syndrome. All patients with Microphthalmia and Linear Skin defects (MLS syndrome) included here had either a heterozygous *HCCS* mutation or Xp22 monosomy. Eye abnormalities include microphthalmia, sclerocornea and anophthalmia. Linear skin defects may be particularly prominent at birth but heal with age (patient 4, photographs F,G,H). Image adapted from (67) and reproduced with permission of the rights holder, BioMed Central Ltd.

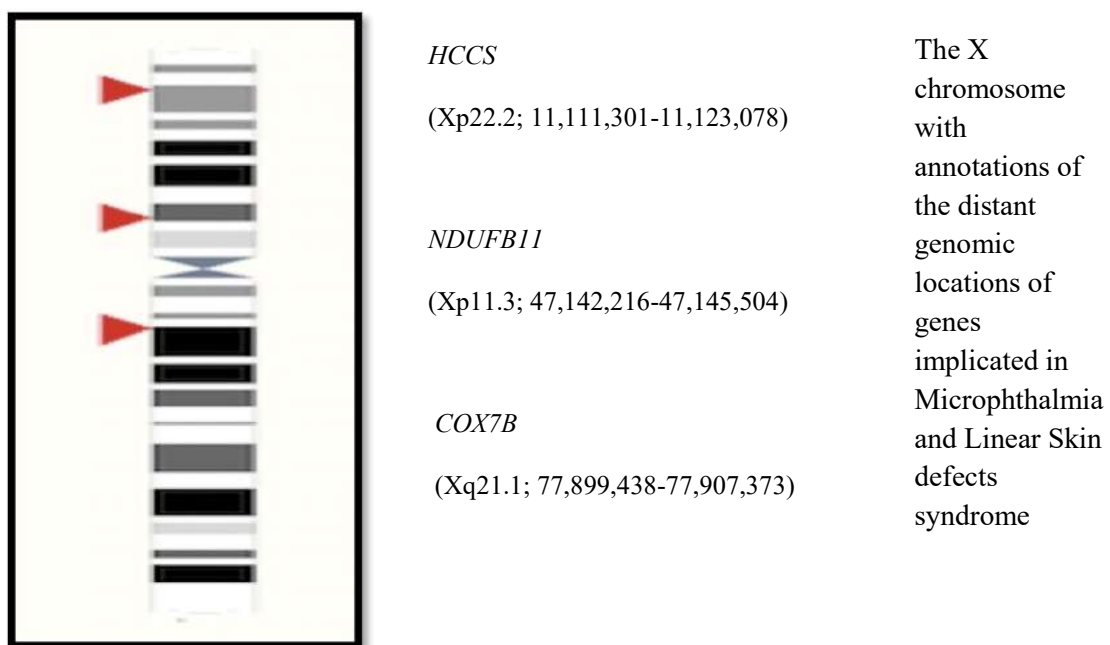
3.4.4 The Genetic basis of MLS syndrome

3.4.4.1 Monosomy Xp22.2

The vast majority of early case reports of patients with MLS were associated with chromosomal abnormalities including terminal deletions and unbalanced translocations of the Xp22.3 region (222, 224, 227), which resulted in monosomy for the Xp22.2 region. Although almost exclusively reported in females, rare males with karyotype abnormalities have been described (221, 222, 227, 228). For example, a new-born male with the typical eye and skin changes of MLS, with the additional findings of hypospadias with chordee, secundum Atrial Septal Defect, anal fistula and agenesis of the corpus callosum, in whom Fluorescence in situ hybridization (FISH) studies with X- and Y- specific probes identified a

derivative X chromosome formed from a translocation between Yp and distal Xp (226). A minimal critical region for MLS of 450-550kb at Xp22.2 was identified by cytogenetic and breakpoint analyses of affected and unaffected individuals (229), see Figure 3-16. The Xp22.2 region contains *HCCS* but not *NDUFB11* or *COX7B*.

Figure 3-16 Ideogram of the X chromosome with annotations of genes implicated in MLS syndrome



3.4.4.2 HCCS

More recently Wimplinger *et al* (2006) examined MLS patients with normal karyotypes and identified short deletions or point mutations in the *HCCS* gene (*HCCS* (OMIN 30056)) which encodes holocytochrome c-type synthase that functions as a heme lyase by covalently adding the prothetic heme group to both apocytochrome c and c1, an important mitochondrially targeted protein. It is involved in oxidative phosphorylation and apoptotic

cell death. Impaired HCCS function in yeast and mice results in OXPHOS defects, supporting a crucial role for HCCS in the formation and function of the MRC. Van Rahden *et al* reported that *HCCS* was implicated in all chromosomal rearrangements reported in MLS syndrome-affected individuals reported to date (67). Using an alternative approach Prakash *et al* (230) engineered mouse models with overlapping deletions similar to the human MLS deletions. Through the generation and characterization of these mutant mice they found that inactivation of the *Hccs* genes results in early embryonic lethality of the mutant mice and that expression of a human *HCCS* transgene rescued the phenotype. The ‘rescued’ male and female animals had no clear phenotype, and they concluded that loss of HCCS causes the male lethality of human MLS syndrome. The authors suggested several possible explanations for the variability in the phenotype of the embryos including variable patterns of X chromosome inactivation. They noted that

‘since *Hccs* is the most important contributor to the lethality, there may be residual mitochondrial oxidative phosphorylation in the early embryo because of the presence of maternally inherited mitochondria’ (230).

3.4.4.3 COX7B

Recognizing the genetic heterogeneity of MLS, which was evident from cases of this characteristic neurodevelopmental phenotype without mutations in *HCCS*, Indrieri *et al* (231) identified possible candidates and analysed the X-linked *COX7B*, in fourteen female cases in whom no *HCCS* mutation, or deletion or translocation of the Xp22 region could be identified (see Figure 3-16). They identified deleterious *de novo* mutations in *COX7B* in two simplex cases and a nonsense mutation, segregating with disease in a familial case (231). *COX7B* encodes a poorly characterized structural subunit of cytochrome c oxidase (COX), the MRC complex IV (see Figure 3-13). Indrieri *et al* determined that *COX7B* is crucial for COX assembly, COX activity, and mitochondrial respiration. Down regulation of the *COX7B* ortholog (*cox7B*) in medaka (*Oryzias latipes*; Japanese Rice Fish) resulted in microcephaly

and microphthalmia, recapitulating the MLS phenotype and determining an essential function of complex IV activity in vertebrate central nervous system development (231). The discovery of *COX7B* which encodes a structural subunit of MRC complex IV, as the second gene (at that time) for MLS provided additional evidence for implication of mitochondrial dysfunction as the underlying cause of MLS syndrome(67).

3.4.5 Diagnostic criteria for Microphthalmia and Linear Skin defects syndrome

MLS syndrome has been associated with widespread systemic effects (see Table 3-13 and Table 3-14). In the last revision of the diagnostic criteria in 2011 (232) Morelo & Franco proposed that a clear diagnosis of MLS could be made in individuals when the two major criteria are evident, although we note that these features are not invariably present and the characteristic skin findings tend to resolve and so may not be appreciated in an adult (see Table 3-13 and

Figure 3-15. Morelo & Franco note that individuals with a molecular diagnosis of MLS (at that time, known causes were either a chromosomal abnormality involving the Xp22 region or a mutation in *HCCS*) in whom only one of the two major criteria was present have been reported; there were some individuals showing characteristic skin defects without any ocular abnormalities and others showing an ocular phenotype without any skin lesions(232). They proposed that minor criteria (see Table 3-14) in the presence of a family history consistent with X-linked inheritance with male lethality was supportive of a clinical diagnosis of MLS syndrome (232). It is notable that many of the associated features reported in MLS syndrome (see Table 3-14) have also been reported in association with cases of Histiocytoid CM (see Table 3-1).

In light of the identification of *COX7B* and *NDUFB11* as causative genes of MLS syndrome revision of the diagnostic criteria would be appropriate to encompass these and place greater

weight on the role of molecular diagnosis. The majority of case reports with a defined molecular aetiology in MLS still relate to individuals with cytogenetic abnormalities involving Xp22 (see Figure 3-16). It may be that some of the diverse minor criteria may relate to other genes involved in the cytogenetic abnormality rather than as clearly delineated features of the MLS syndrome. Further genotype- phenotype correlations studies will clarify this.

Table 3-13 Major diagnostic criteria for diagnosis of Microphthalmia and Linear Skin Defects syndrome

System	Feature	Reference
Ocular	• Microphthalmia and/or Anophthalmia • May be unilateral or bilateral • Reported in 93% of affected individuals	(227, 232) (222) (233)
Skin	• Linear skin defects • Present at birth, usually located on the face and neck • More rarely, may involve the scalp and upper trunk • Characterised by areas of aplastic skin, which heal with age to form hyper-pigmented areas • Reported in 95% of affected individuals	(227, 232)

Major diagnostic criteria for diagnosis of Microphthalmia and Linear Skin Defects syndrome (adapted from Morelo & Franco (232))

Table 3-14 Additional features reported in association with MLS syndrome

Ocular features		CNS features	Cardiac features	Other features	
Sclerocornea (227)	Corneal leukoma (232)	Microcephaly (227)	Histiocytoid CM (203, 232)	Facial dysmorphism or asymmetry (231)	Anterior or imperforate anus (222)
Corneal opacities (221, 222, 227, 233)	Short palpebral fissures (203)	ACC (227)	HCM (225, 232)	Nail dystrophy (231)	Pseudotail (234)
Orbital cysts (222)	Aniridia (232)	Anencephaly (232)	ASD & VSD (232)	Minor cutaneous syndactyly (222)	Bicornuate uterus (232)
Microcornea (232)	Cataracts (232)	Ventriculomegaly (227)	SVT (232)	Short stature (231, 235)	Ambiguous genitalia (232)
Congenital glaucoma with total/peripheral anterior synechiae (232)	Hypopigmented areas of the retinal pigment epithelium (232)	Developmental delay (227)	VF (232)	Renal agenesis and ureteral duplication (231)	Penile hypospadias in rare males with a 46, XX karyotype (232)
Peters anomaly (232)	Eyelid fissures (232)		LVH (231)		

Those considered as minor diagnostic criteria are highlighted in bold. Adapted from diagnostic criteria proposed by Morelo & Franco(232). Last revision was in 2011. ACC; Agenesis of the corpus callosum; ASD; Atrial Septal Defect; LVH; Left Ventricular Hypertrophy; HCM; Hypertrophic Cardiomyopathy; MLS; Microphthalmia and Linear Skin defects Syndrome; SVT; Supraventricular Tachycardia; VF: Ventricular Septal Defect; VSD; Ventricular Septal Defect.

3.4.6 Evidence for a link between Histiocytoid CM and MLS

In considering the role of *NDUFB11* in MLS syndrome and Histiocytoid CM, two disorders which until this point were considered distinct phenotypes it is reasonable to consider that these disorders are not allelic (genetically related) but simply represent rare patients with dual diagnoses. Previous WES studies have unexpectedly revealed two molecular diagnoses of non-overlapping genetic disorders in up to 6% of patients (39). Such examples of patients with “multiple hits” or “blended Phenotypes” are likely to become commonplace as WES and subsequently WGS become more widespread (39). In this case, however, there is substantial background evidence to suggest MLS syndrome and Histiocytoid CM are allelic.

Over twenty years ago Bird *et al* (203) reported an infant girl with the characteristic skin lesions of the then newly recognized MLS syndrome, she died suddenly and unexpectedly at four months of age. Death was attributed to oncocytic cardiomyopathy (Histiocytoid CM), with oncocytic cells also present in the choroid plexus and adrenal gland. There was an antecedent history of Wolf Parkinson White Syndrome. The infant had a normal 46, XX karyotype. The palpebral fissures were relatively short but there were no other corneal defects. The authors noted that the

‘co-existence of two rare conditions, one of which mapped to the X chromosome, and an excess of affected females with oncocytic cardiomyopathy, make it likely that oncocytic cardiomyopathy is also X-linked, with Xp22 being a candidate region. Overlapping manifestations in the two conditions (ocular abnormalities in cases of oncocytic cardiomyopathy and arrhythmias in MLS) offer additional support for this hypothesis’ (203).

Additional cases reported prior to this included a female infant with oncocytic cardiomyopathy, bilateral microphthalmia with corneal opacity and no right lens with complete absence of the corpus callosum and dilation of the posterior horns of the lateral ventricles; there was no mention of linear streaks (170). In addition, a female Caucasian child with Histiocytoid CM who presented with VT, VF and cardiac arrest with seizures at age three years and six months. She had left microphthalmia, bilateral oblong pupils and

hazy corneas (196) Kutsche *et al* (236) noted that both of these cases had normal karyotypes suggesting that the genetic defect that caused the MLS syndrome might also be implicated in Histiocytoid CM and observed the gene for holocytochrome c-type synthetase, *HCCS*, resided in the critical region for MLS syndrome.

3.4.7 Do mutations in *HCCS* and *COX7B* also cause Histiocytoid CM?

Noting that, a missense mutation in the mitochondrial cytochrome b (MT-CYB) gene was found in a patient with Histiocytoid CM (207) and thus implicating the MRC in the pathogenesis of Histiocytoid CM, Kutsche *et al* (236) suggested reduced *HCCS* enzyme activity could give rise to a nuclear-encoded respiratory chain defect which, in turn, may lead to (histiocytoid) cardiomyopathy in MLS cases (236). More recently, Van Rahden *et al* (67) reported the clinical spectrum of six females with *HCCS* mutations including one with a *de novo* nonsense mutation c.589C>T (p. R197*) who died at four months of age. In addition to her diagnostic features of MLS syndrome (linear skin defects on her neck, microphthalmia with unilateral sclerocornea (right), and anophthalmia with optic nerve hypoplasia (left)), she also had abnormal myelination, with a hypoplastic corpus callosum, absence of the septum pellucidum. Cardiac manifestations included ventricular tachycardia, poor contraction of the left ventricle, Histiocytoid CM and eosinophilic cell infiltration(67). We hypothesise that mutations in *HCCS* and *COX7B*, which are implicated in MLS syndrome, may also be found to be causative of apparently isolated Histiocytoid CM. We interrogated our four additional unrelated cases of Histiocytoid CM and reviewed publically available WES data (204) from three additional cases of Histiocytoid CM without *NDUFB11* mutations and found no evidence of *de novo* or rare variants in either *HCCS* or *COX7B* (see Table 3-8), although this cohort (n=7) is too small to definitively exclude a role.

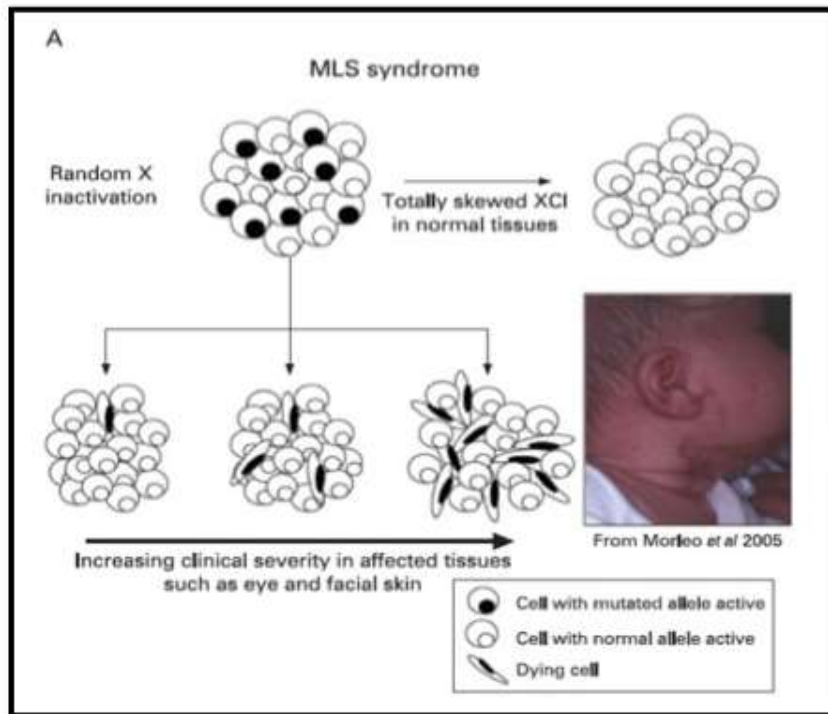
3.4.8 A single mutation with multiple phenotypes

Commonly in humans, different mutations in the same gene cause distinct phenotypes. The molecular mechanisms through which this results remain largely unknown, but possible loss-of-function versus gain-of-function effects are possible mechanisms (237). In recent years molecular diagnostics has determined locus heterogeneity underlying clinical phenotypes which may otherwise be co-attributed as a distinct syndrome based on common clinical features, and also connected phenotypically disparate disorders to a single locus through allelic affinity (98). In this study, an identical mutation in *NDUFB11* appears to be causative of two distinct phenotypes (see Table 3-15). We propose that Histiocytoid CM and MLS are allelic (genetically related) and suggest the development of differing phenotypic effects is influenced by factors other than the specific mutation. Key determinants suggested to explain phenotypic variability in any genetic disorder include, environment, stochastic (chance) factors and genetic background (93). In mitochondrial disease, genetic background of both rare and common variants is particularly likely to impact upon the phenotype as mitochondrial biogenesis and function involves up to 1500 nuclear genes, therefore allowing multiple opportunities for consequences on disease expression (59). The role of environmental factors is also likely to be of particular importance in mitochondrial disease as the activity of mitochondria, which are leading players in numerous essential metabolic pathways, is influenced by factors such as nutrition and exercise (59).

Molecular mechanisms, which could influence phenotype, include somatic mosaicism; it is recognised that mutations occurring in parental gametes can be responsible for a more severe phenotype than those, which arise, post-zygotically in later embryogenesis (164). In their case of MLS with the same mutation as the case of Histiocytoid CM presented here Van Rahden *et al* (219) noted that the mutated base thymine was present in 42 sequence reads and the wild-type base cytosine in 84, suggesting that the mutation was present in mosaic state in lymphocytes. DNA isolated from fibroblasts was also suggestive of mosaicism, with a low

signal for the variant (thymine) superimposed on the wild-type sequence (cytosine). In addition Shehtata *et al* (204) note in one of their cases with a nonsense *NDUFB11* mutation (case GHCG) a second nonsense mutation was detected in *Cytochrome b*, but only at a frequency of 20% and only in the cardiac tissue, possibly indicating a clonal selection of a somatic mutation in the diseased heart and suggest this is one of the potential avenues through which the expressivity of defects in MRC complex 1 activity might be regulated. In the case of genes on the X-chromosome, the pattern of X-Inactivation may also play a crucial role. Several authors have suggested tissue-specific differentially skewed X-inactivation of the X chromosome may play a critical role in the development of MLS syndrome (63, 238, 239), in which most females present with the classical phenotype but where there is significant intra-familial variability and variation in the clinical phenotype of simplex cases. Skewed X inactivation patterns have been detected in 16 out of 17 MLS patients analysed in one case series (63). Most visibly, the neonatal linear skin defects (see Figure 3-15 and Figure 3-17) could be due to a mosaic pattern of X-chromosome inactivation early in tissue development, with alternating regions of tissue skewed toward the normal X and regions skewed toward the abnormal X (63). It has been suggested that in skin the “sick” cells in which the normal X chromosome is inactivated may die and be replaced by cells in which the abnormal X is inactivated over time (63). Hobson *et al* (240) note that the skin lesions of MLS syndrome patients are replaced by hyper-pigmented lesion as they age which may reflect this underlying process. The resulting scars may be minimal in adulthood (241).

Figure 3-17 Schematic representation of X Chromosome inactivation in female somatic cells in MLS syndrome



In normal conditions the expected chimerism results in 50% of cells that inactivate one X, whereas the remaining cells inactivate the other X chromosome. In Microphthalmia and Linear Skin defects (MLS) syndrome, skewing of X inactivation can occur and vary among individuals, tissues and in different stages of development, thus influencing the phenotypic variability observed in this syndrome. Top represents cells undergoing totally skewed X inactivation that forces preferential activation of the unaffected X in normal tissues. Bottom schematizes the increasing severity of clinical manifestations depending on the proportion of “suffering” cells whose normal X is inactivated in the affected tissue, or at a specific time of embryonic development. The picture shows the typical linear skin lesions observed in MLS patients. Image adapted from (63) and reproduced with permission of the rights holder, BMJ Publishing Group Ltd.

