

Congenital Heart Disease Gene Identification by Whole Exome Sequencing

Dr. Verity Laura Hartill

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
Leeds Institute of Biomedical and Clinical Sciences
Faculty of Medicine and Health

August 2017

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Results from Chapter 4 appear in the publication:

Sudden Cardiac Death due to Deficiency of the Mitochondrial Inorganic Pyrophosphatase PPA2. Kennedy H., Haack TB., Hartill VL., *et al.* American Journal of Human Genetics. 2016 Sep 1;99(3):674-82.

The candidate recruited and consented the described family (Family CHD2) to the project, undertook a full medical history, obtained appropriate DNA samples, and undertook whole exome sequencing library preparation and data analysis (including analysis for CNVs and autozygous regions). This led to the identification of *PPA2* as the most likely candidate gene. She then formed a collaboration with an international group of researchers who had recognized mutations in the same gene in three other families. The functional studies undertaken in this paper were not performed by the candidate, but by the other researchers involved in the collaboration, as stated in the text. This included the Western blot, Seahorse assays and pyrophosphatase assays described in this chapter. The candidate was involved in writing some parts of the paper, including clinical descriptions of the family and methods of exome analysis and checking drafts of the final manuscript.

Results from Chapter 5 appear in the publication:

DNAAF1 Links Heart Laterality with the AAA+ ATPase RUVBL1 and Ciliary Intraflagellar Transport. Hartill VL., van de Hoek G., Patel MP., *et al.* Manuscript submitted to Human Molecular Genetics, August 2017.

The candidate undertook whole exome sequencing library preparation and data analysis (including analysis for CNVs and autozygous regions) for the family described in this project (Family CHD1). This led to the identification of *DNAAF1* as the most likely candidate gene and the candidate undertook variant interpretation and arranged further clinical visits with the family. *DNAAF1* mutant and wildtype constructs (FLAG-tagged, N-TAP and PCS2+) were created by the candidate, using gateway cloning techniques. The candidate performed localization and co-localization experiments in RPE1-cells lines, using immunofluorescence microscopy, and siRNA knockdown

experiments, including all confirmations and data analysis. The candidate was involved in mouse embryo dissection and immunofluorescence, with close guidance from Ms. Dena Abdelmottaleb and Prof. Colin Johnson, University of Leeds. TAP experiments were performed by Dr K. Szymanska, University of Leeds and the candidate was involved in TAP data analysis. Dr. M. Patel, University College London, performed the majority of CoIP experiments described in this chapter, but the candidate performed CoIPs with endogenous DNAAF1 pull-down. All of the zebrafish experiments described in this chapter were performed by collaborators in Utrecht, The Netherlands (Dr. G. van de Hoek and Prof. R. Giles). The candidate was involved in oversight of the project, wrote the initial drafts of the paper and was involved in writing the final manuscript.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

© 2017 The University of Leeds and Dr. Verity Laura Hartill

Acknowledgements

This research has been carried out by a team which has included other researchers as detailed below. My own contributions, fully and explicitly indicated in the thesis, have been ascertainment, consent and clinical examination of the families described in this thesis, whole exome sequencing library preparation and data analysis, analysis of whole exome sequencing data for autozygous regions and CNVs, data quality analysis, analysis of whole genome sequencing at low read depth, DNA extraction from cell cultures and paraffin-embedded slides, PCR and Sanger sequencing, gateway cloning and site-directed mutagenesis, over-expression experiments in cell-lines, siRNA knockdown experiments in cell-lines, immunofluorescence and confocal microscopy, and, in close collaboration with other researchers, mouse embryo dissection, mouse embryo whole mount immunofluorescence and light-sheet microscopy.

The other members of the group and their contributions have been as follows:

Dr. H. Kennedy, Prof J. Mayr, Paracelsus Medical University (PMU) Salzburg, Austria and colleagues: Pyrophosphatase assays described in Chapter 4.

Dr. K. Szymanska, Post-Doctoral Research Fellow, Department of Ophthalmology and Neuroscience, University of Leeds: TAP experiments described in Chapter 5.

Dr. M. Patel, Post-Doctoral Research Fellow, University College London: Co-immunoprecipitation experiments described in Chapter 5.

Ms D. Abdelmottaleb, Visiting Researcher, Department of Ophthalmology and Neuroscience, University of Leeds: Mouse embryo dissection and immunofluorescence described in Chapter 5.