Lastly, phenotypic differences may result from differing susceptibility to mitochondrial defects. Despite the evidence that mutations in mitochondrial proteins induce mitochondrial dysfunction, it remains unclear how specific genetic defects result in abnormalities at the level of cells, organs, or the whole organism. A considerable challenge in establishing the effects of specific mutations is that the consequences of mitochondrial dysfunction are influenced by the activity of compensatory stress-response pathways, namely, mitochondrial biogenesis, resulting in an increase in the expression of oxidative phosphorylation proteins, the removal of dysfunctional mitochondria and a switch to a more glycolytic mode of ATP production (70). Van Rahden *et al* (67, 219) suggested that

‘the different ability of developing tissues and organs to cope with cells harbouring a defective OXPHOS system (as a result of an active mutant X chromosome) may account for

the high clinical variability in MLS syndrome and could explain some specific clinical features, such as (Histiocytoid) cardiomyopathy, agenesis of the corpus callosum and deafness which are typically found in OXPHOS disorders’(67).

Table 3-15 Detailed comparison of the clinical features in the case of Histiocytoid CM and a previously reported case with a phenotype of MLS Syndrome and an identical *NDUFB11* mutation (c.262C>T; p. Arg88Ter)

Phenotypic feature	Case reported here (Subject 1)	Case reported by van Rahden(219)
Antenatal History	Healthy, non-consanguineous parents Alternating bradycardia and tachycardia noted at 27 weeks; Emergency Caesarean section at 38 weeks for fetal tachycardia	Healthy, non-consanguineous parents
Sex	Female	Female
Birth weight	3690g (91st centile)	3,060g (10th-25th centile)
Cardiac Arrhythmias	Neonatal episodes of supraventricular tachycardia. Collapse with first documented ventricular tachycardia (VT) at seven months. Continued VT “Storms” necessitating drug treatment, dual chamber implantable cardiac defibrillator, and left thoracic sympathectomy. Cardiac transplantation was carried out at 13 months	Sudden unexpected cardiac arrest aged six months. Repeated treatment for ventricular arrhythmias. Death within several weeks
Cardiomyopathy	Features of Left Ventricular Non-Compaction. Preserved systolic function	-
Cardiac histology	Histiocytoid CM	Histiocytoid CM
Eye	Microphthalmia	-
	Scleroderma	-
	Other eye abnormalities	Intermittent squint. Lacrimal duct atresia
Neuromuscular	Mild to moderate bulbar palsy	Axial hypotonia present from birth
Skin abnormalities	-	Linear skin defects on nose, chin and neck present at birth, disappeared in the first few months of life.
Thyroid abnormalities	Focal histiocytoid change in the thyroid (and also lungs and choroid plexus of the brain)	Oncocytic metaplasia evidence on post mortem
Other anomalies	Severe feeding difficulties gastro-oesophageal reflux, requiring fun Nissen fundoplication and Percutaneous Endoscopic Gastrostomy	Failure to thrive documented from one month of age
Evidence of somatic mosaicism	No evidence in DNA from lymphocytes	Yes, in DNA from lymphocytes and fibroblasts

CM; Cardiomyopathy; MLS; Microphthalmia and Linear Skin Defects syndrome

3.4.9 Whole Exome Sequencing using an overlap approach: Candidate genes identified

The overlap approach generated a greater number of candidate genes (n=16) than the trio approach, despite filtering against population data from ExAC (see Table 3-8). Mono-allelic, bi-allelic and mitochondrial inheritance models were all considered. Close examination of the variants identified in the *ANKRD20A4*, which had initially seemed consistent with mono-allelic inheritance and *FRG1* and *IGHJ6*, which had seemed consistent with bi-allelic inheritance ruled out these genes as candidates (see Table 3-9, Table 3-10 & Table 3-11). Of the mtDNA variants identified (see Table 3-12) only one variant (m.3308T>G) was seen at a low population frequency (reported four times in MITOMAP out of a possible 30589 mitochondrial DNA sequences; 0.0001) and therefore potentially consistent with the rarity and severity of disease in Histiocytoid CM. This locus: MT-ND1 is a sub-unit of complex 1 in the MRC, like *NDUFB11*. The nucleotide change results in a non-synonymous amino acid change (ND1: p. Met3308*). This variant has previously been reported in association with SIDs (see Table 3-16).

Table 3-16 Clinical details from previous reports of the m.3308T>G variant

Case	Reference	Comment
1	(242)	Identified in a cohort of 257 cases of SIDS, Sudden Infant Death attributed to infection and borderline SIDs cases and 102 control living infants where variation in the hyper-variable region1 (HVR-1) and coding regions of mtDNA were studied
2	(242)	
3	(243)	
4	(244)	Identified in 1 patient in a cohort of 8 PD patients and 9 controls who had sequencing of the entire mitochondrial genome of their substantia nigra. Age and sex unspecified

The m.3308T>G variant was identified in one case of Histiocytoid CM through WES of a cohort of four unrelated cases. It has previously been reported in four unrelated cases, including three cases of SIDs

An additional mutation (m. 3308T>C), in the same base pair, which is in the first triplet of the ND1 gene, has also been described. The (m. 3308T>C), has been described both in

association with bilateral striatal necrosis MELAS (245), maternally inherited diabetes and deafness (246) and in controls(247).

The m.3308T>G variant identified in case 20PO03585, represents a strong candidate for Histiocytoid CM. 20PO03585 was a female child who died suddenly and unexpectedly at nine weeks of age. She was born by emergency caesarean section at 34+5 weeks for worsening pre-eclampsia. There was no history of any illness prior to death. The baby's mother had a history of insulin dependent diabetes from one year of age, and there is also a history insulin dependent diabetes in the maternal sister, brother, aunt and grandmother. Previously testing for mitochondrial point mutations on liver DNA from the affected child had failed to detect any abnormality (m.3243A>G, m.8344A>G, m. 8993T>C/G, m.3260A>G). Testing for major mitochondrial DNA rearrangements failed. Chromosome analysis was that of a normal female. Macroscopic examination of the heart showed raised greyish areas over the apex of the heart, with thickened rather mucoid trabeculae left ventricle, with a greyish nodule evident at the tricuspid valve. Histology of the heart showed multiple macroscopically visible nodules to be well-demarcated aggregates of large rounded cells with clear cytoplasm and small central nuclei, in keeping with Histiocytoid CM. The nodules were predominantly sub-endocardial, but also present within the deeper myocardium and the pericardium. The conduction system was involved with small foci in the region of the AV node. A small aggregate of Histiocytoid cells was present at the base of the tricuspid valve.

The differential diagnosis included cardiac rhabdomyomas, which may occur as part of Tuberous Sclerosis Complex. There were no additional brain, kidney or skin findings to support this diagnosis and genetic analysis was undertaken using DHPLC to screen the coding exons of *TSC1* and *TSC2* and MLPA to detect large deletions or duplications of *TSC2*. This would be expected to identify the causative mutation in ~75% of cases. Sanger

validation of m.3308T>G variant in the infant is on going. Maternal DNA samples have been requested.

3.4.10 What is the cause of Histiocytoid CM in the outstanding cases?

The success rate of WES in this cohort of five cases of Histiocytoid CM is 20-40%, which is in line with previous studies, which found an overall success rate of ~25% (39). Failures to identify a molecular diagnosis can be broadly considered as technical or analytical (96).

These may involve issues relating to the sequencing and bioinformatics pipeline; namely, lack of sequence coverage, variant calling issues and misinterpretation of variants or issues relating to the variant; for example, that the cause is located outside the coding sequence or is a large indel or structural variant missed by exome sequencing (38). Some of these issues could be overcome by Whole Genome Sequencing (WGS), which has advantages over WES for superior detection of structural variation, such as copy-number variants e.g. DiGeorge syndrome and detection of variants in non-exonic regulatory regions and common variants in intronic and intergenic regions (40). Further types of DNA variant which are currently not well detected or undetectable by either WES or WGS include repetitive DNA, such as trinucleotide repeats e.g. Fragile X syndrome, aneuploidy e.g. Down's syndrome and epigenetic alterations e.g. Prader-Willi syndrome (40).

Furthermore, certain genes remain challenging for current WES pipelines, and these include genes which contain an internal duplication (e.g. *FLG*) or a nearby pseudogene (e.g. *ANHAK2*), or in cases where recent duplications and /or annotation errors have resulted in the same sequence being assigned to two genes (e.g. *SLX1A* and *SLX1B*). Current pipelines may systematically either under call variation in these genes because of uncertainty of which gene to assign them to leading to reads which are unmapped, or may result in overcalling false variants owing to read misplacement(55).

The coverage in our samples of both nuclear and mitochondrial genes is high (see Table 3-3 & Table 3-4) and therefore it is less likely that we have missed causal variants due to low coverage but it may be that candidate nuclear encoded mitochondrial genes were not present on the targeted exome capture used. Gillisen *et al* noted that their lack of success in identify *MLL2* as the causative gene for Kabuki syndrome (unknown at the time) was due to *MLL2* not being represented on the exome enrichment kit they had used at the time and therefore no sequence data for that gene was available(38).

In this study, we had identified *NDUFB11* as a strong candidate gene at an early stage but it was only with the publication of other reports of similar findings in other cases that we were able to confirm this was the casual variant in subject 1. Similarly, studies of the yield from WES show a substantial proportion (25-30%) of diagnostic success depends on recent progress in the discovery of genes underlying disease (39). Revision of our current data as additional cases of Histiocytoid CM undergo WES, and new disease-gene discoveries are made, may reveal casual variants, which are currently overlooked.

Further explanations for why a genetic cause may not be identified using WES include issues with the initial samples selected- either clinical heterogeneity or incorrect diagnosis. We had limited data on the additional cases recruited through UK regional genetics services used in the overlap approach. Previously successful examples of WES in multiple unrelated affected cases have used highly selected cases, which have been assessed by panels of expert as with the most similar phenotypes (214). Nevertheless, many other WES studies have identified patients with overlapping characteristic clinical presentations where deleterious variants are not detected in identified disease genes e.g. only 26 of the 43 patients with Kabuki syndrome had mutations in the causative gene *MLL2* (166). In reference to this Lu *et al* suggest that “In addition to indicating locus heterogeneity, these results suggest that complex genetic mechanisms involving oligogenic inheritance, with multiple causative

alleles, modifier alleles, or both, are probably more common than previously appreciated”(45).

Additional nuclear encoded or mtDNA genes within the MRC are good candidates for further causes of both Histiocytoid CM and MLS syndrome. As the fraction of captured mitochondrial sequences is linked to the relative abundance of the corresponding mitochondrial genome in the original total DNA extract, then consideration should be given to the most appropriate sample types to further investigate this. For example heart and skeletal muscle contain more mtDNA per cell than peripheral blood (248).

3.4.11 Treatment options in Histiocytoid CM

The diagnosis of Histiocytoid CM is almost exclusively a post-mortem diagnosis with few reports of survival. Surgical intervention in a female infant, presenting at eight months of age with malignant ventricular and supraventricular arrhythmias was reported over 30 years ago. During exploration of the heart on cardiopulmonary bypass, a 1cm diameter raised whitish yellow nodular area in the membranous ventricular septum was identified and removed, complete heart block was created and a permanent pacemaker inserted. The authors report the child was well with no evidence of psychomotor delay at seven years (McGregor et al, 1984). No follow-up reports on whether she survived to adulthood or reproduced are evident in the literature.

Zangwill *et al* reported the first case of recovery in a child with Histiocytoid CM following orthotopic heart transplantation for global ventricular dysfunction (200). Two further cases of successful orthotopic heart transplant; in infants with Histiocytoid CM, with LVNC were recently reported (187). The first of these cases was a ten-month-old female Asian infant who presented with fatigue and feeding intolerance, where investigations revealed left ventricular dysfunction with LVNC, atrial flutter and ventricular tachycardia were also noted. She underwent orthotopic heart transplantation, with the explanted heart showing typical

features of LVNC and macroscopic and histological features of Histiocytoid CM. The authors report she is doing well eight years post transplant(187). The second case they present is less convincing in the absence of a molecular diagnosis; a six week old African American male infant presenting with a two day history of feeding intolerance and tachypnoea in whom investigations revealed LVNC and LVEF ~27%. Genetic investigation revealed a Variant of Uncertain Significance (VUS) in *TTN* c.56984delC. He deteriorated clinically and was placed on Extracorporeal Membrane Oxygenation (ECMO) support, with subsequent implantation of a Left Ventricular Assist Device (LVAD) with resection of endocardial fibroelastosis and atrial septal defect closure. Biopsy at that time of the left ventricle was reported to show bundles of abnormal myocytes with abundant, foamy cytoplasm consistent with Histiocytoid CM. The patient underwent cardiac transplantation and examination of the explanted heart confirmed LVNC but did not reveal any further evidence of Histiocytoid CM (187).

Kearney *et al* reported their experience with eleven infants, aged two years or less who presented with malignant tachyarrhythmias. Nine of the eleven patients who underwent electrophysiological mapping and surgical excision of lesions consistent with Histiocytoid CM, which the authors termed “myocardial hamartomas”, survived. Follow-up periods ranged from one month to six years. Malhotra *et al* note that such reports of clinical manifestations of the Histiocytoid CM being abolished by surgical excision of nodules of Histiocytoid cells, goes against the concept of Histiocytoid CM as a mitochondrial disorder as generally mitochondrial cardiomyopathies show progressive clinical deterioration with age(180). With our current knowledge that Histiocytoid CM may be due to a nuclear gene mutation on the X chromosome, we hypothesize that the pattern of X-inactivation within the cardiac tissue could lead to localised areas of affected tissues, within otherwise ‘normal’ cardiac tissue. Histological descriptions would be consistent with this ‘microscopic sections of the heart showed multiple foci of sub-endocardial collections of large, plump, and foamy-appearing cells. Intervening normal myocardium was seen’(249).

It has been recognised for over twenty years that occasional patients with mitochondrial disorders, particularly with a myopathic presentation, may show a clinical response to supplementation with riboflavin (vitamin B₂)(85). Haack *et al* reported on one of their cases with compound heterozygous mutations in *ACAD9* (a member of the mitochondrial acyl-CoA dehydrogenase protein family) causing a complex 1 deficiency with cardiomyopathy, where there was a clinical response to treatment with daily riboflavin treatment (100mg). The authors note that beneficial effects of such treatment have repeatedly been reported in individuals with mutations encoding enzymes, which contain FAD and flavin mononucleotide (FMN). They hypothesise that

‘the identification of pathogenic mutations in *ACAD9*, which encodes a FAD-containing flavoprotein, offers a rational, mechanistic explanation of our clinical observation’(92).

The underlying molecular basis of the MRC deficiency may impact upon the potential success of treatment. Gotz *et al* reported of an infant with homozygous missense mutations in *AARS2*, who presented at 3.5 months with failure to thrive, poor feeding, delayed motor development and generalized weakness, and was found to have HCM. Muscle biopsy showed scattered cytochrome c oxidase (COX, MRC complex IV)- deficient muscle fibers, suggesting generalized muscle dysfunction and this was considered as contraindication for heart transplantation. She died at ten months of age despite intensive treatment for heart failure and supplementation of carnitine, CoQ, riboflavin and medium chain triglycerides(29). Fassone *et al* propose that a therapeutic trial of riboflavin should be mandatory for all patients with complex I deficiency, accepting that most patients are unlikely to respond (85). The molecular diagnosis in our proband (subject 1) was made after death and so no such therapeutic trial was possible, but this may be appropriate to consider in future cases identified with an *NDUFB11* mutation.

3.4.12 The role of WES or WGS in severely ill infants

WES was not available in our institution or through alternative routes in the health service to the proband (subject 1) and her parents during life. This situation is currently changing, as WES technology is more widely adopted and evidence gathers for the role that rapid genetic diagnosis can make in severely ill infants. In the first example of the utility of whole exome (or whole genome) sequencing to make a genetic diagnosis in an undiagnosed illness, Choi *et al* used WES in a child referred with a suspected diagnosis of Bartter syndrome (a renal salt-wasting disease) and detected novel homozygous missense variants (Asp652Asn) in solute carrier family 26, member 3 (*SLC26A3*), a gene known to cause a congenital chloride-losing diarrhoea. The unanticipated finding was subsequently confirmed with further clinical evaluation. Similar findings were replicated in five additional cases (out of 39 cases referred with suspected Bartter syndrome in whom the common gene mutations had not been identified).

Subsequently reports have emerged of molecular diagnosis made through WES, which have changed clinical management. For example, in a 15-month-old male child with severe inflammatory bowel disease in whom a novel hemizygous missense variant (Cys203Tyr) in X-linked inhibitor of apoptosis (XIAP) a known cause of X-linked lymphoproliferative syndrome type 2 (XLP2), was discovered, in which severe colitis was an unusual symptom. The diagnosis directed a specific treatment, namely allogenic haematopoietic progenitor cell transplant) that would not otherwise have been considered and appears to have lead to a successful outcome(250). More recently Willig *et al* described their experience with rapid WGS in 35 infants in their neonatal and paediatric intensive care units. In cases selected as likely to have a genetic basis for their conditions, they identified a diagnosis in 57% of cases. In 13/20 (65%) of these diagnoses were deemed of “acute clinical use” (251). The main rationale for use of WGS (which is more expensive) in preference for WES in such cases is that by omitting the capture step necessary for WES, time is saved, which may be critical. In

addition, WES costs are not directly proportional to the fraction of the genome targeted (~2%), as a result of several aspects of the process including imperfect capture specificity, skewing in the uniformity of target coverage which is introduced by the capture step and the costs associated with the additional step of exome capture (96). Beyond acutely ill infants, within the DDD project, analysis of the first 1133 children identified five children with a diagnostic variant in a treatable intellectual disability gene (*DHCR7*, *IVD*, *LMBRD1*, *MTR*, and *SLC2A1*)(97).

It is unclear whether an earlier molecular diagnosis would have affected the outcome in our patient, but certainly, moving forward, the ability to diagnose Histiocytoid CM early, on molecular grounds rather than requiring histological examination of the heart could direct clinical management. In addition and importantly it could have significantly shortened the time to diagnosis and avoided the “diagnostic odyssey” which is commonly seen in patients with ultra-rare diseases. Current costs of exome or genome sequencing for clinical diagnostic purposes are in the range of \$4,000 to \$15,000 per patient, with some laboratories offering lower per-person charges for family testing. As a comparison, the per-person charge for sequencing of an exome may be only two to four times the charge for some single-gene sequencing tests, making WES potentially more cost-efficient in a number of clinical scenarios(40). Such costs compare very favourably with the total cost of other investigations used to investigate this infant (see Table 3-2 and appendix Table 4-7 & Table 4-9). Applied earlier in the disease course WES could have been a cost-effective alternative. Extensive investigations including those directed on identifying mtDNA abnormalities had been undertaken. As in this case, most children with complex 1 deficiency have only minor and non-specific abnormalities in muscle histology e.g. mild lipid accumulation or fibre type disproportion (85). It is possible that if additional tests had been undertaken on cardiac tissue rather than the skeletal muscle biopsy, then they may have been informative, but at that point the diagnosis was evident from histology.

Benefits of a molecular diagnosis in the proband for this family have included allowing accurate recurrence risks for the parents, the option of prenatal testing and the ability to discharge first -degree relatives from clinical cardiac screening.