Dr. G. van de Hoek and Prof. R. Giles, University Medical Center, Utrecht, The Netherlands: Zebrafish whole mount in situ hybridization described in Chapter 5.

The Leeds Sequencing Facility, Leeds Teaching Hospitals NHS Trust: Preparation of samples for Whole Genome Sequencing at low read depth and WGS and WES sample sequencing following library preparation on Illumina HiSeq 2500 and 3000.

Dr. C. Watson, Leeds Teaching Hospitals NHS Trust: Whole genome sequencing and Sanger breakpoint analysis in mapping an identified deletion in *DNAAF1* described in Chapter 5.

Dr L. Southgate: Sanger sequencing of variants identified in family CHD9b, described in Chapter 6.

Dr C. Bennett, Dr A. Dobbie, Prof. E. Sheridan, Dr K. Prescott, Dr. E. Hobson, Dr. R. Jewell, Ms K. Ashcroft (Department of Clinical Genetics, Leeds Teaching Hospitals NHS Trust), Dr K. English, Dr. J. Oliver, Dr. E. Brown, Dr. S. Barwick (Department of

Cardiology, Leeds Teaching Hospitals NHS Trust), Prof. J. Goodship (Institute for Human Genetics, Newcastle-Upon-Tyne), Dr. A. Brady, Dr. N. Canham (Northwest Thames Regional Genetics Center, London Northwest Healthcare NHS Trust): Referral of patients to the project.

I would like to thank my supervisors, Prof. Colin Johnson and Prof. Eamonn Sheridan, for their patience, dedication, knowledge and oversight in supporting me through this process. Thanks must also go to the other members of Team Meckel and members of the department of Ophthalmology and Neuroscience, who have taught me all of the included laboratory methods. In particular I would like to thank Kasia Szymanska, Clare Logan, Dave Parry and James Poulter who have guided this journey from clinician to aspiring scientist.

I have huge gratitude to the families who have been involved in this project and, without whom, this thesis would not exist. Many thanks go to our collaborators who have supported this project with time and resources and have allowed open and honest discussion to move these projects forward towards publication. Thanks must go to the British Heart Foundation for funding this work.

Finally I would like to thank my long-suffering husband, Andy, and treasured son, Charlie, for their support over the past 4 years.

Abstract

Congenital Heart Disease (CHD) is the most common congenital defect, but the genetic aetiology of a large proportion of CHD is unexplained. This project aimed to delineate novel genetic causes of CHD using Whole Exome Sequencing (WES) in a family-based approach.

Sixteen families were recruited to the study. WES data analysis followed a standardized pipeline and candidate variants were prioritized on the basis of *in silico* pathogenicity prediction tools and review of the current literature.

Known and candidate genes in CHD were successfully identified using WES. In one family, a mutation in *PIGV* was identified, providing a diagnosis of Hyperphosphatasia and Mental Retardation syndrome and expanding the known phenotype of this condition. In a family with early-onset cardiomyopathy, a mutation in *PPA2* was identified, encoding a mitochondria-specific pyrophosphatase enzyme. Through collaboration this gene was identified to be causative in three further families and mutation pathogenicity was investigated by functional studies.

In a further family, a missense mutation in *DNAAF1* was associated with heterotaxy, in the absence of clinical features of Primary Ciliary Dyskinesia, the phenotype usually associated with this gene. Zebrafish studies supported the pathogenicity of this variant and functional experiments identified novel interactants of DNAAF1 to include Pontin, Reptin and IFT88. Pontin was found to be expressed on the left side of the embryonic node in mice and zebrafish, a pattern which was abolished in *dnaaf1*^{-/-} mutant fish, suggesting DNAAF1 and Pontin to be involved in the development of early laterality. In two families with athelia, choanal atresia and CHD a candidate variant in *KMT2D* was identified. The phenotype was distinct from Kabuki syndrome and is likely to represent a novel *KMT2D*-related disorder.

WES was a successful tool in gene identification in CHD and, coupled with functional studies, has provided novel insights into the pathogenesis of CHD.

Contents

Acknowledgements	iii
Abstract	v
Contents.....	vi
List of Figures	xiv
List of Tables.....	xvii
Abbreviations	xix
1 Introduction.....	1
1.1 General Introduction.....	1
1.2 Formation of the Foetal Heart.....	2
1.3 Classification of Congenital Heart Disease.....	4
1.4 Management of CHD.....	5
1.5 The Known Genetic Aetiology of Congenital Heart Disease	6
1.5.1 Copy Number Variation	6
1.5.2 Single Gene Variants.....	8
1.5.2.1 Transcription Factors	8
1.5.2.2 Signaling Pathways.....	10
1.5.2.3 Laterality in the Early Embryo	11
1.5.2.4 Variants in Structural Genes.....	11
1.5.3 Epigenetics	12
1.5.4 MicroRNAs.....	14
1.6 Cilia and CHD.....	16
1.6.1 Cilia	16
1.6.2 Defects in Primary Cilia are Associated with CHD	17
1.6.3 Defects in Motile Cilia are Associated with CHD.....	17
1.7 Gene Identification in Congenital Heart Disease	18
1.7.1 Next Generation Sequencing Technologies.....	18
1.7.2 Whole Exome Sequencing.....	19
1.7.2.1 Whole Exome Sequencing in Consanguineous Unions.....	20
1.7.3 Whole Genome Sequencing.....	21
1.7.4 GWAS	21
1.8 Clinical Utility of Gene Identification	22
1.9 Aims.....	23

2	Materials and Methods	25
2.1	Materials	25
2.1.1	General Reagents	25
2.1.2	Solutions	25
2.1.2.1	PBST	25
2.1.2.2	Tris-EDTA (TE) Buffer 1x	25
2.1.2.3	Tris-Acetate-EDTA (TAE) Buffer 50x	25
2.1.2.4	Tris-Borate-EDTA (TBE) Buffer 10x	25
2.1.2.5	Gel Loading Dye 2x	25
2.1.2.6	NP40 Lysis Buffer	26
2.1.2.7	IP Wash Buffer	26
2.1.2.8	IP Incubation Buffer	26
2.1.2.9	SDS Loading Buffer	26
2.1.2.10	Lysis Buffer for Cellular DNA Extraction	26
2.1.2.11	TAP Wash Buffer	26
2.1.2.12	D-desthiobiotin Elution Buffer	26
2.1.3	Cell Lines	27
2.1.4	Primary Antibodies	27
2.2	Methods	29
2.2.1	Patient Identification	29
2.2.2	Consent	29
2.2.3	Ethical Approval	30
2.2.4	Clinical Assessment	30
2.2.5	Clinical Investigation for PCD	30
2.2.6	DNA Sampling	30
2.2.7	DNA Extraction	30
2.2.7.1	Peripheral Blood Samples	30
2.2.7.2	Saliva Samples	31
2.2.7.3	Cell Culture	31
2.2.7.4	Paraffin-Embedded Slides	31
2.2.8	Whole Genome Amplification	31
2.2.9	RNA extraction	32
2.2.10	Polymerase Chain Reaction	32
2.2.10.1	Primer Design	32
2.2.10.2	PCR Reaction	32
2.2.10.3	Agarose Gel Electrophoresis	33
2.2.11	Purification with ExoSAP	33
2.2.12	Sanger Sequencing	33

2.2.13	Genotyping with SNP array	34
2.2.14	Whole Exome Sequencing	34
2.2.14.1	Sure Select XT method	35
2.2.14.2	Sure Select QXT Method	38
2.2.14.3	WES Data Analysis	39
2.2.14.4	Assessment of Mutation Pathogenicity	42
2.2.14.5	WES Assessment of Quality of Data	43
2.2.14.6	Autozygosity Mapping from WES Data	43
2.2.14.7	Copy Number Variation from WES Data	44
2.2.15	Whole Genome Sequencing at Low Read Depth	44
2.2.16	Whole Genome Sequencing and Breakpoint Analysis	45
2.2.17	Microbiology	46
2.2.17.1	Gateway Cloning	46
2.2.17.2	Bacterial Transformation	48
2.2.17.3	Site-Directed Mutagenesis	48
2.2.17.4	Mini Preps of Plasmid DNA	49
2.2.17.5	Maxi Preps of Plasmid DNA	49
2.2.18	Restriction Enzyme Digest	50
2.2.19	Cell Culture	50
2.2.20	Cell Passage and Harvesting	50
2.2.21	Transfection	51
2.2.21.1	Over-expression	51
2.2.21.2	siRNA Knockdown	51
2.2.22	Immunofluorescence and Confocal Microscopy	52
2.2.23	Western Blotting	52
2.2.24	Co-Immunoprecipitation	54
2.2.25	Tandem Affinity Purification	54
2.2.26	Mouse Embryo Dissection	55
2.2.27	Mouse Embryo Whole Mount Immunofluorescence	55
2.2.28	Light Sheet Microscopy	56
2.2.29	Statistical Tests	56