3.5 Conclusion

Histiocytoid CM is a genetically heterogeneous disorder. A *de novo* nonsense mutation in *NDUFB11* is the cause of Histiocytoid CM in subject 1 (the affected child in the family ‘trio’; 10CM02571). Histiocytoid CM and MLS are allelic (genetically related) and we hypothesize that additional genes implicated in MLS syndrome (*HCCS* and *COX7B*) are candidate genes for isolated Histiocytoid CM, though no variants have been identified in the small number of cases with data currently available to us (n=7). We propose that all female infants presenting with evidence of severe arrhythmias +/- cardiomyopathy should be investigated for mutations in *NDUFB11* (+/- *HCCS* and *COX7B*). A therapeutic trial of riboflavin should be considered in these patients. We would advise that cardiac screening for evidence of cardiomyopathy and arrhythmia should be considered in those individuals with a diagnosis of MLS syndrome. We expect that any live born affected males with pathogenic variants of *NDUFB11* will be mosaic for the pathogenic variant and we would expect that non-mosaic hemizygous males are not viable. *NDUFB11* mutations may therefore be a cause of male miscarriage. In subjects with clinical evidence of MLS syndrome or Histiocytoid CM, molecular testing of *NDUFB11*, *HCCS* and *COX7B* and screening for cytogenetic abnormalities is appropriate. Molecular diagnosis of an increasing number of cases of both MLS syndrome and Histiocytoid CM will delineate the range of phenotypic variability attributable to mutations in *NDUFB11*, *HCCS* and *COX7B* rather than to larger cytogenetic abnormalities, aiding genotype-phenotype correlations.

The m.3308T>G variant identified in case 20PO03585 represents a further strong candidate gene for Histiocytoid CM and further investigations are on-going.

3.5.1 Outline of future work

Sanger sequencing of the m.3308T>G variant, identified in case 20PO03585 is being undertaken. This represents a further strong candidate gene for Histiocytoid CM.

Additional analysis of the data on mtDNA variants from the ‘overlap’ approach, filtering for single rare variants in any mtDNA gene could potentially identify further candidates.

Additional data will now be available from the DDD study and this will be further interrogated to identify any rare sequence variants in *NDUFB11*, *HCCS* or *COX7B*.

References

1. Roger VL. The heart failure epidemic. *Int J Environ Res Public Health*. 2010;7(4):1807-30.
2. McNally EM, Barefield DY, Puckelwartz MJ. The genetic landscape of cardiomyopathy and its role in heart failure. *Cell Metab*. 2015;21(2):174-82.
3. Liew CC, Dzau VJ. Molecular genetics and genomics of heart failure. *Nat Rev Genet*. 2004;5(11):811-25.
4. Cahill TJ, Ashrafian H, Watkins H. Genetic cardiomyopathies causing heart failure. *Circ Res*. 2013;113(6):660-75.
5. Members ATF, McMurray JJV, Adamopoulos S, Anker SD, Auricchio A, Böhm M, et al. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012. 2012.
6. Camici PG, Prasad SK, Rimoldi OE. Stunning, hibernation, and assessment of myocardial viability. *Circulation*. 2008;117(1):103-14.
7. Braunwald E. Heart failure. *JACC Heart Fail*. 2013;1(1):1-20.
8. Nabel EG, Braunwald E. A tale of coronary artery disease and myocardial infarction. *N Engl J Med*. 2012;366(1):54-63.
9. Maron BJ, Towbin JA, Thiene G, Antzelevitch C, Corrado D, Arnett D, 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(14):1807-16.
10. Elliott P, Andersson B, Arbustini E, Bilinska Z, Cecchi F, Charron P, 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(2):270-6.
11. Geisterfer-Lowrance AA, Kass S, Tanigawa G, Vosberg HP, McKenna W, Seidman CE, et al. A molecular basis for familial hypertrophic cardiomyopathy: a beta cardiac myosin heavy chain gene missense mutation. *Cell*. 1990;62(5):999-1006.
12. Ganesh SK, Arnett DK, Assimes TL, Basson CT, Chakravarti A, Ellinor PT, et al. Genetics and genomics for the prevention and treatment of cardiovascular disease: update: a scientific statement from the American Heart Association. *Circulation*. 2013;128(25):2813-51.
13. Li L, Bainbridge MN, Tan Y, Willerson JT, Marian AJ. A Potential Oligogenic Etiology of Hypertrophic Cardiomyopathy: A Classic Single-Gene Disorder. *Circ Res*. 2017;120(7):1084-90.
14. Authors/Task Force m, Elliott PM, Anastakis A, Borger MA, Borggrefe M, Cecchi F, et al. 2014 ESC Guidelines on diagnosis and management of hypertrophic cardiomyopathy: the Task Force for the Diagnosis and Management of Hypertrophic Cardiomyopathy of the European Society of Cardiology (ESC). *Eur Heart J*. 2014;35(39):2733-79.
15. Hershberger RE, Hedges DJ, Morales A. Dilated cardiomyopathy: the complexity of a diverse genetic architecture. *Nature Reviews Cardiology*. 2013;10(9):531-47.
16. McNally EM, Golbus JR, Puckelwartz MJ. Genetic mutations and mechanisms in dilated cardiomyopathy. *J Clin Invest*. 2013;123(1):19-26.
17. Probst S, Oechslin E, Schuler P, Greutmann M, Boye P, Knirsch W, et al. Sarcomere gene mutations in isolated left ventricular noncompaction cardiomyopathy do not predict clinical phenotype. *Circ Cardiovasc Genet*. 2011;4(4):367-74.
18. Almeida AG, Pinto FJ. Non-compaction cardiomyopathy. *Heart*. 2013;99(20):1535-42.
19. Towbin JA, Lorts A, Jefferies JL. Left ventricular non-compaction cardiomyopathy. *Lancet*. 2015;386(9995):813-25.

20. Kohli SK, Pantazis AA, Shah JS, Adeyemi B, Jackson G, McKenna WJ, et al. Diagnosis of left-ventricular non-compaction in patients with left-ventricular systolic dysfunction: time for a reappraisal of diagnostic criteria? *European Heart Journal*. 2008;29(1):89-95.
21. Peters F, Khandheria BK, Santos Cd, Matioda H, Maharaj N, Libhaber E, et al. Isolated Left Ventricular Noncompaction in Sub-Saharan Africa: A Clinical and Echocardiographic Perspective. 2012.
22. Thavendiranathan P, Dahiya A, Phelan D, Desai MY, Tang WH. Isolated left ventricular non-compaction controversies in diagnostic criteria, adverse outcomes and management. *Heart*. 2013;99(10):681-9.
23. Andrews RE, Fenton MJ, Ridout DA, Burch M, Assoc BCC. New-onset heart failure due to heart muscle disease in childhood - A prospective study in the United Kingdom and Ireland. *Circulation*. 2008;117(1):79-84.
24. Lipshultz SE, Sleeper LA, Towbin JA, Lowe AM, Orav EJ, Cox GF, et al. The incidence of pediatric cardiomyopathy in two regions of the United States. *New England Journal of Medicine*. 2003;348(17):1647-55.
25. Nugent AW, Daubeney PEF, Chondros P, Carlin JB, Cheung M, Wilkinson LC, et al. The epidemiology of childhood cardiomyopathy in Australia. *New England Journal of Medicine*. 2003;348(17):1639-46.
26. Fatkin D, Lam L, Herman DS, Benson CC, Felkin LE, Barton PJR, et al. Titin truncating mutations: A rare cause of dilated cardiomyopathy in the young. *Progress in Pediatric Cardiology*. 2016;40:41-5.
27. Pugh TJ, Kelly MA, Gowrisankar S, Hynes E, Seidman MA, Baxter SM, et al. The landscape of genetic variation in dilated cardiomyopathy as surveyed by clinical DNA sequencing. *Genet Med*. 2014;16(8):601-8.
28. Towbin JA, Lowe AM, Colan SD, Sleeper LA, Orav EJ, Clunie S, et al. Incidence, causes, and outcomes of dilated cardiomyopathy in children. *Jama-Journal of the American Medical Association*. 2006;296(15):1867-76.
29. Gotz A, Tynismaa H, Euro L, Ellonen P, Hyotylainen T, Ojala T, et al. Exome sequencing identifies mitochondrial alanyl-tRNA synthetase mutations in infantile mitochondrial cardiomyopathy. *Am J Hum Genet*. 2011;88(5):635-42.
30. Chang KT, Taylor GP, Meschino WS, Kantor PF, Cutz E. Mitogenic cardiomyopathy: a lethal neonatal familial dilated cardiomyopathy characterized by myocyte hyperplasia and proliferation. *Hum Pathol*. 2010;41(7):1002-8.
31. Feero W, Alan E. Gutmacher, AE and Collins FS. Genomic Medicine — An Updated Primer. *NEJM*. 2010;362(21):2001-11.
32. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet*. 2010;11(1):31-46.
33. Hennekam RCM, Biesecker LG. Next-Generation Sequencing Demands Next-Generation Phenotyping. *Human Mutation*. 2012;33(5):884-6.
34. Mardis ER. Next-Generation Sequencing Platforms. *Annual Review of Analytical Chemistry*, Vol 6. 2013;6:287-303.
35. Tucker T, Marra M, Friedman JM. Massively parallel sequencing: the next big thing in genetic medicine. *Am J Hum Genet*. 2009;85(2):142-54.
36. Ware JS, Roberts AM, Cook SA. Next generation sequencing for clinical diagnostics and personalised medicine: implications for the next generation cardiologist. *Heart*. 2012;98(4):276-81.
37. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*. 2009;461(7261):272-6.
38. Gilissen C, Hoischen A, Brunner HG, Veltman JA. Disease gene identification strategies for exome sequencing. *Eur J Hum Genet*. 2012;20(5):490-7.
39. Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, et al. Molecular Findings Among Patients Referred for Clinical Whole-Exome Sequencing. *JAMA*. 2014.
40. Biesecker LG, Green RC. Diagnostic clinical genome and exome sequencing. *N Engl J Med*. 2014;370(25):2418-25.

41. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med*. 2013;15(7):565-74.
42. Cooper GM, Shendure J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat Rev Genet*. 2011;12(9):628-40.
43. Vissers LE, Veltman JA. Standardized phenotyping enhances Mendelian disease gene identification. *Nat Genet*. 2015;47(11):1222-4.
44. MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, et al. Guidelines for investigating causality of sequence variants in human disease. *Nature*. 2014;508(7497):469-76.
45. Lu JT, Campeau PM, Lee BH. Genotype-phenotype correlation--promiscuity in the era of next-generation sequencing. *N Engl J Med*. 2014;371(7):593-6.
46. Flannick J, Beer NL, Bick AG, Agarwala V, Molnes J, Gupta N, et al. Assessing the phenotypic effects in the general population of rare variants in genes for a dominant Mendelian form of diabetes. *Nat Genet*. 2013.
47. 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*. 2010;26(16):2069-70.
48. Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, et al. Mapping and sequencing of structural variation from eight human genomes. *Nature*. 2008;453(7191):56-64.
49. Tabor HK, Auer PL, Jamal SM, Chong JX, Yu JH, Gordon AS, et al. Pathogenic variants for Mendelian and complex traits in exomes of 6,517 European and African Americans: implications for the return of incidental results. *Am J Hum Genet*. 2014;95(2):183-93.
50. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.
51. Exome Aggregation Consortium (ExAC) [Internet]. [cited 1st November 2015].
52. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14(2):178-92.
53. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29(1):24-6.
54. Duzkale H, Shen J, McLaughlin H, Alfares A, Kelly MA, Pugh TJ, et al. A systematic approach to assessing the clinical significance of genetic variants. *Clin Genet*. 2013;84(5):453-63.
55. Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, et al. A framework for the interpretation of de novo mutation in human disease. *Nat Genet*. 2014;46(9):944-50.
56. Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet*. 2013;9(8):e1003709.
57. Huang N, Wellcome Trust Sanger Institute WTGC, Cambridge, United Kingdom, Lee I, Center for Systems and Synthetic Biology DoCaB, Institute for Cellular and Molecular Biology, University of Texas, Austin, Texas, United States of America, Department of Biotechnology CoLSaB, Yonsei University, Seoul, South Korea, Marcotte EM, et al. Characterising and Predicting Haploinsufficiency in the Human Genome. *PLOS Genetics*. 2010;6(10).
58. Antonarakis SE. CpG Dinucleotides and Human Disorders. 2006.
59. Benit P, El-Khoury R, Schiff M, Sainsard-Chanet A, Rustin P. Genetic background influences mitochondrial function: modeling mitochondrial disease for therapeutic development. *Trends Mol Med*. 2010;16(5):210-7.
60. Wang J, Schmitt ES, Landsverk ML, Zhang VW, Li FY, Graham BH, et al. An integrated approach for classifying mitochondrial DNA variants: one clinical diagnostic laboratory's experience. *Genet Med*. 2012;14(6):620-6.
61. Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature*. 2005;434(7031):400-4.
62. Puck JM, Willard HF. X inactivation in females with X-linked disease. *N Engl J Med*. 1998;338(5):325-8.
63. Morleo M, Franco B. Dosage compensation of the mammalian X chromosome influences the phenotypic variability of X-linked dominant male-lethal disorders. *J Med Genet*. 2008;45(7):401-8.

64. Dobyns WB, Filauro A, Tomson BN, Chan AS, Ho AW, Ting NT, et al. Inheritance of most X-linked traits is not dominant or recessive, just X-linked. *American Journal of Medical Genetics Part A*. 2004;129a(2):136-43.
65. chromosome SoX. Sequencing of X chromosome. *Nature*. 2005;434.
66. Lyon MF. X-chromosome inactivation: a repeat hypothesis. *Cytogenet Cell Genet*. 1998;80(1-4):133-7.
67. van Rahden VA, Rau I, Fuchs S, Kosyna FK, de Almeida HL, Jr., Fryssira H, et al. Clinical spectrum of females with HCCS mutation: from no clinical signs to a neonatal lethal form of the microphthalmia with linear skin defects (MLS) syndrome. *Orphanet J Rare Dis*. 2014;9:53.
68. Plenge RM, Stevenson RA, Lubs HA, Schwartz CE, Willard HF. Skewed X-chromosome inactivation is a common feature of X-linked mental retardation disorders. *American Journal of Human Genetics*. 2002;71(1):168-73.
69. Belmont JW. Insights into lymphocyte development from X-linked immune deficiencies. *Trends Genet*. 1995;11(3):112-6.
70. Koopman WJ, Willems PH, Smeitink JA. Monogenic mitochondrial disorders. *N Engl J Med*. 2012;366(12):1132-41.
71. Schon EA, DiMauro S, Hirano M. Human mitochondrial DNA: roles of inherited and somatic mutations. *Nat Rev Genet*. 2012;13(12):878-90.
72. Opdal SH. Mitochondrial DNA and sudden infant death syndrome. *Acta Paediatr*. 2014;103(7):685-6.
73. Kujoth GC, Bradshaw PC, Haroon S, Prolla TA. The role of mitochondrial DNA mutations in mammalian aging. *PLoS Genet*. 2007;3(2):e24.
74. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. *Nature*. 1981;290(5806):457-65.
75. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet*. 1999;23(2):147.
76. Bandelt HJ, Kloss-Brandstatter A, Richards MB, Yao YG, Logan I. The case for the continuing use of the revised Cambridge Reference Sequence (rCRS) and the standardization of notation in human mitochondrial DNA studies. *J Hum Genet*. 2014;59(2):66-77.
77. Vafai SB, Mootha VK. Mitochondrial disorders as windows into an ancient organelle. *Nature*. 2012;491(7424):374-83.
78. Vallance HD, Jevon G, Wallace DC, Brown MD. A case of sporadic infantile histiocytoid cardiomyopathy caused by the A8344G (MERRF) mitochondrial DNA mutation. *Pediatr Cardiol*. 2004;25(5):538-40.
79. Stewart JB, Chinnery PF. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nat Rev Genet*. 2015;16(9):530-42.
80. Gorman GS, Schaefer AM, Ng Y, Gomez N, Blakely EL, Alston CL, et al. Prevalence of nuclear and mitochondrial DNA mutations related to adult mitochondrial disease. *Ann Neurol*. 2015;77(5):753-9.
81. Bates MG, Bourke JP, Giordano C, d'Amati G, Turnbull DM, Taylor RW. Cardiac involvement in mitochondrial DNA disease: clinical spectrum, diagnosis, and management. *Eur Heart J*. 2012;33(24):3023-33.
82. Rotig A, Munnich A. Genetic features of mitochondrial respiratory chain disorders. *J Am Soc Nephrol*. 2003;14(12):2995-3007.
83. Chinnery PF, Johnson MA, Wardell TM, Singh-Kler R, Hayes C, Brown DT, et al. The epidemiology of pathogenic mitochondrial DNA mutations. *Ann Neurol*. 2000;48(2):188-93.
84. Griffin HR, Pyle A, Blakely EL, Alston CL, Duff J, Hudson G, et al. Accurate mitochondrial DNA sequencing using off-target reads provides a single test to identify pathogenic point mutations. *Genet Med*. 2014;16(12):962-71.
85. Fassone E, Rahman S. Complex I deficiency: clinical features, biochemistry and molecular genetics. *J Med Genet*. 2012;49(9):578-90.
86. Calvo SE, Clauser KR, Mootha VK. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res*. 2016;44(D1):D1251-7.

87. Wallace DC, Singh G, Lott MT, Hodge JA, Schurr TG, Lezza AM, et al. Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. *Science*. 1988;242(4884):1427-30.
88. Holt IJ, Harding AE, Morgan-Hughes JA. Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. *Nature*. 1988;331(6158):717-9.
89. Bourgeron T, Rustin P, Chretien D, Birch-Machin M, Bourgeois M, Viegas-Pequignot E, et al. Mutation of a nuclear succinate dehydrogenase gene results in mitochondrial respiratory chain deficiency. *Nat Genet*. 1995;11(2):144-9.
90. Finsterer J, Kothari S. Cardiac manifestations of primary mitochondrial disorders. *Int J Cardiol*. 2014;177(3):754-63.
91. Yapfilito-Lee J, Weintraub R, Jamsen K, Chow CW, Thorburn DR, Boneh A. Cardiac manifestations in oxidative phosphorylation disorders of childhood. *Journal of Pediatrics*. 2007;150(4):407-11.
92. Haack TB, Danhauser K, Haberberger B, Hoser J, Strecker V, Boehm D, et al. Exome sequencing identifies ACAD9 mutations as a cause of complex I deficiency. *Nat Genet*. 2010;42(12):1131-4.
93. Brunner HG. The Variability of Genetic Disease. *N Engl J Med*. 2012;367(14):1349-50.
94. Baird PA, Anderson TW, Newcombe HB, Lowry RB. Genetic-Disorders in Children and Young-Adults - a Population Study. *American Journal of Human Genetics*. 1988;42(5):677-93.
95. Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobreira N, Smith JD, et al. The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *American Journal of Human Genetics*. 2015;97(2):199-215.
96. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet*. 2011;12(11):745-55.
97. Wright CF, Fitzgerald TW, Jones WD, Clayton S, Mcrae JF, van Kogelenberg M, et al. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet*. 2015;385(9975):1305-14.
98. Bainbridge MN, Hu H, Muzny DM, Musante L, Lupski JR, Graham BH, et al. De novo truncating mutations in ASXL3 are associated with a novel clinical phenotype with similarities to Bohring-Opitz syndrome. *Genome Medicine*. 2013;5.
99. Gilissen C, Hoischen A, Brunner HG, Veltman JA. Unlocking Mendelian disease using exome sequencing. *Genome Biol*. 2011;12(9):228.
100. Vissers LELM, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, et al. A de novo paradigm for mental retardation. *Nature Genetics*. 2010;42(12):1109-+.
101. Nachman MW, Crowell SL. Estimate of the mutation rate per nucleotide in humans. *Genetics*. 2000;156(1):297-304.
102. Conrad DF, Keebler JEM, DePristo MA, Lindsay SJ, Zhang YJ, Casals F, et al. Variation in genome-wide mutation rates within and between human families. *Nature Genetics*. 2011;43(7):712-U137.
103. Lynch M. Rate, molecular spectrum, and consequences of human mutation. *Proc Natl Acad Sci U S A*. 2010;107(3):961-8.
104. Siu BL, Niimura H, Osborne JA, Fatkin D, MacRae C, Solomon S, et al. Familial Dilated Cardiomyopathy Locus Maps to Chromosome 2q31. *Circulation*. 1999;99(8):1022-6.
105. Gerull B, Atherton J, Geupel A, Sasse-Klaassen S, Heuser A, Frenneaux M, et al. Identification of a novel frameshift mutation in the giant muscle filament titin in a large Australian family with dilated cardiomyopathy. *J Mol Med (Berl)*. 2006;84(6):478-83.
106. Gerull B, Gramlich M, Atherton J, McNabb M, Trombitas K, Sasse-Klaassen S, et al. Mutations of TTN, encoding the giant muscle filament titin, cause familial dilated cardiomyopathy. *Nature Genetics*. 2002;30(2):201-4.
107. Herman DS, Lam L, Taylor MR, Wang L, Teekakirikul P, Christodoulou D, et al. Truncations of titin causing dilated cardiomyopathy. *N Engl J Med*. 2012;366(7):619-28.
108. Taylor M, Graw S, Sinagra G, Barnes C, Slavov D, Brun F, et al. Genetic variation in titin in arrhythmogenic right ventricular cardiomyopathy-overlap syndromes. *Circulation*. 2011;124(8):876-85.