3	Whole Exome Sequencing can Identify Known and Candidate Genes in Families with Congenital Heart Disease	57
3.1	Introduction	57
3.2	Overall Results	57
3.2.1	Patient Recruitment and Phenotyping	57
3.2.2	Whole Exome Sequencing	58

3.2.2.1	Assessment of WES Library Preparation	58
3.2.2.2	WES Data Quality Scores	60
3.2.2.2.1	Depth of Coverage	60
3.2.2.2.2	PICARD's Collect Multiple Metrics	62
3.3	Family with a Mutation in a Known Gene	65
3.3.1	Family CHD14	65
3.3.1.1	Clinical Phenotype.....	65
3.3.1.1.1	Autozygosity Mapping.....	65
3.3.1.1.2	Whole Exome Sequencing.....	66
3.3.1.2	Sanger Sequencing.....	68
3.3.1.3	Discussion of Candidate Variants	68
3.4	Families Affected with CHD	69
3.4.1	Clinical Phenotypes.....	69
3.4.2	Whole Exome Sequencing.....	79
3.4.3	CNV analysis.....	86
3.4.3.1	Whole Genome Sequencing at Low Read Depth	86
3.4.3.2	CNV Analysis Using Exome Depth	88
3.4.4	Comparison of Variants to Ciliopathy Databases	88
3.4.5	Fetal RNA Data Analysis	89
3.4.6	Sanger sequencing	90
3.4.7	Discussion of Families with CHD	92
3.4.7.1	Discussion of Recruited Families.....	92
3.4.7.2	Discussion of Identified Candidate Variants in CHD Families.....	94
3.4.7.3	Family CHD3	94
3.4.7.4	Family CHD4.....	96
3.4.7.5	Family CHD5	97
3.4.7.6	Family CHD6.....	98
3.4.7.7	Family CHD7	99
3.4.7.8	Family CHD8.....	102
3.4.7.9	Family CHD12	102
3.4.7.10	Family CHD13.....	102
3.4.7.11	Family CHD15.....	103
3.4.7.12	Family CHD16.....	104
3.4.7.13	Overlapping Findings in CHD Patients	104
3.5	Patients with Laterality Disorders from the DDD Project	105
3.5.1	Introduction.....	105
3.5.2	Results	105
3.5.2.1	Clinical Summary.....	105

3.5.2.2	Identification of Known Disease-Causing Genes.....	105
3.5.2.3	Candidate Laterality Variants	108
3.5.3	Discussion of Findings from the DDD Project	108
3.6	Conclusions.....	109
4	<i>PPA2</i> Mutations Cause Early Onset Cardiomyopathy and Sudden Infant	
Death	110
4.1	Introduction	110
4.2	Results	110
4.2.1	Clinical Phenotype.....	110
4.2.2	Whole Exome Sequencing.....	112
4.2.3	Autozygosity Mapping.....	115
4.2.4	CNV analysis.....	116
4.2.5	Bioinformatic Analysis of the <i>PPA2</i> variant	116
4.2.6	Sanger Sequencing.....	118
4.2.7	Identification of Further Families with Mutations in <i>PPA2</i>	118
4.2.8	Mutation Analysis.....	120
4.2.9	Pyrophosphatase Activity	123
4.3	Discussion	125
4.3.1	Candidate Variants in Sudden Unexpected Death with Cardiomyopathy	125
4.3.2	Mitochondrial Function	126
4.3.3	The Function of <i>PPA2</i>	127
4.3.4	Discussion of Functional Studies Performed in <i>PPA2</i>	128
4.3.5	Identified Mutations in <i>PPA2</i> are Hypomorphic	129
4.3.6	How is <i>PPA2</i> Dysfunction Linked to the Clinical Phenotype?.....	130
4.3.7	The Importance of the Identification of Mutations in <i>PPA2</i>	131
4.4	Conclusions.....	132
5	Mutations in <i>DNAAF1</i> are Associated with Congenital Heart Disease in the	
Absence of Primary Ciliary Dyskinesia.....		134
5.1	Introduction	134
5.2	Results	135
5.2.1	Clinical Phenotype.....	135
5.2.2	SNP microarray	136
5.2.3	Whole Exome Sequencing.....	136
5.2.4	CNV analysis.....	139
5.2.5	Analysis of the Missense Variant in <i>DNAAF1</i>	139