109. Hackman P, Richard I, Vihola A, Haravuori H, Marchand S, Labeit S, et al. Tibial muscular dystrophy (TMD), 2q31 linked myopathy: Sequencing and functional studies of the titin (TTN) gene. *Journal of the Neurological Sciences*. 2002;199:S35-S.
110. Norton N, Li D, Rieder MJ, Siegfried JD, Rampersaud E, Züchner S, et al. Genome-wide studies of copy number variation and exome sequencing identify rare variants in BAG3 as a cause of dilated cardiomyopathy. *Am J Hum Genet*. 2011;88(3):273-82.
111. McNally EM. Genetics: broken giant linked to heart failure. *Nature*. 2012;483(7389):281-2.
112. Trinick J, Knight P, Whiting A. Purification and properties of native titin. *J Mol Biol*. 1984;180(2):331-56.
113. LeWinter MM, Granzier HL. Titin is a major human disease gene. *Circulation*. 2013;127(8):938-44.
114. Tskhovrebova L, Trinick J, Sleep JA, Simmons RM. Elasticity and unfolding of single molecules of the giant muscle protein titin. *Nature*. 1997;387(6630):308-12.
115. Watkins H. Tackling the achilles' heel of genetic testing. *Sci Transl Med*. 2015;7(270):270fs1.
116. Hidalgo C, Granzier H. Tuning the molecular giant titin through phosphorylation: role in health and disease. *Trends Cardiovasc Med*. 2013;23(5):165-71.
117. Roberts AM, Ware JS, Herman DS, Schafer S, Baksi J, Bick AG, et al. Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease. *Sci Transl Med*. 2015;7(270):270ra6.
118. Roberts AM. Titin: an analysis of genetic variation and cardiac phenotype [PhD Thesis]: Imperial College London; 2015.
119. Ware JS, Li J, Mazaika E, Yasso CM, DeSouza T, Cappola TP, et al. Shared Genetic Predisposition in Peripartum and Dilated Cardiomyopathies. *N Engl J Med*. 2016.
120. Golbus JR, Puckelwartz MJ, Fahrenbach JP, Dellefave-Castillo LM, Wolfgeher D, McNally EM. Population-based variation in cardiomyopathy genes. *Circ Cardiovasc Genet*. 2012;5(4):391-9.
121. Roncarati R, Viviani Anselmi C, Krawitz P, Lattanzi G, von Kodolitsch Y, Perrot A, et al. Doubly heterozygous LMNA and TTN mutations revealed by exome sequencing in a severe form of dilated cardiomyopathy. *Eur J Hum Genet*. 2013;21(10):1105-11.
122. Mann DL, Barger PM, Burkoff D. Myocardial recovery and the failing heart: myth, magic, or molecular target? *J Am Coll Cardiol*. 2012;60(24):2465-72.
123. Felkin LE, Walsh R, Ware JS, Yacoub MH, Birks EJ, Barton PJR, et al. Recovery of Cardiac Function in Cardiomyopathy Caused by Titin Truncation. *JAMA Cardiology*. 2016;1(2):234.
124. Ohlsson M, Hedberg C, Bradvik B, Lindberg C, Tajsharghi H, Danielsson O, et al. Hereditary myopathy with early respiratory failure associated with a mutation in A-band titin. *Brain*. 2012;135(Pt 6):1682-94.
125. Hinson JT, Chopra A, Nafissi N, Polacheck WJ, Benson CC, Swist S, et al. HEART DISEASE. Titin mutations in iPS cells define sarcomere insufficiency as a cause of dilated cardiomyopathy. *Science*. 2015;349(6251):982-6.
126. Chauveau C, Rowell J, Ferreira A. A Rising Titan: TTN Review and Mutation Update. *Human Mutation*. 2014;35(9):1046-59.
127. Begay RL, Graw S, Sinagra G, Merlo M, Slavov D, Gowan K, et al. Role of Titin Missense Variants in Dilated Cardiomyopathy. *J Am Heart Assoc*. 2015;4(11).
128. Towbin JA. Inherited Cardiomyopathies. *Circulation Journal*. 2014;78(10):2347-56.
129. Walsh R, Thomson KL, Ware JS, Funke BH, Woodley J, McGuire KJ, et al. Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples. *Genet Med*. 2016.
130. A HRM. Dilated Cardiomyopathy Overview Seattle (WA): University of Washington, Seattle.: Seattle (WA): University of Washington, Seattle.; 2007 Jul 27 [Updated 2015 Sep 24]. [Available from: Available from: <http://www.ncbi.nlm.nih.gov/books/NBK1309/>.
131. Kwong RY, Korlakunta H. Diagnostic and prognostic value of cardiac magnetic resonance imaging in assessing myocardial viability. *Top Magn Reson Imaging*. 2008;19(1):15-24.
132. Schinkel AF, Poldermans D, Elhendy A, Bax JJ. Assessment of myocardial viability in patients with heart failure. *J Nucl Med*. 2007;48(7):1135-46.
133. Kim RJ, Albert TS, Wible JH, Elliott MD, Allen JC, Lee JC, et al. Performance of delayed-enhancement magnetic resonance imaging with gadoversetamide contrast for the detection and

- assessment of myocardial infarction: an international, multicenter, double-blinded, randomized trial. *Circulation*. 2008;117(5):629-37.
134. Amado LC, Gerber BL, Gupta SN, Rettmann DW, Szarf G, Schock R, et al. Accurate and objective infarct sizing by contrast-enhanced magnetic resonance imaging in a canine myocardial infarction model. *J Am Coll Cardiol*. 2004;44(12):2383-9.
 135. Mollet NR, Dymarkowski S, Volders W, Wathiong J, Herbots L, Rademakers FE, et al. Visualization of ventricular thrombi with contrast-enhanced magnetic resonance imaging in patients with ischemic heart disease. *Circulation*. 2002;106(23):2873-6.
 136. Srichai MB, Junor C, Rodriguez LL, Stillman AE, Grimm RA, Lieber ML, et al. Clinical, imaging, and pathological characteristics of left ventricular thrombus: a comparison of contrast-enhanced magnetic resonance imaging, transthoracic echocardiography, and transesophageal echocardiography with surgical or pathological validation. *Am Heart J*. 2006;152(1):75-84.
 137. Becker S, Walter S, Witzke O, Kreuter A, Kribben A, Mitchell A. Application of gadolinium-based contrast agents and prevalence of nephrogenic systemic fibrosis in a cohort of end-stage renal disease patients on hemodialysis. *Nephron Clin Pract*. 2012;121(1-2):c91-4.
 138. Thygesen K, Alpert JS, White HD, Joint ESCAAHAWHFTFtRoMI. Universal definition of myocardial infarction. *Eur Heart J*. 2007;28(20):2525-38.
 139. Alpert JS, Thygesen K, Antman E, Bassand JP. Myocardial infarction redefined--a consensus document of The Joint European Society of Cardiology/American College of Cardiology Committee for the redefinition of myocardial infarction. *J Am Coll Cardiol*. 2000;36(3):959-69.
 140. Pasotti M, Prati F, Arbustini E. The pathology of myocardial infarction in the pre- and post-interventional era. *Heart*. 2006;92(11):1552-6.
 141. Thygesen K, Alpert JS, Jaffe AS, Simoons ML, Chaitman BR, White HD, et al. Third universal definition of myocardial infarction. *Circulation*. 2012;126(16):2020-35.
 142. Deedwania PC, Carbajal EV. Silent myocardial ischemia. A clinical perspective. *Arch Intern Med*. 1991;151(12):2373-82.
 143. Ambale-Venkatesh B, Lima JA. Cardiac MRI: a central prognostic tool in myocardial fibrosis. *Nat Rev Cardiol*. 2015;12(1):18-29.
 144. McCrohon JA, Moon JC, Prasad SK, McKenna WJ, Lorenz CH, Coats AJ, et al. Differentiation of heart failure related to dilated cardiomyopathy and coronary artery disease using gadolinium-enhanced cardiovascular magnetic resonance. *Circulation*. 2003;108(1):54-9.
 145. Soriano CJ, Ridocci F, Estornell J, Jimenez J, Martinez V, De Velasco JA. Noninvasive diagnosis of coronary artery disease in patients with heart failure and systolic dysfunction of uncertain etiology, using late gadolinium-enhanced cardiovascular magnetic resonance. *J Am Coll Cardiol*. 2005;45(5):743-8.
 146. Cappola TP. Molecular remodeling in human heart failure. *J Am Coll Cardiol*. 2008;51(2):137-8.
 147. Kim RJ, Fieno DS, Parrish TB, Harris K, Chen EL, Simonetti O, et al. Relationship of MRI delayed contrast enhancement to irreversible injury, infarct age, and contractile function. *Circulation*. 1999;100(19):1992-2002.
 148. Fieno DS, Kim RJ, Chen EL, Lomasney JW, Klocke FJ, Judd RM. Contrast-enhanced magnetic resonance imaging of myocardium at risk: distinction between reversible and irreversible injury throughout infarct healing. *J Am Coll Cardiol*. 2000;36(6):1985-91.
 149. Vanderheyden M, Mullens W, Delrue L, Goethals M, de Bruyne B, Wijns W, et al. Myocardial gene expression in heart failure patients treated with cardiac resynchronization therapy. *Journal of the American College of Cardiology*. 2008;51(2):129-36.
 150. Zipes DP, Camm AJ, Borggrefe M, Buxton AE, Chaitman B, Fromer M, et al. ACC/AHA/ESC 2006 Guidelines for Management of Patients With Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death: a report of the American College of Cardiology/American Heart Association Task Force and the European Society of Cardiology Committee for Practice Guidelines (writing committee to develop Guidelines for Management of Patients With Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death): developed in collaboration with the European Heart Rhythm Association and the Heart Rhythm Society. *Circulation*. 2006;114(10):e385-484.
 151. Buckley U, Shivkumar K. Implantable cardioverter defibrillators: even better than we thought? *Eur Heart J*. 2015;36(26):1646-8.

152. Raphael CE, Finegold JA, Barron AJ, Whinnett ZI, Mayet J, Linde C, et al. The effect of duration of follow-up and presence of competing risk on lifespan-gain from implantable cardioverter defibrillator therapy: who benefits the most? *Eur Heart J*. 2015;36(26):1676-88.
153. Gajarsa JJ, Kloner RA. Left ventricular remodeling in the post-infarction heart: a review of cellular, molecular mechanisms, and therapeutic modalities. *Heart Fail Rev*. 2011;16(1):13-21.
154. Cohn JN, Ferrari R, Sharpe N. Cardiac remodeling--concepts and clinical implications: a consensus paper from an international forum on cardiac remodeling. Behalf of an International Forum on Cardiac Remodeling. *J Am Coll Cardiol*. 2000;35(3):569-82.
155. Dagres N, Hindricks G. Risk stratification after myocardial infarction: is left ventricular ejection fraction enough to prevent sudden cardiac death? *Eur Heart J*. 2013;34(26):1964-71.
156. Grothues F, Smith GC, Moon JC, Bellenger NG, Collins P, Klein HU, 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. *Am J Cardiol*. 2002;90(1):29-34.
157. Gulati A, Jabbour A, Ismail TF, Guha K, Khwaja J, Raza S, et al. Association of fibrosis with mortality and sudden cardiac death in patients with nonischemic dilated cardiomyopathy. *JAMA*. 2013;309(9):896-908.
158. Thiele H, Kappl MJ, Conradi S, Niebauer J, Hambrecht R, Schuler G. Reproducibility of chronic and acute infarct size measurement by delayed enhancement-magnetic resonance imaging. *J Am Coll Cardiol*. 2006;47(8):1641-5.
159. Cerqueira MD, Weissman NJ, Dilsizian V, Jacobs AK, Kaul S, Laskey WK, et al. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart - A statement for healthcare professionals from the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association. *Circulation*. 2002;105(4):539-42.
160. den Dunnen JT, Antonarakis SE. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat*. 2000;15(1):7-12.
161. Rios D, McLaren WM, Chen Y, Birney E, Stabenau A, Flicek P, et al. A database and API for variation, dense genotyping and resequencing data. *BMC Bioinformatics*. 2010;11:238.
162. Ajay SS, Parker SC, Abaan HO, Fajardo KV, Margulies EH. Accurate and comprehensive sequencing of personal genomes. *Genome Res*. 2011;21(9):1498-505.
163. Banner NR, Bonser RS, Clark AL, Clark S, Cowburn PJ, Gardner RS, et al. UK guidelines for referral and assessment of adults for heart transplantation. *Heart*. 2011;97(18):1520-7.
164. Lindhurst MJ, Sapp JC, Teer JK, Johnston JJ, Finn EM, Peters K, et al. A Mosaic Activating Mutation in AKT1 Associated with the Proteus Syndrome. *New England Journal of Medicine*. 2011;365(7):611-9.
165. Hennekam RC. Care for patients with ultra-rare disorders. *Eur J Med Genet*. 2011;54(3):220-4.
166. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nature Genetics*. 2010;42(9):790-U85.
167. Louw JJ, Corveleyn A, Jia Y, Iqbal S, Boshoff D, Gewillig M, et al. Homozygous loss-of-function mutation in ALMS1 causes the lethal disorder mitogenic cardiomyopathy in two siblings. *Eur J Med Genet*. 2014;57(9):532-5.
168. Voth D. [On arachnocytopsis of the myocardium (A contribution to the problem of rhabdomyoma of the heart)]. *Frankf Z Pathol*. 1962;71:646-56.
169. Ferrans VJ, McAllister HA, Jr., Haese WH. Infantile cardiomyopathy with histiocytoid change in cardiac muscle cells. Report of six patients. *Circ*. 1976;53(4):708-19.
170. Silver MM, Burns JE, Sethi RK, Rowe RD. Oncocytic cardiomyopathy in an infant with oncocytosis in exocrine and endocrine glands. *Hum Pathol*. 1980;11(6):598-605.
171. Gelb AB, Van Meter SH, Billingham ME, Berry GJ, Rouse RV. Infantile histiocytoid cardiomyopathy—Myocardial or conduction system hamartoma: What is the cell type involved? *Human Pathology*. 1993;24(11):1226-31.
172. Radford DJ, Chalk SM. Infantile xanthomatous cardiomyopathy. *Aust Paediatr J*. 1980;16(2):123-5.
173. MacMahon HE. Infantile xanthomatous cardiomyopathy. *Pediatrics*. 1971;48(2):312-5.

174. Bove KE, Schwartz DC. Focal lipid cardiomyopathy in an infant with paroxysmal atrial tachycardia. *Arch Pathol.* 1973;95(1):26-36.
175. Ross CF, Belton EM. A case of isolated cardiac lipidosi. *Br Heart J.* 1968;30(5):726-8.
176. Gelb AB, Vanmeter SH, Billingham ME, Rouse RV. Is Infantile Cardiomyopathy with Histiocytoid Change Really Purkinje-Cell Hyperplasia. *Laboratory Investigation.* 1993;68(1):A23-A.
177. Suarez V, Fuggle WJ, Cameron AH, French TA, Hollingworth T. Foamy myocardial transformation of infancy: an inherited disease. *J Clin Pathol.* 1987;40(3):329-34.
178. Haese WH, Maron BJ, Mirowski M, Rowe RD, Hutchins GM. Peculiar focal myocardial degeneration and fatal ventricular arrhythmias in a child. *N Engl J Med.* 1972;287(4):180-1.
179. Kauffman SL, Chandra N, Peress NS, Rodriguez-Torres R. Idiopathic infantile cardiomyopathy with involvement of the conduction system. *Am J Cardiol.* 1972;30(6):648-52.
180. Malhotra V, Ferrans VJ, Virmani R. Infantile histiocytoid cardiomyopathy: three cases and literature review. *Am Heart J.* 1994;128(5):1009-21.
181. Jain D, Maleszewski JJ, Halushka MK. Benign cardiac tumors and tumorlike conditions. *Ann Diagn Pathol.* 2010;14(3):215-30.
182. Zimmermann A, Diem P, Cottier H. Congenital "histiocytoid" cardiomyopathy: evidence suggesting a developmental disorder of the Purkinje cell system of the heart. *Virchows Arch A Pathol Anat Histol.* 1982;396(2):187-95.
183. Tsuruda T, Hatakeyama K, Nagamachi S, Sekita Y, Sakamoto S, Endo GJ, et al. Inhibition of development of abdominal aortic aneurysm by glycolysis restriction. *Arterioscler Thromb Vasc Biol.* 2012;32(6):1410-7.
184. Ashworth M. Cardiomyopathy in Childhood: Histopathological and Genetic Features. *The Open Pathology Journal.* 2010;4:80-93.
185. Cabana MD, Becher O, Smith A. Histiocytoid cardiomyopathy presenting with Wolff-Parkinson-White syndrome. *Heart.* 2000;83(1):98-9.
186. Finsterer J, Stollberger C. Is mitochondrial disease the common cause of histiocytoid cardiomyopathy and non-compaction? *Int J Legal Med.* 2009;123(6):507-8.
187. Siehr SL, Bernstein D, Yeh J, Berry GJ, Rosenthal DN, Hollander SA. Orthotopic heart transplantation in two infants with histiocytoid cardiomyopathy and left ventricular non-compaction. *Pediatr Transplant.* 2013;17(7):E165-7.
188. Gilbert-Barness E, Barness LA. Festschrift for Dr. John M. Opitz: Pathogenesis of cardiac conduction disorders in children genetic and histopathologic aspects. *American Journal of Medical Genetics Part A.* 2006;140A(19):1993-2006.
189. Amini M, Bosman C, Marino B. Histiocytoid cardiomyopathy in infancy: a new hypothesis? *Chest.* 1980;77(4):556-8.
190. Shehata BM, Patterson K, Thomas JE, Scala-Barnett D, Dasu S, Robinson HB. Histiocytoid cardiomyopathy: three new cases and a review of the literature. *Pediatr Dev Pathol.* 1998;1(1):56-69.
191. Coulibaly B, Piercecchi-Marti MD, Fernandez C, Wasier AP, Viard L, Fraisse A, et al. [A rare cause of sudden cardiac failure: histiocytoid cardiomyopathy]. *Ann Pathol.* 2011;31(2):93-7.
192. Bruton D, Herdson PB, Becroft DMO. Histiocytoid Cardiomyopathy of Infancy - Unexplained Myofiber Degeneration. *Pathology.* 1977;9(2):115-22.
193. Ilina MV, Kepron CA, Taylor GP, Perrin DG, Kantor PF, Somers GR. Undiagnosed heart disease leading to sudden unexpected death in childhood: a retrospective study. *Pediatrics.* 2011;128(3):e513-20.
194. Cunningham NE, Stewart J. A rare cause cot death--infantile xanthomatous cardiomyopathy. *Med Sci Law.* 1985;25(2):149-52.
195. Shehata BM, Bouzyk M, Shulman SC, Tang W, Steelman CK, Davis GK, et al. Identification of candidate genes for histiocytoid cardiomyopathy (HC) using whole genome expression analysis: analyzing material from the HC registry. *Pediatr Dev Pathol.* 2011;14(5):370-7.
196. Kearney DL, Titus JL, Hawkins EP, Ott DA, Garson A, Jr. Pathologic features of myocardial hamartomas causing childhood tachyarrhythmias. *Circulation.* 1987;75(4):705-10.
197. Burke A, Mont E, Kutys R, Virmani R. Left ventricular noncompaction: a pathological study of 14 cases. *Hum Pathol.* 2005;36(4):403-11.
198. Grech V, Ellul B, Montalto SA. Sudden cardiac death in infancy due to histiocytoid cardiomyopathy. *Cardiol Young.* 2000;10(1):49-51.

199. Witzleben CL, Pinto M. Foamy Myocardial Transformation of Infancy - Lipid or Histiocytoid Myocardiopathy. *Archives of Pathology & Laboratory Medicine*. 1978;102(6):306-11.
200. Zangwill SD, Trost BA, Zlotocha J, Tweddell JS, Jaquiss RD, Berger S. Orthotopic heart transplantation in a child with histiocytoid cardiomyopathy. *Journal of Heart and Lung Transplantation*. 2004;23(7):902-4.
201. Franciosi RA, Singh A. Oncocytic cardiomyopathy syndrome. *Hum Pathol*. 1988;19(11):1361-2.
202. Deacon JS, Gilbert EF, Viseskul C, Herrmann J, Angevine JM, Opitz JM, et al. Familial cardiac lipodosis. *Birth Defects Orig Artic Ser*. 1974;10(8):181-95.
203. Bird LM, Krous HF, Eichenfield LF, Swalwell CI, Jones MC. Female infant with oncocytic cardiomyopathy and microphthalmia with linear skin defects (MLS): a clue to the pathogenesis of oncocytic cardiomyopathy? *Am J Med Genet*. 1994;53(2):141-8.
204. Shehata BM, Cundiff CA, Lee K, Sabharwal A, Lalwani MK, Davis AK, et al. Exome sequencing of patients with histiocytoid cardiomyopathy reveals a de novo NDUFB11 mutation that plays a role in the pathogenesis of histiocytoid cardiomyopathy. *Am J Med Genet A*. 2015;167A(9):2114-21.
205. Finsterer J. Histiocytoid cardiomyopathy: a mitochondrial disorder. *Clin Cardiol*. 2008;31(5):225-7.
206. Papadimitriou A, Neustein HB, Dimauro S, Stanton R, Bresolin N. Histiocytoid cardiomyopathy of infancy: deficiency of reducible cytochrome b in heart mitochondria. *Pediatr Res*. 1984;18(10):1023-8.
207. Andreu AL, Checcarelli N, Iwata S, Shanske S, DiMauro S. A missense mutation in the mitochondrial cytochrome b gene in a revisited case with histiocytoid cardiomyopathy. *Pediatr Res*. 2000;48(3):311-4.
208. Cataldo S, Annoni GA, Marziliano N. The perfect storm? Histiocytoid cardiomyopathy and compound CACNA2D1 and RANGRF mutation in a baby. *Cardiol Young*. 2015;25(1):174-6.
209. Risgaard B, Jabbari R, Refsgaard L, Holst AG, Haunso S, Sadjadieh A, et al. High prevalence of genetic variants previously associated with Brugada syndrome in new exome data. *Clin Genet*. 2013;84(5):489-95.
210. Bagnall RD, Das KJ, Duflou J, Semsarian C. Exome analysis-based molecular autopsy in cases of sudden unexplained death in the young. *Heart Rhythm*. 2014;11(4):655-62.
211. Vergult S, Dheedene A, Meurs A, Faes F, Isidor B, Janssens S, et al. Genomic aberrations of the CACNA2D1 gene in three patients with epilepsy and intellectual disability. *Eur J Hum Genet*. 2015;23(5):628-32.
212. Campuzano O, Berne P, Selga E, Allegue C, Iglesias A, Brugada J, et al. Brugada syndrome and p.E61X_RANGRF. *Cardiol J*. 2014;21(2):121-7.
213. Baillie T, Chan YF, Koelmeyer TD, Cluroe AD. Test and teach. Ill-defined subendocardial nodules in an infant. Histiocytoid cardiomyopathy. *Pathology*. 2001;33(2):230-4.
214. Hoischen A, van Bon BW, Rodriguez-Santiago B, Gilissen C, Vissers LE, de Vries P, et al. De novo nonsense mutations in ASXL1 cause Bohring-Opitz syndrome. *Nat Genet*. 2011;43(8):729-31.
215. Vallance HD, Jevon G, Brown MD. A case of sporadic infantile histiocytoid cardiomyopathy caused by the 8344 (MERRF) mtDNA mutation. *Journal of Inherited Metabolic Disease*. 2000;23(Supplement 1):146-.
216. Rea G, Homfray T, Till J, Roses-Noguer F, Buchan RJ, Wilkinson S, et al. Histiocytoid cardiomyopathy and microphthalmia with linear skin defects syndrome: phenotypes linked by truncating variants in NDUFB11. *Cold Spring Harb Mol Case Stud*. 2017;3(1):a001271.
217. Gelberg HB. Purkinje fiber dysplasia (histiocytoid cardiomyopathy) with ventricular noncompaction in a savannah kitten. *Vet Pathol*. 2009;46(4):693-7.
218. Edston E, Perskvist N. Histiocytoid cardiomyopathy and ventricular non-compaction in a case of sudden death in a female infant. *Int J Legal Med*. 2009;123(1):47-53.
219. van Rahden VA, Fernandez-Vizarra E, Alawi M, Brand K, Fellmann F, Horn D, et al. Mutations in NDUFB11, encoding a complex I component of the mitochondrial respiratory chain, cause microphthalmia with linear skin defects syndrome. *Am J Hum Genet*. 2015;96(4):640-50.

220. Gurok U, Bork K, Nuber U, Sporle R, Nohring S, Horstkorte R. Expression of Ndufb11 encoding the neuronal protein 15.6 during neurite outgrowth and development. *Gene Expr Patterns*. 2007;7(3):370-4.
221. Al-Gazali LI MR, Caine A, Dennis N, Antoniou A, Fitchett M, Insley J, Goodfellow PG, & Hulten M. An XX male and two t (X;Y) females with linear skin defects and congenital microphthalmia: a new syndrome at Xp22.3. *J Med Genet*. 1988;25(9):638-9.
222. al-Gazali LI, Mueller RF, Caine A, Antoniou A, McCartney A, Fitchett M, et al. Two 46,XX,t(X;Y) females with linear skin defects and congenital microphthalmia: a new syndrome at Xp22.3. *J Med Genet*. 1990;27(1):59-63.
223. McLeod SD, Sugar J, Elejalde BR, Eng A, Lebel RR. Gazali-Temple syndrome. *Arch Ophthalmol*. 1994;112(6):851-2.
224. Temple IK, Hurst JA, Hing S, Butler L, Baraitser M. De novo deletion of Xp22.2-pter in a female with linear skin lesions of the face and neck, microphthalmia, and anterior chamber eye anomalies. *J Med Genet*. 1990;27(1):56-8.
225. Happle R, Daniels O, Koopman RJ. MIDAS syndrome (microphthalmia, dermal aplasia, and sclerocornea): an X-linked phenotype distinct from Goltz syndrome. *Am J Med Genet*. 1993;47(5):710-3.
226. Robert F. Stratton CAW, Brent R. Paulgar, Mary E. Price, and Charleen M. Moore. Second 46,XX Male With MLS Syndrome. *American Journal of Medical Genetics*. 1998;76:37-41.
227. Morleo M, Pramparo T, Perone L, Gregato G, Le Caignec C, Mueller RF, et al. Microphthalmia with linear skin defects (MLS) syndrome: clinical, cytogenetic, and molecular characterization of 11 cases. *Am J Med Genet A*. 2005;137(2):190-8.
228. Lindsay EA, Grillo A, Ferrero GB, Roth EJ, Magenis E, Grompe M, et al. Microphthalmia with linear skin defects (MLS) syndrome: clinical, cytogenetic, and molecular characterization. *Am J Med Genet*. 1994;49(2):229-34.
229. Schaefer L, Wapenaar MC, Bassi MT, Ferrero GB, Grillo A, Roth RJ, et al. Characterization and Cloning of the Critical Region for the Microphthalmia with Linear Skin Defects Syndrome (Mls). *Journal of Cellular Biochemistry*. 1994:208-.
230. Prakash SK. Loss of holocytochrome c-type synthetase causes the male lethality of X-linked dominant micro-phthlalmia with linear skin defects (MLS) syndrome. *Human Molecular Genetics*. 2002;11(25):3237-48.
231. Indrieri A, van Rahden VA, Tiranti V, Morleo M, Iaconis D, Tammara R, et al. Mutations in COX7B cause microphthalmia with linear skin lesions, an unconventional mitochondrial disease. *Am J Hum Genet*. 2012;91(5):942-9.
232. Morleo M FB. Microphthalmia with Linear Skin Defects Syndrome. In: Pagon RA, Adam MP, Ardinger HH, et al, editors *GeneReviews®* [Internet] Seattle (WA): University of Washington, Seattle; 1993-2015 Available from: <http://www.ncbi.nlm.nih.gov/books/NBK7041/>. 2009 [Updated 2011].
233. Allanson J, Richter S. Linear skin defects and congenital microphthalmia: a new syndrome at Xp22.2. *Journal of Medical Genetics*. 1991;28(2):143-4.
234. Alberry MS, Juvanic G, Crolla J, Soothill P, Newbury-Ecob R. Pseudotail as a feature of microphthalmia with linear skin defects syndrome. *Clin Dysmorphol*. 2011;20(2):111-3.
235. Sharma VM, Ruiz de Luzuriaga AM, Waggoner D, Greenwald M, Stein SL. Microphthalmia with linear skin defects: a case report and review. *Pediatr Dermatol*. 2008;25(5):548-52.
236. Kutsche K, Werner W, Bartsch O, von der Wense A, Meinecke P, Gal A. Microphthalmia with linear skin defects syndrome (MLS): a male with a mosaic paracentric inversion of Xp. *Cytogenet Genome Res*. 2002;99(1-4):297-302.
237. Inoue K, Khajavi M, Ohyama T, Hirabayashi S, Wilson J, Reggin JD, et al. Molecular mechanism for distinct neurological phenotypes conveyed by allelic truncating mutations. *Nat Genet*. 2004;36(4):361-9.
238. Van den Veyver IB. Skewed X inactivation in X-linked disorders. *Semin Reprod Med*. 2001;19(2):183-91.
239. Franco B, Ballabio A. X-inactivation and human disease: X-linked dominant male-lethal disorders. *Curr Opin Genet Dev*. 2006;16(3):254-9.

240. Hobson GM, Gibson CW, Aragon M, Yuan ZA, Davis-Williams A, Banser L, et al. A large X-chromosomal deletion is associated with microphthalmia with linear skin defects (MLS) and amelogenesis imperfecta (XAI). *Am J Med Genet A*. 2009;149A(8):1698-705.
241. Vergult S, Leroy B, Claerhout I, Menten B. Familial cases of a submicroscopic Xp22.2 deletion: genotype-phenotype correlation in microphthalmia with linear skin defects syndrome. *Mol Vis*. 2013;19:311-8.
242. Opdal SH, Rognum TO, Torgersen H, Vege A. Mitochondrial DNA point mutations detected in four cases of sudden infant death syndrome. *Acta Paediatr*. 1999;88(9):957-60.
243. Opdal SH, Vege A, Egeland T, Musse MA, Rognum TO. Possible role of mtDNA mutations in sudden infant death. *Pediatr Neurol*. 2002;27(1):23-9.
244. Vives-Bauza C, Andreu AL, Manfredi G, Beal MF, Janetzky B, Gruenewald TH, et al. Sequence analysis of the entire mitochondrial genome in Parkinson's disease. *Biochem Biophys Res Commun*. 2002;290(5):1593-601.
245. Campos Y, Martin MA, Rubio JC, Gutierrez del Olmo MC, Cabello A, Arenas J. Bilateral striatal necrosis and MELAS associated with a new T3308C mutation in the mitochondrial ND1 gene. *Biochem Biophys Res Commun*. 1997;238(2):323-5.
246. Mezghani N, Mnif M, Mkaouar-Rebai E, Kallel N, Charfi N, Abid M, et al. A maternally inherited diabetes and deafness patient with the 12S rRNA m.1555A>G and the ND1 m.3308T>C mutations associated with multiple mitochondrial deletions. *Biochem Biophys Res Commun*. 2013;431(4):670-4.
247. Vilarinho L, Choro R, Cardoso ML, Rocha H, Nogueira C, Santorelli FM. The ND1 T3308C mutation may be a mtDNA polymorphism. Report of two Portuguese patients. *J Inherit Metab Dis*. 1999;22(1):90-1.
248. Picardi E, Pesole G. Mitochondrial genomes gleaned from human whole-exome sequencing. *Nat Methods*. 2012;9(6):523-4.
249. Jain D, Chopra P. Histiocytoid cardiomyopathy: does it exist in the fetal-age group? *Cardiovasc Pathol*. 2011;20(6):386-7.
250. Worthey EA, Mayer AN, Syverson GD, Helbling D, Bonacci BB, Decker B, et al. Making a definitive diagnosis: Successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genetics in Medicine*. 2011;13(3):255-62.
251. Willig LK, Petrikin JE, Smith LD, Saunders CJ, Thiffault I, Miller NA, et al. Whole-genome sequencing for identification of Mendelian disorders in critically ill infants: a retrospective analysis of diagnostic and clinical findings. *Lancet Respiratory Medicine*. 2015;3(5):377-87.
252. Fihn SD, Gardin JM, Abrams J, Berra K, Blankenship JC, Dallas AP, et al. 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease: a report of the American College of Cardiology Foundation/American Heart Association task force on practice guidelines, and the American College of Physicians, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. *Circulation*. 2012;126(25):e354-471.
253. Cannon CP, Brindis RG, Chaitman BR, Cohen DJ, Cross JT, Jr., Drozda JP, Jr., et al. 2013 ACCF/AHA key data elements and definitions for measuring the clinical management and outcomes of patients with acute coronary syndromes and coronary artery disease: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Clinical Data Standards (Writing Committee to Develop Acute Coronary Syndromes and Coronary Artery Disease Clinical Data Standards). *Circulation*. 2013;127(9):1052-89.
254. Huang Y, Yang J, Ying D, Zhang Y, Shotelersuk V, Hirankarn N, et al. HLAreporter: a tool for HLA typing from next generation sequencing data. *Genome Med*. 2015;7(1):25.

4 Appendix

Figure 4-1 Definition of Myocardial Infarction

Criteria for acute myocardial infarction
The term acute myocardial infarction (MI) should be used when there is evidence of myocardial necrosis in a clinical setting consistent with acute myocardial ischaemia. Under these conditions any one of the following criteria meets the diagnosis for MI: • Detection of a rise and/or fall of cardiac biomarker values [preferably cardiac troponin (cTn)] with at least one value above the 99th percentile upper reference limit (URL) and with at least one of the following: • Symptoms of ischaemia. • New or presumed new significant ST-segment-T wave (ST-T) changes or new left bundle branch block (LBBB). • Development of pathological Q waves in the ECG. • Imaging evidence of new loss of viable myocardium or new regional wall motion abnormality. • Identification of an intracoronary thrombus by angiography or autopsy. • Cardiac death with symptoms suggestive of myocardial ischaemia and presumed new ischaemic ECG changes or new LBBB, but death occurred before cardiac biomarkers were obtained, or before cardiac biomarker values would be increased. • Percutaneous coronary intervention (PCI) related MI is arbitrarily defined by elevation of cTn values (>5 x 99th percentile URL) in patients with normal baseline values (≤99th percentile URL) or a rise of cTn values >20% if the baseline values are elevated and are stable or falling. In addition, either (i) symptoms suggestive of myocardial ischaemia or (ii) new ischaemic ECG changes or (iii) angiographic findings consistent with a procedural complication or (iv) imaging demonstration of new loss of viable myocardium or new regional wall motion abnormality are required. • Stent thrombosis associated with MI when detected by coronary angiography or autopsy in the setting of myocardial ischaemia and with a rise and/or fall of cardiac biomarker values with at least one value above the 99th percentile URL. • Coronary artery bypass grafting (CABG) related MI is arbitrarily defined by elevation of cardiac biomarker values (>10 x 99th percentile URL) in patients with normal baseline cTn values (≤99th percentile URL). In addition, either (i) new pathological Q waves or new LBBB, or (ii) angiographic documented new graft or new native coronary artery occlusion, or (iii) imaging evidence of new loss of viable myocardium or new regional wall motion abnormality.
Criteria for prior myocardial infarction
Any one of the following criteria meets the diagnosis for prior MI: • Pathological Q waves with or without symptoms in the absence of non-ischaemic causes. • Imaging evidence of a region of loss of viable myocardium that is thinned and fails to contract, in the absence of a non-ischaemic cause. • Pathological findings of a prior MI.

Taken from the Third Universal Definition of MI (141)

Figure 4-2 Screenshots of proforma used to capture information on MI cohort

Recording information of prior revascularisation and indication for CMR

History of prior revascularisation :	☐ Yes ☐ No ☒ Unknown
History of CABG :	☐ Yes ☐ No ☒ Unknown
History of PCI :	☐ Yes ☐ No ☒ Unknown
Indication for CMR :	☐ Assessment of ischaemia/viability ☐ Diagnostic uncertainty ☐ Prior to device insertion (ICD/CRT) ☐ Assess myocardial fibrosis ☐ Other ☒ Unknown

Recording information on Myocardial Infarction

Infarction :	
Infarction :	☒ Yes ☐ No ☐ Unknown
Most affected territory :	☐ LAD ☐ LCx ☐ PDA ☐ RCA ☒ Unknown LCx=posterior/infero-lateral, PDA=inferior, RCA=if RCA on angio
Wall thickness in most affected territory (mm) :	
Multi-territory infarct :	☐ Yes ☐ No ☒ Unknown
Wall motion of most affected territory :	☐ normal ☐ hypo ☐ aknetic ☐ dyskinetic ☒ Unknown
Mitral regurgitation :	☐ None ☐ Minimal ☐ Moderate ☐ Severe ☒ Unknown
Infarcted infero-medial papillary :	☐ Yes ☐ No ☒ Unknown
Infarct type :	☐ full thickness ☐ partial thickness ☒ Unknown partial=<50%
Infarct thickness :	☐ 0-25 % ☐ 26-50 % ☐ 51-75 % ☐ 76-100 % ☒ Unknown
total infarcted segments :	
age when had infarct :	

Recording information on perfusion and hibernation

Perfusion :	
Perfusion studies done :	☒ Yes ☐ No ☐ Unknown
Inducible perfusion deficit :	☒ Yes ☐ No ☐ Unknown
Main perfusion deficit territory :	☐ LAD ☐ LCx ☐ PDA ☐ RCA ☒ Unknown
Multi-territory perfusion deficit (non syndrome X) :	☐ Yes ☐ No ☒ Unknown
Microvascular ischaemia (syndrome X) :	☐ Yes ☐ No ☒ Unknown
Non-viable segments :	
Hibernating segments :	
Viable segments :	

Recording information on arrhythmia, family history, presence and degree of significant coronary artery disease.

Arrhythmia :	
Atrial fibrillation :	☐ Yes ☐ No ☒ Unknown
Non sustained VT :	☐ Yes ☐ No ☒ Unknown
Sustained VT :	☐ Yes ☐ No ☒ Unknown
Ventricular fibrillation :	☐ Yes ☐ No ☒ Unknown
Family history of SCD :	☐ Yes ☐ No ☒ Unknown
Family history of DCM :	☐ Yes ☐ No ☒ Unknown
Family history of HCM :	☐ Yes ☐ No ☒ Unknown
Family history of ARVC :	☐ Yes ☐ No ☒ Unknown
Family history of premature coronary artery disease :	☐ Yes ☐ No ☒ Unknown
Significant Coronary Artery Disease :	☒ Yes ☐ No ☐ Unknown Angiographically significant CAD (Any major epicardial artery with more than >70%, or Left Main or proximal LAD >50% stenosis)
Significant Coronary Artery Disease Options :	☐ 1 vessel ☐ 2 Vessel ☐ 3 vessel ☒ Unknown

Recording date of diagnosis and date of CMR scan

Date of diagnosis :	01/01/1985 
Date of first diagnostic CMR :	

Table 4-1 Standardized clinical definitions used for data capture in MI cohort

Term		Standardized clinical definition
Significant Coronary Artery Disease (CAD)		The most commonly used nomenclature for defining coronary artery anatomy is based on the assumption that there are three major coronary arteries; the left anterior descending (LAD), the Circumflex (Cx or LCx) and the right coronary artery (RCA), with a right-dominant, left-dominant or co-dominant circulation. The extent of disease is defined as one- vessel, two-vessel or three vessel, with a significant stenosis indicating a diameter reduction of $\geq 70\%$ in any of the LAD, LCx and RCA or 50% in the left main stem (LMS) (252) on coronary angiography concurrent with time of CMR.
Non-sustained Ventricular Tachycardia (NSVT)		Yes = Documented history of NSVT (defined as three or more consecutive ventricular beats at a rate of greater than 100 beats/min with a duration of less than 30 seconds). No = No documented history of NSVT and at least one 24 Holter with no evidence of NSVT. Unknown = Either no clinical information available or history of symptoms suggestive of NSVT.
Sustained Ventricular Tachycardia (VT)		Yes = Documented history of Sustained VT. No = No documented history of Sustained VT and no history suggestive of symptoms of Sustained VT. Unknown = Either no clinical information available or history of symptoms suggestive of NSVT in the absence of negative investigations.
Ventricular Fibrillation (VF)		Yes = Documented history of VF. No = No documented history of VF and no history suggestive of VF. Unknown = Either no clinical information available or history of symptoms suggestive of VF in the absence of negative investigations.
Atrial Fibrillation (AF)		Yes = Chronic AF/history of AF/ history of PAF/ ECG at recruitment showing AF. No = No documented history of AF or history suggestive of PAF and an ECG in SR. Unknown = No ECG available or no clinical information available or history of symptoms suggestive of AF in the absence of negative investigations (such as normal Holter monitor).
Family History	DCM	Documented family history of 'DCM' in one or more 1st or 2nd degree relative(s).
	HCM	Documented family history of 'HCM' in one or more 1st or 2nd degree relative(s).
	ARVC	Documented family history of 'ARVC' in one or more 1st or 2nd degree relative(s).
	SCD	Documented family history of 'sudden or unexplained in one or more 1st or 2nd degree relative(s).
	Premature coronary artery disease-	Documented family history of premature coronary artery disease: Any direct blood relatives (parents, siblings, children) who have had any of the following (Angina, Acute MI, Sudden Cardiac Death without obvious cause, CABG surgery or PCI) at age less than 55 years for male relatives or less than 65 years for female relatives(253).

ARVC, Arrhythmogenic Right Ventricular Cardiomyopathy; BPM, beats per minute; CABG, Coronary Artery Bypass Grafting; DCM, Dilated Cardiomyopathy; HCM, Hypertrophic Cardiomyopathy; MI, Myocardial Infarction; PCI, Percutaneous Coronary Intervention; SCD, Sudden Cardiac Death.

Table 4-2 BRU ID's of cohort of subjects with CMR evidence of MI (n=335)

10AB00687	10CI02867	10DT00394	10GR0016	10JW00348	10MK0039	10PM01459	10SB03509	12DB00482
10AB00696	10CJ02554	10DT01655	10GS01545	10JW01010	10MK0199	10PN00319	10SB03869	12DH0037
10AB03502	10CM0022	10DW0277	10GT0256	10JW01409	10MK0216	10PO02159	10SC03754	12GP00265
10AC02352	10CM0244	10DW0398	10GT0313	10JW01763	10MK0388	10PV00951	10SD00339	12GW0049
10AD00920	10CS00065	10EA02520	10HG0028	10JW02122	10MM0339	10PW0072	10SD01832	12JB00489
10AF00538	10CS01696	10EA03688	10HH0185	10KD02161	10MR00010	10RA01508	10SG03954	12JB00913
10AF01727	10CT01570	10EB03683	10HM0243	10KK03464	10MR01583	10RA03155	10SH03588	12JM00502
10AG02024	10CT02604	10EB03903	10HM0271	10KK03648	10MR01609	10RB00013	10SL02652	12JR00759
10AH02364	10CW0176	10EC00851	10HS00919	10KM03207	10MS00277	10RB04162	10SM00353	12KE00823
10AL02308	10DB01184	10EC01486	10IA03680	10KO01523	10MS01391	10RC00821	10SR00862	12KM0089
10AM00076	10DB01765	10EC02142	10IF03153	10KV00564	10MS01537	10RC01446	10SS00959	12KP00632
10AM00335	10DB01945	10EC03303	10IK01960	10KW0196	10MS01596	10RC01684	10SS01234	12MC0009
10AM01947	10DB02297	10ED00373	10IW01863	10LB01168	10MS04201	10RD01169	10SS02150	12MF0107
10AM03622	10DC00108	10ED03350	10JA00334	10LB01752	10MT00451	10RD01245	10TC03169	12MG0112
10AM03743	10DC00250	10ED03769	10JA00523	10LB01930	10MW0091	10RG01738	10TC03228	12MM0039
10AO00220	10DC00852	10EF01560	10JA01257	10LB02463	10MW0163	10RG01932	10TG03111	12MS0052
10AR01826	10DC01840	10EF02449	10JB00006	10LC01673	10MW0353	10RG02677	10TH02584	12NL00793
10AS01828	10DC03840	10EH02372	10JB02380	10LD03018	10MW0389	10RG03304	10TH03736	12PE00573
10AS02486	10DD03746	10EL01442	10JB02694	10LE03030	10MW0391	10RH01983	10TJ00305	12PP00470
10AW0252	10DE02195	10EO01601	10JB03170	10LG01360	10NC00942	10RH02545	10TM00255	12RL00443
10AY02776	10DE03996	10ET03706	10JC02003	10LH03368	10NM0374	10RJ00370	10TP02441	12RP00785
10BC00144	10DG00129	10EW01698	10JC03016	10LP00661	10NS02690	10RJ03089	10TR01729	12WH0034
10BD03332	10DG02418	10EW02106	10JD01605	10LP02901	10NW0047	10RK01578	10TS01875	12WR0078
10BH01334	10DG03750	10FA00081	10JD03330	10MA0032	10PA01615	10RK03448	10TS02716	
10BM00207	10DH01587	10FK00294	10JE01764	10MA0072	10PA02042	10RL01743	10TV02198	
10BM01710	10DH01604	10FS04063	10JG03410	10MA0160	10PA02115	10RM0096	10WA0395	
10BM01879	10DJ02128	10FT01640	10JH00608	10MA03305	10PB02271	10RM0266	10WD0180	
10BM02061	10DM0007	10GA01966	10JJ01842	10MB00827	10PB03542	10RM0319	10WE01880	
10BM03713	10DM0147	10GA03600	10JK03140	10MB04098	10PC02470	10RM0315	10WH0348	
10BN00817	10DM0152	10GB00995	10JK03684	10MC01984	10PD01420	10RP00337	10WO0257	
10BP01773	10DM0313	10GG01869	10JL00430	10MC02141	10PD01519	10RP00728	10YR01980	
10BR02781	10DN02289	10GH02468	10JM01264	10MD00960	10PE00955	10RP01661	11AA00009	
10BS02643	10DO01720	10GJ00662	10JM01489	10ME01915	10PF00346	10RR03024	11DR00004	
10BT00657	10DP03156	10GL01853	10JN01775	10MF02496	10PF00941	10RS01431	11RB00012	
10BW01691	10DR01522	10GL03747	10JO02565	10MF03044	10PG02005	10RT01588	12AD00057	
10CB01376	10DR02311	10GM0334	10JO03535	10MF03179	10PH01694	10RT02183	12AE00377	
10CB02323	10DS00746	10GP00249	10JP00958	10MG00681	10PL00702	10SA00676	12BH00494	
10CG01982	10DS01757	10GP01554	10JR01995	10MG00682	10PL03579	10SA01645	12BH01228	
10CI01632	10DS03752	10GP01914	10JS02182	10MG01925	10PM00393	10SA02761	12CH00267	

Table 4-3 BRU ID's of cohort of Healthy Volunteers (HVOLs) included in study (n= 431)

14AA01658	14AR01530	14CM01680	14EB01858	14HA01434	14JF01463	14KN01429	14MM01805	14PA01545	14RS01641	14SS01768
14AB01788	14AS01383	14CM01936	14EC01433	14HA01815	14JF01908	14KR01871	14MM01837	14PC01309	14RS01752	14SS01830
14AB01833	14AS01535	14CP01466	14EC01935	14HB01499	14JG01730	14KS01749	14MM01848	14PE01950	14RS01774	14SS01929
14AB01912	14AS01679	14CR01440	14ED01866	14HB01699	14JG01839	14KT01842	14MO01335	14PF01454	14RS01889	14ST01686
14AB01963	14AS01772	14CS01449	14EG01498	14HB01804	14JH01338	14KW01281	14MO01427	14PI01855	14RW01763	14SV01502
14AC01351	14AS01934	14CS01674	14EH01413	14HB01884	14JH01887	14KW01902	14MO01954	14PK01881	14RW01854	14SV01835
14AC01671	14AS01947	14CS01905	14EH01503	14HD01295	14JJ01753	14LB01343	14MP01525	14PL01949	14RZ01770	14SW01733
14AC01903	14AT01349	14CW01426	14EH01808	14HD01838	14JK01282	14LB01748	14MQ01722	14PM01297	14SA01523	14SY01504
14AD01439	14AT01818	14CW01826	14EH01870	14HD01849	14JK01732	14LD01382	14MR01765	14PM01421	14SB01516	14TA01341
14AD01695	14AV01806	14CY01518	14EI01892	14HH01442	14JL01668	14LD01897	14MR01893	14PM01944	14SB01729	14TC01664
14AD01814	14AW01807	14DA01414	14EK01363	14HH01797	14JM01448	14LH01943	14MS01573	14PN01744	14SB01898	14TD01766
14AD01860	14BA01438	14DB01755	14EM01845	14HH01872	14JM01450	14LL01420	14MS01648	14PN01746	14SB01907	14TG01576
14AD01879	14BB01425	14DB01809	14EO01332	14HH01874	14JM01711	14LL01542	14MS01728	14PR01694	14SC01373	14TH01900
14AF01374	14BB01843	14DB01918	14EP01551	14HJ01863	14JM01775	14LM01759	14MS01754	14PR01783	14SC01488	14TL01784
14AF01495	14BC01350	14DC01669	14FA01412	14HL01455	14JM01801	14LM01820	14MS01846	14PS01275	14SC01555	14TL01923
14AG01422	14BC01859	14DC01867	14FC01802	14HN01532	14JN01819	14LR01537	14MS01891	14PT01873	14SC01798	14TM01521
14AG01575	14BC01906	14DD01640	14FH01568	14HN01816	14JP01471	14LR01675	14MS01939	14RA01645	14SH01831	14TO01364
14AG01689	14BC01925	14DF01386	14FH01776	14HS01821	14JP01577	14LS01747	14NA01875	14RA01868	14SH01885	14TO01750
14AG01693	14BH01847	14DF01490	14FM01457	14HS01941	14JP01644	14LT01876	14NB01824	14RB01428	14SH01895	14TP01423
14AG01764	14BM01886	14DF01712	14FN01464	14IA01705	14JS01544	14LW01500	14NC01452	14RB01899	14SH01956	14TP01928
14AG01850	14BO01441	14DG01670	14FP01865	14IC01953	14JS01825	14LW01713	14NF01841	14RC01685	14SJ01795	14TR01696
14AH01458	14BO01702	14DG01684	14FZ01507	14IH01751	14JS01942	14MA01655	14NK01654	14RC01817	14SJ01828	14TR01743
14AH01539	14BP01522	14DH01731	14GC01853	14IL01719	14JS01959	14MA01932	14NK01767	14RC01909	14SK01758	14TS01337

Continued on next page

Table 4-3 continued from previous page

14AI01781	14BS01890	14DH01857	14GC01877	14IL01948	14JT01493	14MB01643	14NK01829	14RD01952	14SL01665	14TW01786
14AI01787	14BT01840	14DK01945	14GJ01656	14IM01913	14JT01571	14MB01710	14NL01717	14RE01447	14SM01435	14UK01501
14AK01726	14BW01869	14DL01739	14GJ01799	14IR01773	14JT01836	14MC01553	14NL01955	14RF01737	14SM01546	14VA01467
14AL01844	14CA01938	14DM01509	14GJ01811	14JA01436	14JV01397	14MC01687	14NM01701	14RG01777	14SM01690	14VD01418
14AM01272	14CB01517	14DN01375	14GL01794	14JA01659	14JW01741	14MC01878	14NM01734	14RG01946	14SM01708	14VH01362
14AM01486	14CB01779	14DO01445	14GM01331	14JB01547	14JW01920	14ME01310	14NM01793	14RH01353	14SM01792	14VH01813
14AM01672	14CD01396	14DP01411	14GM01515	14JB01691	14KC01470	14ME01431	14NM01852	14RH01398	14SM01810	14VL01355
14AM01937	14CG01678	14DR01762	14GM01822	14JB01698	14KC01681	14MG01958	14NP01827	14RK01494	14SM01894	14VM01832
14AN01543	14CH01329	14DS01385	14GM01931	14JC01930	14KD01677	14MJ01834	14NS01888	14RK01727	14SM01919	14VR01714
14AO01372	14CH01359	14DS01443	14GP01692	14JD01384	14KD01957	14MK01357	14NT01465	14RL01725	14SM01940	14WH01862
14AP01771	14CH01921	14DS01636	14GR01437	14JD01724	14KF01492	14MK01926	14NT01703	14RM01419	14SN01330	14YB01780
14AP01789	14CI01430	14DU01356	14GR01723	14JD01790	14KH01415	14ML01336	14OB01856	14RM01497	14SP01704	14YL01484
14AR01416	14CJ01735	14DV01462	14GR01778	14JD01823	14KH01485	14ML01800	14OL01738	14RM01513	14SP01761	14ZD01444
14AR01432	14CJ01782	14DW01572	14GR01914	14JD01896	14KH01791	14MM01469	14OL01785	14RP01505	14SP01851	14ZN01340
14AR01483	14CK01531	14EB01652	14GS01381	14JD01904	14KM01456	14MM01512	14ON01682	14RR01901	14SR01769	14ZW01519
14AR01514	14CM01424	14EB01736	14GT01365	14JD01922	14KN01296	14MM01527	14PA01489	14RS01274	14SS01496	

Table 4-4 BRU ID's of cohort with end stage ischaemic cardiomyopathy (n=95)

20AB01023	20CW01004	20GF01026	20KC01024	20PB01073	20RT00959	Note six additional cases failed sequencing outright and were excluded from the analysis
20AC00985	20CW01010	20GM01045	20KH00980	20PF01006	20SB01047	
20AF00990	20DD01039	20GP00998	20KP00984	20PF01054	20SK00965	
20AG00992	20DG00962	20GT01066	20KS01068	20PM01021	20SM01041	
20AH01009	20DJ01069	20GW01002	20LB01034	20PP01063	20SR00974	
20AS00996	20DP00967	20HF00986	20LV01037	20PS00960	20SW01028	
20BD01051	20DP01053	20HP00970	20MH01048	20PT01036	20TB00977	
20BQ01061	20DS00975	20HP01005	20MM00987	20RB00978	20TD01030	
20BS00997	20DS00981	20IC00991	20MM00993	20RB01059	20TH01032	
20BS01064	20DW00961	20JA00971	20MM01001	20RC01042	20TH01049	
20BT00983	20EC01060	20JF01020	20MP01019	20RG00968	20TN01043	
20BT01055	20ED01031	20JM00972	20MR00979	20RH00989	20TR01011	
20CB00973	20EJ00982	20JN01050	20MS00994	20RM01038	20VK01008	
20CB01018	20GB01046	20JS01003	20NM01072	20RN01022	20WL01033	
20CC00976	20GC00999	20JS01029	20OM01062	20RP01027	20YH00988	
20CP01040	20GD01035	20KA01017	20PB01000	20RS01007		

Table 4-5 Details of Amplicon PCR primers used for Sanger sequencing of *TTN*vs

BRU ID	Variant	Primer Name	Start	End	Forward primer	Reverse primer
10AF00538	c.66010delG	TTN_ex314	179446901	179447592	GTGGGAGAAAGACATTCATATTCTAATG	GCTAATAACAGGAGGATCACAGC
10AO0020	c.77421C>A	TTN_ex326a	179433239	179433856	GTTGCAGTTATTACTTTAGAGTCACAGC	TGGCAACTGTGATTCCATATTGT
10BC00144	c.45756dupA	14F	179485419	179485750	CATACTATAAAGTCTTGTCTGGTTGC	GAAAACGTATGAAAAAATCTCTAGTGG
10BM02061	c.44816-1G>A	TTN_ex243	179487291	179487663	GC AAAAGCATCATACCTGATG	GGTGACCAGAGAAGTTGTGACTAA
10CL03599	c.29006_29015delGAAGTGACAC	TTN_100 (GR)	179572133	179572764	ATGTGATGTGTTCTTTGCTTGT	CAGATGAGCAACAACCTGATATTT
10DG00129	c.79645G>T	TTN_326b	179431009	179431499	GATCATGAGTATGAATTCAGGGTC	TTGAATGGAATGTGAAITCTGG
10DG00129	c.37231G>T	125F	179526330	179526689	GGTACTTGTCAATTTGACCTTAAGC	TATACATACGTAAACAACCTGCACG
10DG00129	c.36646G>T	71F	179527850	179528733	ACTGACTTTAGGCTTCAGAAGACC	CCACCAAGATATTTTGGATAGC
10DM00077	c.81262_81269delCAGATGCT	26F	179429086	179429886	ATATCTTGGGAACCTCCAGC	AGATGGTTCATAAAGTTTCCAGC
10DM03183	c.7498C>T	32F	179412451	179413076	TAAATGAGTATGGTGTGGTGAGC	ACGCTAAATGTTTTAACACCAGC
10EF02449	c.45344delT	TTN_245_2 (GR)	179485978	179486290	GGTCATTCAGAACGCTCACCC	TCATCCCTCTCCACTGGAC
10JC03016	c.31426+1G>C	116F	179559051	179559835	TTGTACCCAGAGTTCAAAGTTGTG	AATGCATTGCATAGTGAAGTTAGC
10KW01906	c.95415_95416+2delCAGT	73F	179410887	179410310	GAAAGAATGATCTTCATCATGGC	CTAGGTAGTTGTGATTTTCAATGC
10PB02271	c.37770dupA	TTN_186(GR)	179518498	179523147	CTCACAATCACTGTAATATATACCTTCA	AGAGGAGATATCTCTTCTGCAGAAT
10RB00013	c.105289G>T	TTN_358	179395728	179396424	TTAAATGGTACCACAATGGTGT	TTTTCTACCACCACGCTGTAAT
10RP00728	c.37731delA	TTN_186	179518498	179523147	CTCACAATCACTGTAATATATACCTTCA	AGAGGAGATATCTCTTCTGCAGAAT
10TM00255	c.85417G>T	108F	179424969	179425818	AGTTAAAGATGATGTTGAGCCTCC	AGCATTTTCTGCATAAACACG
12CH00267	c.1624G>T	TTN_10	179656559	179657245	GACAGAAATAGTACAGAAATAATATT	ATAGATGTGCCCTAAACTGTTCTGT
12GP00265	c.36964G>T	TTN_177	179518379	179518896	TCTGTACCCCTCACAAACACT	AGGAGCCTTAGGAACTTTCTTTT
14AB01833	c.37285+1G>C	125F	179526330	179526689	GGTACTTGTCAATTTGACCTTAAGC	TATACATACGTAAACAACCTGCACG
14AH015392	c.32554+1G>C	99F	179549511	179549878	GTCTACGTTTGTGACAAACACATG	GATATGGAAAACACTAAACACAGGC
14AO013721	c.11254+2T>C	138F	179620782	179621635	AACACTCGTCTTTCATCTTCATCTTCC	ATAATCATGGCCAAATGACTTGG
14CR01440	c.32588dupA	141F	179485654	179486221	TGGCAAAAGTCAAAGTACATGG	TTGAAGACAACCTTCTGTGTTC
14CS01449	c.63464delGinsAGCCC	144F	79452481	179453061	GACACTAATTGAAAAATTGGATGAC	TATGACCAGTTTCTCATGCTG
14EC01433	c.30803-2A>G	140F	179560499	179561202	ATCCTCAAATCAAGAGTAGAATGC	ATAAACTGCAATCAGGTTTCATAGC
14JC01930	c.98506C>T	86F	179404081	179404789	ACATATGATCATTTTGCTTCTGTC	ATAATTTTGTGGTTGAAAGGGC
14JD01896	c.21142C>T	139F	179588507	179588981	TTCTGCTGTTTTCTTCTGTTG	AAATGTTTGTCAACATAGATGTGC
14JM01448	c.67159delA	47F	179444807	179445426	TAAATGTATTTTAGGAGTGACAAGGC	TAGACCAGGATTTCTCTCTGTC
14MO01427	c.45391delA	TTN_245_2(GR)	179485978	179486290	GGTCATTCAGAACGCTCACCC	TCATCCCTCTCCACTGGAC
14PM01944	c.106532-1G>T	33F	179393560	179394183	CTTAAGTCATATGAATCATCTTTGGC	AGTACCCATTTACATCAGTGG
14PS01275	c.6802G>T	145F	179638791	179639349	TTTATCTCTCAGCCATCAGC	ACAGACTTTTTGGTCACTAAGTCC

Table 4-5 Continued from previous page

14RH013531	c.11254+2T>C	138F	179620782	179621635	<i>AAC</i> TACGTCTTTCATCTTCATCTTCC	<i>ATA</i> ATCATGGCCAATGACTTGG
14SM01546	c.67348C>T	143F	79444517	179445071	<i>TAC</i> CAATTGTTGTGAAAGTGCTTG	<i>GCT</i> ACAGGATGACTTAGATACAGACC
14SS01830	c.10852C>T	138F	179620782	179621635	<i>AAC</i> TACGTCTTTCATCTTCATCTTCC	<i>ATA</i> ATCATGGCCAATGACTTGG
14ZN01340	c.17508dupA	T1N_ex60	179595601	179596010	<i>TGG</i> GTTTAGGAAAGAAATGC	<i>AAG</i> AAAGGARCATGGAAATGC
20KP00984	11183dupG	T1N_46 (GR)	179620713	179621211	<i>GC</i> ATCTTCAGGACGTCACTG	<i>CTT</i> TGCTTCCCGAACCTACG
20KP00984	17823delA	T1N_61 (GR)	179595057	179595696	<i>ATG</i> ATGTTGGGAGGAGCAGC	<i>TG</i> ACATCTTGTGACTGTGGC
20LB01034	c.13606_13607insAACG	T1N_48 (GR)	179603882	179604623	<i>CTT</i> CGGCAAAGTCTGTAAACAG	<i>GTG</i> AGTTTGGCTGAAGTTGC
20RM01038	c.106542delTinsGT	33F	179393560	179394183	<i>CTT</i> AAGTCATATGAATCATCTTTGGC	<i>AGT</i> ACCCATTTACATCAGTGG

Table 4-6 Maximum, minimum and mean callability of all genes (n=202) sequenced using SureSelect and SOLiD across cohorts

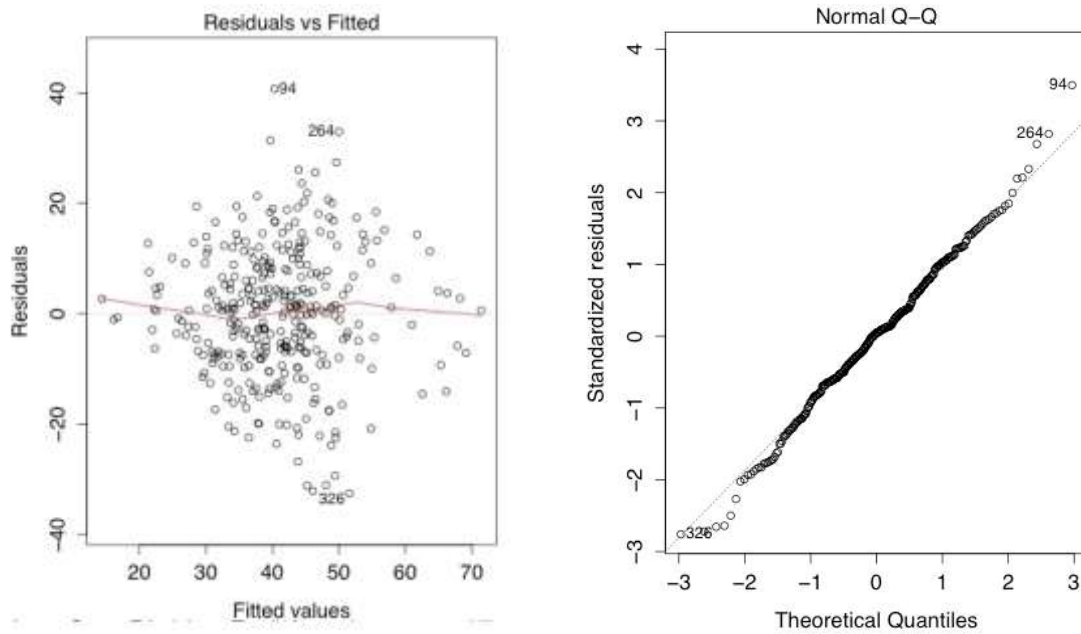
	Cohort of subjects with CMR evidence of MI			HVOL cohort			End-stage Ischaemic cardiomyopathy cohort		
	Maximum % Callable	Minimum % Callable	Mean % Callable	Maximum % Callable	Minimum % Callable	Mean % Callable	Maximum % Callable	Minimum % Callable	Mean % Callable
Gene ID									
<i>ABCC8</i>	100	83.2	99.8	100	24.9	95.6	100	97.8	99.8
<i>ABCC9</i>	100	97.0	100.0	100	30.8	96.1	100	88.4	99.7
<i>ACTC1</i>	100	89	99.9	100	26.5	96.5	100	95.7	99.9
<i>ACTN2</i>	100	94.5	99.9	100	25.7	94.6	100	97.2	99.9
<i>ADRA2B</i>	100	69.5	99.6	100	28.7	96.4	100	89.5	99.1
<i>ADRB2</i>	100	95.1	99.9	100	25.8	94.7	100	95.2	99.8
<i>AGL</i>	100	0	99.4	100	27.1	95.3	100	71.6	99.2
<i>AKAP9</i>	100	97.1	99.9	100	26.9	96.3	100	55.6	98.8
<i>ALMS1</i>	99.7	98.3	99.4	100	19.6	94.6	99.9	93.6	99.4
<i>ANK2</i>	100	94.1	100.0	100	19.6	95.1	100	98.3	99.9
<i>ANKRD1</i>	100	95.5	99.6	100	31.0	97.6	100	89.1	99.0
<i>APOA1</i>	100	65.3	88.1	100	27.2	95.3	100	73.6	92.9
<i>ATG5</i>	100	98.9	100.0	100	26.5	94.5	100	53.6	98.7
<i>ATP1B1</i>	100	92.6	99.9	100	26.3	96.0	100	99.9	100.0
<i>ATP2A2</i>	100	0	99.4	100	21.8	95.2	100	96.4	100.0
<i>BAG3</i>	100	0	98.4	100	25.0	95.0	100	92.2	98.7
<i>C21orf7</i>	100	42.4	99.6	100	23.8	95.1	100	96.2	99.8
<i>CACNA1C</i>	100	86.0	99.5	100	20.3	94.5	100	96.8	99.6
<i>CACNA1D</i>	100	29.9	99.2	100	25.3	95.1	100	97.4	99.8
<i>CACNA2D1</i>	100	1.2	98.9	100	30.8	96.7	100	56.1	96.7
<i>CACNB2</i>	100	93.2	99.8	100	21.7	96.0	100	92.4	99.8
<i>CALM1</i>	100	0	99.4	100	23.8	95.3	100	93.6	99.9
<i>CALM2</i>	100	0	97.4	100	19.8	95.3	100	73.4	93.4
<i>CALM3</i>	100	0	98.8	100	28.5	96.5	100	98.9	99.4
<i>CALR3</i>	100	92.2	99.9	100	28.3	96.6	100	97.3	99.9
<i>CAPN1</i>	100	78.7	99.8	100	23.7	95.4	100	97.5	99.9
<i>CASQ2</i>	100	100	100	100	30.6	97.1	100	97.1	99.9
<i>CAV3</i>	100	100	100	100	19.6	93.6	100	100	100
<i>CCT2</i>	100	99.3	100.0	100	21.4	95.4	100	90.2	99.8
<i>CCT3</i>	100	87.3	97.7	100	27.6	94.8	100	92.7	96.9
<i>CCT4</i>	100	93.9	99.8	100	25.1	96.2	100	82.4	99.1
<i>CCT5</i>	100	99.1	100.0	100	23.8	94.9	100	95.7	99.9
<i>CCT6A</i>	100	91.3	99.3	100	19.6	94.9	100	83.2	99.1
<i>CCT6B</i>	100	97.2	100.0	100	26.4	95.9	100	57.9	98.2
<i>CCT7</i>	100	98.7	100.0	100	31.0	96.5	100	100.0	100.0
<i>CCT8</i>	100	96.4	100.0	100	19.6	94.3	100	76.4	99.5
<i>CNOT1</i>	100	99.3	100.0	100	20.2	92.5	100	93.9	99.9
<i>CNOT3</i>	100	72.4	94.5	100	24.5	95.7	100	89.4	96.9
<i>CNOT4</i>	100	99.7	100.0	100	26.5	95.3	100	95.2	99.9
<i>CRYAB</i>	100	93.0	100.0	100	22.5	95.4	100	100.0	100.0
<i>CSRP3</i>	100	100	100	100	25.2	96.2	100	97.4	99.9
<i>CTF1</i>	39.1	19.5	27.6	100	19.6	92.8	41.8	19.6	32.0
<i>DES</i>	99.5	58.9	79.8	100	19.6	94.6	100.0	64.1	87.5
<i>DMD</i>	100.0	97.9	100.0	100	27.9	96.1	100.0	91.5	99.4
<i>DNAJC1</i>	96.7	81.0	93.6	100	19.6	95.3	97.5	70.2	93.0
<i>DNAJC19</i>	100.0	98.0	100.0	100	23.4	95.6	100.0	47.9	97.9
<i>DPP6</i>	97.8	84.2	95.4	100	22.1	94.8	98.2	90.8	96.0
<i>DSC2</i>	100.0	96.8	98.9	100	30.8	97.4	100.0	76.2	98.8

<i>DSG2</i>	100.0	97.8	98.9	100	24.2	95.1	100.0	84.6	98.8
<i>DSP</i>	100.0	98.0	99.8	100	23.9	96.1	100.0	98.0	99.9
<i>DTNA</i>	100	0	99.4	100	22.4	94.7	100.0	98.9	100.0
<i>EMD</i>	100	68.2	93.9	100	23.9	93.9	100.0	80.8	95.9
<i>ENDOG</i>	72.0	43.8	51.7	100	29.6	96.5	79.1	44.0	60.6
<i>EYA4</i>	100	100.0	100.0	100	19.6	94.6	100.0	95.0	99.9
<i>F7</i>	100	45.4	96.1	100	25.4	95.7	100.0	84.4	97.9
<i>FBXO32</i>	100	85.6	99.7	100	28.3	96.1	100.0	96.5	99.7
<i>FHL2</i>	100	79.8	89.8	100	25.3	95.5	100.0	83.0	92.4
<i>FKRP</i>	97.0	29.4	63.8	100	24.6	96.4	97.0	45.1	77.8
<i>FKTN</i>	100	98.6	100.0	100	19.6	94.9	100.0	70.5	99.3
<i>FLT1</i>	100	96.5	98.8	100	29.6	95.6	100.0	95.9	99.1
<i>FOXD4</i>	93.3	22.0	59.2	100	20.8	94.8	95.7	29.7	66.9
<i>FOXO3</i>	83.1	45.9	63.2	100	20.0	93.1	84.8	48.7	63.9
<i>FXN</i>	85.0	0.0	75.3	100	29.0	96.1	89.5	73.6	76.7
<i>GAA</i>	100.0	0.0	92.3	100	26.6	94.9	100	39.5	90.8
<i>GATA4</i>	69.8	49.4	58.4	100	24.6	95.6	65.5	55.2	59.5
<i>GINS3</i>	100.0	77.9	92.9	100	26.9	96.4	100.0	76.8	90.2
<i>GLA</i>	100.0	0.0	99.4	100	25.0	95.9	100.0	92.6	99.7
<i>GPDI1L</i>	100.0	94.7	98.5	100	26.5	96.9	100.0	92.2	99.3
<i>GPR183</i>	100.0	100.0	100.0	100	26.0	96.0	100.0	99.2	100.0
<i>HADHA</i>	100.0	97.3	100.0	100	19.6	93.9	100.0	97.6	100.0
<i>HCN4</i>	91.6	0.0	76.4	100	24.5	96.1	88.5	55.5	80.0
<i>HFE</i>	100.0	91.8	99.9	100	19.6	95.3	100.0	98.4	100.0
<i>HNRNPK</i>	100.0	93.8	100.0	100	28.6	97.4	100.0	83.0	99.6
<i>HNRNPM</i>	100.0	67.0	97.2	100	21.0	94.4	100.0	83.7	98.3
<i>HNRNPU</i>	97.0	85.0	92.9	100	26.1	96.0	98.4	86.9	93.7
<i>HOPX</i>	100.0	65.3	99.0	100	29.5	96.9	100.0	78.1	98.1
<i>HSP90AA1</i>	100.0	96.9	99.7	100	19.6	94.5	100.0	93.5	99.7
<i>HSP90AB1</i>	100.0	99.3	100.0	100	27.0	94.8	100.0	94	99.9
<i>HSPB1</i>	99.4	42.3	83.1	100	19.6	93.2	100.0	74.1	90.5
<i>HSPB6</i>	93.5	9.1	41.0	100	34.5	97.2	94.4	21.5	52.7
<i>HSPB7</i>	100.0	8.4	98.9	100	19.6	94.1	100.0	91.5	99.2
<i>HSPB8</i>	100.0	0.0	98.5	100	19.6	91.9	100.0	89.5	98.4
<i>IL18</i>	100.0	98.1	100.0	100	19.6	93.3	100.0	13.8	98.3
<i>ILK</i>	100.0	98.6	100.0	100	23.4	95.4	100.0	98.6	99.9
<i>JPH2</i>	91.9	37.3	61.3	100	28.8	96.9	85.8	47.9	70.0
<i>JUP</i>	100.0	72.1	99.4	100	26.0	94.7	100	92.5	99.4
<i>KCNE1</i>	100.0	90.6	100.0	100	26.1	94.7	100	100	100.0
<i>KCNE2</i>	100.0	100.0	100.0	100	22.4	94.8	100	100	100.0
<i>KCNE3</i>	100.0	100.0	100.0	100	21.4	95.0	100	95.51	99.9
<i>KCNH2</i>	94.6	35.9	79.4	100	24.2	96.1	95.2	70.3	85.0
<i>KCNJ11</i>	100.0	57.0	99.8	100	31.1	96.2	100.0	91.9	99.5
<i>KCNJ2</i>	100.0	100.0	100.0	100	26.2	96.1	100.0	100.0	100.0
<i>KCNJ5</i>	100.0	0.0	99.4	100	23.4	95.3	100.0	97.5	99.7
<i>KCNJ8</i>	100	99.6	100.0	100	24.9	95.3	100.0	99.9	100.0
<i>KCNQ1</i>	89.4	50.1	85.7	100	23.1	95.3	89.6	77.1	87.3
<i>KCNQ2</i>	98.9	35.9	85.0	100	24.9	96.2	98.8	71.6	89.5
<i>KRAS</i>	100	0.0	99.3	100	23.7	95.2	100.0	52.6	97.4
<i>LAMA2</i>	100	99.4	100.0	100	29.6	96.7	100.0	95.6	99.9
<i>LAMA4</i>	100	97.1	100.0	100	22.8	95.5	100	95.1	99.9
<i>LAMP2</i>	100	0.0	98.0	100	31.9	95.2	100	84.6	96.1
<i>LDB3</i>	100	55.7	95.9	100	25.9	96.2	100	84.3	96.9
<i>LIPC</i>	100	87.8	99.8	100	23.8	95.4	100	93.2	99.5
<i>LITAF</i>	100	89.4	99.9	100	24.9	95.5	100	93.2	99.6

<i>LMNA</i>	100	50.9	96.8	100	29.2	96.2	100	85.4	97.9
<i>LSMI4A</i>	100	91.5	99.9	100	19.6	94.7	100	88.6	99.7
<i>MBNL2</i>	100	98.2	99.8	100	25.3	95.7	100	90.0	99.6
<i>MDM2</i>	100	94.4	99.9	100	19.6	92.6	100	64.5	98.4
<i>MYBPC3</i>	100	60.5	99.6	100	23.6	95.2	100	95.0	99.5
<i>MYH6</i>	100	90.0	96.8	100	24.1	95.1	99.1	93.7	96.9
<i>MYH7</i>	100	90.6	97.0	100	19.6	93.8	99.3	93.0	97.0
<i>MYL2</i>	100	96.6	100.0	100	21.0	94.1	100	100	100
<i>MYL3</i>	100.0	89.4	100.0	100	20.1	94.6	100	95.8	99.9
<i>MYLK2</i>	100.0	86.9	99.8	100	19.6	92.5	100	97.0	99.9
<i>MYOZ2</i>	100.0	100.0	100.0	100	23.7	95.6	100	92.6	99.9
<i>MYPN</i>	100.0	0.0	99.4	100	22.9	95.3	100	97.5	99.9
<i>NDRG4</i>	97.5	48.2	89.6	100	24.6	95.5	98.0	69.7	91.4
<i>NEXN</i>	100.0	95.5	99.7	100	28.5	96.3	100	63.2	98.5
<i>NFKB1</i>	100.0	99.5	100.0	100	23.1	95.2	100	89.7	99.7
<i>NOS1AP</i>	100.0	71.3	99.5	100	19.6	95.1	100	90.6	99.5
<i>NR1H2</i>	100.0	50.6	94.8	100	26.9	96.5	100	76.3	96.2
<i>NR1H3</i>	100.0	89.7	99.5	100	25.9	94.7	100	92.3	98.8
<i>PDLIM3</i>	100.0	92.3	99.9	100	23.9	95.3	100	97.4	100.0
<i>PFDN1</i>	100.0	94.7	100.0	100	20.5	94.5	100	78.7	99.7
<i>PFDN2</i>	100.0	90.2	99.9	100	26.0	95.8	100	99.8	100.0
<i>PFDN4</i>	100.0	89.2	99.2	100	28.4	96.4	100	47.0	97.0
<i>PFDN5</i>	100.0	99.2	100.0	100	25.1	95.2	100	92.3	99.6
<i>PKP2</i>	100.0	85.7	93.4	100	19.6	94.5	100	88.3	95.6
<i>PLEC</i>	98.2	28.4	79.1	100	22.4	92.7	98.4	61.9	86.3
<i>PLN</i>	100.0	100.0	100.0	100	27.7	96.6	100.0	100.0	100.0
<i>POLR2F</i>	100.0	86.5	97.9	100	24.4	94.9	100.0	91.6	98.1
<i>PRKAG2</i>	100.0	0.0	95.8	100	20.0	95.2	100.0	84.1	96.0
<i>PRPF8</i>	100.0	98.4	99.7	100	25.2	95.8	100.0	98.2	99.6
<i>PSEN1</i>	100.0	94.1	100.0	100	31.6	97.2	100.0	98.9	100.0
<i>PSEN2</i>	100.0	90.7	99.9	100	20.6	95.5	100.0	97.1	99.8
<i>PTBP1</i>	100.0	33.5	93.8	100	26.7	96.3	100.0	78.1	96.2
<i>PTPN11</i>	99.5	0.0	98.6	100	24.5	95.1	100.0	95.6	99.0
<i>RAF1</i>	100.0	18.7	99.5	100	21.0	95.7	100.0	99.9	100.0
<i>RANGRF</i>	100.0	82.1	99.9	100	22.3	93.5	100.0	100.0	100.0
<i>RBM20</i>	98.9	93.3	96.0	100	22.7	95.3	99.7	95.3	96.6
<i>RBM39</i>	100.0	98.6	100.0	100	24.7	96.9	100.0	86.2	99.7
<i>RBM4</i>	100.0	97.6	99.8	100	27.7	96.4	100.0	97.6	99.6
<i>RBM45</i>	100.0	79.3	99.9	100	26.4	95.6	100.0	74.2	99.3
<i>RBM46</i>	100.0	87.6	99.9	100	29.2	96.8	100.0	78.2	99.7
<i>RBM4B</i>	100.0	100.0	100.0	100	23.8	95.3	100.0	98.9	100.0
<i>RBM7</i>	100.0	94.9	100.0	100	32.9	97.0	100.0	93.9	99.8
<i>RNF207</i>	97.4	37.7	87.7	100	19.6	93.1	99.1	57.8	90.3
<i>RXRA</i>	98.1	50.9	96.2	100	24.4	95.2	98.0	79.6	96.2
<i>RXRB</i>	98.9	79.8	89.3	100	25.9	95.6	98.3	82.1	91.5
<i>RXRG</i>	100.0	99.0	100.0	100	29.0	96.4	100.0	98.5	100.0
<i>RYR2</i>	100.0	98.6	99.9	100	23.8	95.2	100.0	89.4	99.4
<i>SCN10A</i>	100.0	40.9	99.2	100	24.2	94.0	100.0	99.4	100.0
<i>SCN1B</i>	96.2	73.1	95.5	100	27.8	97.1	96.1	86.3	95.2
<i>SCN3B</i>	100.0	0.0	99.4	100	26.9	96.0	100.0	98.5	100.0
<i>SCN4B</i>	100.0	90.9	99.8	100	25.6	95.5	100.0	98.1	99.9
<i>SCN5A</i>	100.0	89.1	99.7	100	21.1	95.8	100.0	95.8	99.6
<i>SDHA</i>	98.9	90.1	95.1	100	19.6	92.2	99.6	87.4	95.1
<i>SELP</i>	100.0	100.0	100.0	100	21.5	94.5	100.0	98.8	100.0
<i>SERPINE1</i>	100.0	98.6	100.0	100	29.0	97.4	100.0	97.8	99.9

<i>SF3B3</i>	100.0	99.8	100.0	100	25.0	95.1	100.0	99.6	100.0
<i>SGCA</i>	100.0	83.9	99.8	100	20.6	95.0	100.0	97.8	99.9
<i>SGCB</i>	96.8	96.3	96.8	100	20.3	94.9	96.8	94.5	96.5
<i>SGCD</i>	100.0	94.4	100.0	100	23.0	94.5	100.0	95.0	99.9
<i>SGCG</i>	100.0	100.0	100.0	100	26.6	96.1	100.0	99.8	100.0
<i>SLC8A1</i>	100.0	0.0	99.4	100	22.6	95.2	100.0	99.3	99.9
<i>SNRNP70</i>	100.0	58.0	80.1	100	29.3	97.3	100.0	65.1	90.5
<i>SNRPA</i>	100.0	79.3	99.1	100	28.7	95.3	100.0	81.0	98.7
<i>SNRPB</i>	100.0	96.9	100.0	100	29.4	96.4	100.0	93.5	99.6
<i>SNTA1</i>	90.5	52.8	78.8	100	25.1	96.3	85.1	65.0	79.2
<i>SOD2</i>	100.0	74.6	92.3	100	22.3	95.0	100	87.0	95.4
<i>SOS1</i>	100.0	0.0	97.2	100	19.6	93.6	99.8	60.1	96.9
<i>SRL</i>	100.0	0.0	97.0	100	22.2	94.8	100.0	81.7	96.3
<i>STUB1</i>	100.0	48.3	85.7	100	27.1	95.2	100.0	69.2	92.0
<i>SYNE1</i>	100.0	99.8	100.0	100	30.8	97.4	100.0	94.3	99.8
<i>SYNM</i>	91.9	82.7	85.6	100	28.3	96.2	92.1	82.8	87.3
<i>TAZ</i>	98.8	0.0	86.2	100	29.2	95.9	97.7	62.3	85.9
<i>TBX20</i>	100.0	90.6	99.5	100	29.5	97.0	100.0	91.4	98.5
<i>TCAP</i>	100.0	21.7	98.6	100	28.4	96.9	100.0	89.7	99.0
<i>TCP1</i>	100.0	96.3	100.0	100	28.0	96.6	100.0	97.2	100.0
<i>TGFB3</i>	100.0	99.3	100.0	100	27.9	96.4	100.0	98.8	100.0
<i>TMEM43</i>	100	91.9	99.8	100	25.4	95.9	100.0	94.8	99.6
<i>TMPO</i>	99.0	95.3	97.8	100	27.0	95.7	99.9	88.9	98.1
<i>TNNC1</i>	100	99.6	100.0	100	29.0	96.2	100.0	99.6	100.0
<i>TNNI3</i>	100	65.3	98.8	100	21.9	94.8	100.0	87.4	98.7
<i>TNNT2</i>	100	96.0	100.0	100	23.4	95.0	100.0	99.7	100.0
<i>TP63</i>	100	0.0	97.9	100	19.8	95.2	100.0	91.4	97.2
<i>TPM1</i>	100	64.0	92.1	100	23.6	95.0	100.0	80.1	95.2
<i>TRIM54</i>	100	75.5	99.0	100	22.6	94.5	100.0	91.5	99.3
<i>TRIM55</i>	100	99.7	100.0	100	25.6	96.0	100.0	94.3	99.9
<i>TRIM63</i>	100	96.2	100.0	100	19.6	94.1	100.0	96.6	99.8
<i>TTN</i>	99.5	97.9	99.1	100	27.5	96.9	99.4	92.4	99.0
<i>TTR</i>	100	0.0	99.4	100	25.1	95.9	100	100.0	100.0
<i>UBE2D1</i>	100	78.8	98.2	100	23.1	95.6	100	43.0	96.7
<i>UBE2D2</i>	100	93.0	98.4	100	29.3	95.9	100	47.3	98.2
<i>UBE2D3</i>	100	94.5	100.0	100	27.4	96.7	100	50.5	99.0
<i>UBE2D4</i>	100	84.5	99.8	100	29.4	96.6	100	93.5	99.6
<i>UBE4A</i>	100	99.9	100.0	100	26.8	95.7	100	94.5	99.9
<i>UBE4B</i>	100	98.8	100.0	100	25.0	95.3	100	96.0	99.9
<i>UNC45B</i>	100	96.3	100.0	100	27.4	96.3	100	97.6	99.8
<i>VBPI</i>	100	86.2	99.8	100	29.5	96.4	100	70.9	98.3
<i>VCL</i>	100	95.1	99.7	100	23.7	95.3	100	96.1	99.6
<i>ZBTB17</i>	100	43.9	97.5	100	24.8	95.1	100	84.4	97.9

Figure 4-3 Formal assessment of multiple regression model predicting Left Ventricular Ejection Fraction (LVEF)



Formal assessment of multiple regression model predicting Left Ventricular Ejection Fraction (LVEF) by age, sex, ethnicity, total number of infarcted segments, number of hibernating segments, wall motion of most affected territory, hibernating segments, history of prior revascularization, mitral regurgitation and presence of a *TTNtv* to check that assumptions of linearity, randomness and homoscedasticity have been met. The model appears accurate and generalizable to the population.

Table 4-7 Metabolic investigations of subject 1 (affected child with Histiocytoid CM) showed some minor non-specific abnormalities

Metabolic Investigation	Age at time of investigation	Result
Urine Amino Acids	8/12	Mild to moderate non-specific increase in most amino acids consistent with a generalised aminoaciduria. ? Proximal tubule dysfunction
Plasma Amino Acids	8/12	Normal
Urine Organic acids	8/12	Moderately raised 2-oxoglutarate? Significant? Renal tubular leak, ? Secondary to disturbed carbohydrate metabolism. Suggested repeat measurements and plasma/blood lactate etc.
Repeat Urine Organic Acids	10/12	Normal 2-oxoglutarate on this occasion, with mildly raised suberate with mildly raised octanoate and 7-hydroxyoctanoate ? Secondary to medium-chain triglycerides therapy. Mildly raised pyruvate? Renal tubular leak
Carnitine profile	3/12	Normal
Lactate profile	Not available	Normal

Table 4-8 Investigations for mitochondrial DNA mutations (mtDNA) in skeletal muscle of subject 1 (affected child with Histiocytoid CM), showed no evidence of any abnormality

Test	Mutation/Gene	Disorder	Result
Direct sequencing of skeletal muscle DNA	MTCYB (cytochrome b of complex III)	Mitochondrial complex III deficiency	No mutation detected
mtDNA mutation in Skeletal muscle Pyrosequencing assay	m.3243A>G	Mitochondrial Encephalomyopathy, Lactic Acidosis and Stroke-Like Episodes (MELAS) or CPEO and Maternally-inherited diabetes and deafness (MIDD)	No mutation detected
Direct sequencing of the mitochondrial-encoded tRNA ^{Ile} gene	MTTI	HCM and myopathy	No mtDNA mutations in the MTTI gene
? mtDNA rearrangement DNA samples from skeletal muscle screened for possible mtDNA rearrangements using long-range PCR assay spanning the majority of the mitochondrial genome	Screening of affected tissues for mitochondrial DNA (mtDNA) rearrangement disorders, including both sporadic, large scale single mtDNA deletions and nuclear-driven, multiple mtDNA deletions		No evidence of major mtDNA rearrangements in the tissue

Table 4-9 Further histopathology, immunohistochemistry and immunoblotting investigations in the skeletal muscle of subject 1 (affected child with Histiocytoid CM) showing no specific abnormality

Type of testing	Result	Interpretation
Histopathology of skeletal muscle biopsy	Immunoblot for PTRF and a limited panel of immunolabelling was undertaken	These sections were affected by mechanical artefacts
Immunohistochemistry	Labelling for β -spectrin was very weak and patchy whereas, taking into account the mechanical artefacts present, labelling for caveolin 3 appeared essentially normal and labelling for α -dystroglycan showed some very mild variation	No gross abnormalities were seen in the limited labelling undertaken. In the absence of age matched controls in the immunohistochemistry labelling is difficult to put into context
Immunoblotting	The band PTRF was indistinguishable from the normal control	

Table 4-10 Primers used for Sanger sequencing of *NDUFB11* and *FAM135A*

Gene	Chromosome	Forward primer	Reverse primer	Start	End	Size	Exon	Name
<i>NDUFB11</i>	X	<i>TTGAGACATTGCACCAACCC</i>	<i>CTCTTTGCCTCCAGTTGTCTG</i>	47004236	47005010	775bp	1	NDUFB11_1A
		<i>TACTAAGGAAGCGCGCTCTG</i>	<i>AAGAACGTCTCTACAGCGC</i>	47003778	47004623	846bp	1	NDUFB11_1B
		<i>TCCAGCCATGACTAGAGCTG</i>	<i>TCATCTCAGCTCCCCATTCC</i>	47001896	47002352	457bp	2	NDUFB11_2
		<i>AATGGGGAGCTGAGATGAGG</i>	<i>AGTCCCAGCCACCATCTAAC</i>	47001372	47001913	542bp	3	NDUFB11_3
<i>FAM135A</i>	6	<i>GTTGTACTGCAGCCTTGTAATAAAC</i>	<i>CAGCCTGAAGAACCATGACC</i>	71186519	71187151	633bp	8	FAM135A_8

Table 4-11 The two apparently *de novo* protein altering variants* identified in the affected child (subject 1) with Histiocytoid CM and absent in the unaffected parents which did not validate, from the 'trio' approach to WES

Genes	Chromosome	Genomic Position	Reference	Alternate	Category of Variation	Comment
HLA-DQA1	6	32609312	A	C	Missense	The variant was homozygous in the child (in both WES samples), but is also present in both parents and so not <i>de novo</i> . The complexity of HLA genes makes read-alignment extremely challenging; many short sequencing reads from HLA genes are either not properly mapped or are labelled as unmapped because of significant allelic differences and sequence similarity between paralogous genes (254).
CD163L1	12	7521563	T	TGA	Frameshift	The variant occurs in a very repetitive region, which may cause polymerase jumping. Both parents also have evidence of the same frameshift – either T> GA or T>GAGA, but at a lower allelic balance than in the child so the variant was called as <i>de novo</i> . Either the variant is not real in the child (although it was also present in the other WES data of subject 1) or it was inherited.

* Nonsense, frameshift, essential splice site, missense and in-frame. Variants did not validate by visual inspection of sequencing reads using Integrative Genome Viewer (IGV) in both the child and both parents.

Table 4-12 All non-synonymous protein coding mtDNA variants identified in mtDNA genes which had either variants in either two, three or four of the overlap cohort of four unrelated individuals with Histiocytoid CM

Mitochondrial DNA locus	Variant (position According to HG19)	Nucleotide position according to HG38 (ChrM)	Nucleotide Change	Frequency in MITOMAP (GB sequences)	Sample ID			
					20PO03585	20EA03620	20EM03618	20TH03619
ATP6	ChrM:8702 G >A	ChrM:8701	A-G	10072	A/A		A/A	
CYB	ChrM:14794 A >G	ChrM:14793	A-G	713			G/G	
	ChrM:15219 A >G	ChrM:15218	A-G	615			G/G	
	ChrM:15432 G >A	ChrM:15431	G-A	499				A/A
	ChrM:15453 C>A	ChrM:15452	C-A	2888	A/A			
ND1	ChrM:3309 T >G	ChrM:3308	T-G	4 *	G/G			
	ChrM:4217 T >C	ChrM:4216	T-C	3064	C/C			
ND3	ChrM:10399 G >A	ChrM:10398	A-G	13400	A/A		G/A	
ND5	ChrM:12851 G>A	ChrM:12850	A	12	A/A	A/A	A/A	A/A

Both the variant position according to HG19 and as derived from HG38 are given (see text for details). Where a non-reference call was made, this is indicated, along with the sample in which it was found. The frequency of each variant in MITOMAP GenBank (<http://www.mitomap.org/MITOMAP>) is included.

Table 4-13 Permissions for figures reproduced within thesis

Page No.	Type of work:	Name of work	Source of work	Copyright holder and contact	Permission requested	Permission yes /no	Permission note
29	Figure	Figure 1-1 Next Generation Sequencing workflow	Ware JS, Roberts AM, Cook SA. Next generation sequencing for clinical diagnostics and personalised medicine: implications for the next generation cardiologist. <i>Heart</i> . 2012;98(4):276-81.	BMJ Publishing Group Ltd. Contact Rightslink Copyright Clearance Center.	17/8/16	yes	License number 3931420488829
32	Figure	Figure 1-2 Variant consequences predictions	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. <i>Bioinformatics</i> . 2010;26(16):2069-70.	Oxford University Press. Contact Rightslink Copyright Clearance Center.	17/8/16	yes	License number 3931421313166
42	Figure	Figure 1-3 Mitochondrial DNA heteroplasmy and the threshold effect	Stewart JB, Chinnery PF. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. <i>Nature Reviews Genetics</i> . 2015;16(9):530-42.	Nature Publishing Group. Contact Rightslink Copyright Clearance Center.	17/8/16	yes	License number 3931430365650
43	Figure	Figure 1-4 The Human Mitochondrial Respiratory Chain	Schon EA, DiMauro S, Hirano M. Human mitochondrial DNA: roles of inherited and somatic mutations. <i>Nat Rev Genet</i> . 2012;13(12):878-90.	Nature Publishing Group. Contact Rightslink Copyright Clearance Center.	17/8/16	yes	License number 3931440076335
53	Figure	Figure 1-5 The role of TTN in the sarcomere	McNally EM, Golbus JR, Puckelwartz MJ. Genetic mutations and mechanisms in dilated cardiomyopathy. <i>J Clin Invest</i> . 2013;123(1):19-26.	Journal of Clinical Investigation. Online Contact Rightslink Copyright Clearance Center.	17/8/16	yes	License number 3931440717739

Table 4-13 continued

74	Figure	Figure 2-3 Standard 17 segment Cardiac Model	Cerqueira MD, Weissman NJ, Dilsizian V, Jacobs AK, Kaul S, Laskey WK, et al. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart - A statement for healthcare professionals from the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association. Circulation. 2002;105(4):539-42.	Wolters Kluwer Health, Inc. Contact Rightslink Copyright Clearance Center.	17/8/16	yes	License number 3931441012978
75	Figure	Figure 2-4 Coronary Artery Territories	Cerqueira MD, Weissman NJ, Dilsizian V, Jacobs AK, Kaul S, Laskey WK, et al. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart - A statement for healthcare professionals from the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association. Circulation. 2002;105(4):539-42.	Wolters Kluwer Health, Inc. Contact Rightslink Copyright Clearance Center.	17/8/16	yes	License number 3931441012978
117	Figure	Figure 3-1(A) Histology of Histiocytoid Cardiomyopathy	Ashworth M. Cardiomyopathy in Childhood: Histopathological and Genetic Features. The Open Pathology Journal,. 2010;4:80-93.	Open journal	17/8/16	yes	Permission sought from author
117	Figure	Figure 3-1(B) Histology of Histiocytoid Cardiomyopathy	Cabana MD, Becher O, Smith A. Histiocytoid cardiomyopathy presenting with Wolff-Parkinson-White syndrome. Heart. 2000;83(1):98-9.	BMJ Publishing Group Ltd Contact Rightslink Copyright Clearance Center.	17/8/16	yes	License number 3931460138024
118	Figure	Figure 3-2(A) Macroscopic findings in Histiocytoid Cardiomyopathy	Cabana MD, Becher O, Smith A. Histiocytoid cardiomyopathy presenting with Wolff-Parkinson-White syndrome. Heart. 2000;83(1):98-9.	BMJ Publishing Group Ltd Contact Rightslink Copyright Clearance Center.	17/8/16	yes	License number 3931460138024

Table 4-13 continued

118	Figure	Figure 3-2(B) Macroscopic findings in Histiocytoid Cardiomyopathy	Siehr SL, Bernstein D, Yeh J, Berry GJ, Rosenthal DN, Hollander SA. Orthotopic heart transplantation in two infants with histiocytoid cardiomyopathy and left ventricular non-compaction. <i>Pediatr Transplant</i> . 2013;17(7):E165-7.	John Wiley and Sons Contact Rightslink Copyright Clearance Center.	17/8/16	yes	License number 3931460547748
129	Figure	Figure 3-4 Histology of native heart from proband (Subject 1) with Histiocytoid CM	Rea G, Homfray T, Till J, Roses-Noguer F, Buchan RJ, Wilkinson S et al. Histiocytoid cardiomyopathy and microphthalmia with linear skin defects syndrome: phenotypes linked by truncating variants in <i>NDUFB11</i> . <i>Cold Spring Harb Mol Case Stud</i> 2017; 3(1)	Open access under the Creative Commons Attribution License	Not applicable	Not applicable	Not applicable
156	Figure	Figure 3-13 The KEGG Pathway of Oxidative Phosphorylation	KEGG pathways URL given	Not applicable	Not applicable	Not applicable	Not applicable
160	Figure	3-15 Typical skin lesions and eye features of patients with MLS syndrome	van Rahden VA, Rau I, Fuchs S, Kosyna FK, de Almeida HL, Jr., Fryssira H, et al. Clinical spectrum of females with HCCS mutation: from no clinical signs to a neonatal lethal form of the microphthalmia with linear skin defects (MLS) syndrome. <i>Orphanet J Rare Dis</i> . 2014;9:53.	BioMedCentral Ltd	Not applicable	Not applicable	Open access item
170	Figure	3-17 Schematic representation of X Chromosome inactivation in female somatic cells in MLS syndrome	Morleo M, Franco B. Dosage compensation of the mammalian X chromosome influences the phenotypic variability of X-linked dominant male-lethal disorders. <i>J Med Genet</i> . 2008;45(7):401-8.	BMJ Publishing Group Ltd Contact Rightslink Copyright Clearance Center.	17/8/16	yes	License number 3931460804812

i