5.2.6	Sanger Sequencing Confirmation of Variant in <i>DNAAF1</i>	141
5.2.7	Investigation of PCD in Family CHD1.....	141
5.2.8	Investigation of Mutations in <i>DNAAF1</i> in Other Heterotaxy Patients.....	142
5.2.9	Identification of Further Families with Mutations in <i>DNAAF1</i> Through Collaborating Centres.....	143
5.2.10	Creation of <i>DNAAF1</i> Constructs	146
5.2.11	Rescue of Heart Looping Defect in <i>dnaf1</i> ^{-/-} Zebrafish	148
5.2.12	Tandem Affinity Purification to Identify Interacting Partners of <i>DNAAF1</i> .	149
5.2.13	Co-Immunoprecipitation to Confirm Interactants of <i>DNAAF1</i>	152
5.2.14	Cellular Localization Studies to Assess Co-localisation of <i>DNAAF1</i> with Identified Interacting Partners	155
5.2.15	siRNA Knockdown of Pontin (RUVBL1) and Reptin (RUVBL2).....	160
5.2.16	Mouse Embryo Whole Mount Immunofluorescence	163
5.2.17	Expression of <i>DNAAF1</i> and Pontin at Kupffer's Vesicle in Zebrafish Embryos 165	
5.3	Discussion	168
5.3.1	Discussion of Candidate Variants in Family CHD1.....	168
5.3.2	Primary Ciliary Dyskinesia is a Disease of the Motile Cilia.....	168
5.3.3	The Link between Primary Ciliary Dyskinesia and Congenital Heart Disease 169	
5.3.4	The Known Functions of <i>DNAAF1</i> and Its Role in PCD	169
5.3.5	A Phenotypic Continuum Between CHD and PCD	171
5.3.6	Novel Interactants of <i>DNAAF1</i> and Insights into its Function	172
5.3.7	Asymmetry of Pontin at the Embryonic Node	175
5.4	Conclusions.....	177
6	A Variant in <i>KMT2D</i> is Associated with a Syndrome of Athelia, Choanal atresia and Congenital Heart Disease	178
6.1	Introduction	178
6.2	Results	178
6.2.1	Family CHD9	178
6.2.1.1	Clinical Phenotype.....	178
6.2.1.2	Whole Exome Sequencing.....	179
6.2.1.3	CNV analysis	182
6.2.1.4	Confirmation of Variant in <i>KMT2D</i> by Sanger Sequencing.....	182
6.2.2	Identification of a Second Family affected with Athelia and Choanal Atresia: Family CHD9b.....	183

6.2.2.1	Clinical Phenotype.....	183
6.2.2.2	Data analysis.....	183
6.2.2.3	Sanger Sequencing.....	184
6.2.3	Family CHD10	184
6.2.3.1	Clinical Phenotype.....	184
6.2.3.2	Whole Exome Sequencing Data Analysis	185
6.2.4	Family CHD11	185
6.2.4.1	Clinical Phenotype.....	186
6.2.4.2	DNA Extraction.....	186
6.2.4.3	Whole Exome Sequencing.....	187
6.2.4.4	Sanger sequencing	187
6.3	Discussion	188
6.3.1.1	Clinical Reports of Syndromes Consisting of Athelia and Choanal Atresia	188
6.3.1.2	<i>KMT2D</i> is a Known Cause of Kabuki Syndrome	193
6.3.1.3	Mutations in <i>KMT2D</i> May Cause a Phenotype of Choanal Atresia, Athelia and Other Congenital Abnormalities, that is Distinct to Kabuki Syndrome	194
6.3.1.4	<i>PTPRM</i> as a Candidate Gene for Athelia.....	196
6.4	Conclusions.....	197
7	Discussion.....	198
7.1	Summary of Key Findings	198
7.2	The Difficulties of Candidate Gene Identification in CHD	200
7.3	WES Filtering Strategies.....	201
7.4	The Impact of Understanding the Genetic Cause of Disease	203
7.5	WES in the Clinical Setting	206
7.6	The Future of CHD Genetics	207
7.7	Potential Therapeutics in Genetic Disease	211
7.8	Plans for Future Research	214
7.9	Overall Conclusions	215
	References	216
	Appendix A	241
	Appendix B	242
	Appendix C.....	245
	Appendix D.....	247
	Appendix E.....	252

Appendix F	254
Appendix G	256
Appendix H	257
Appendix I	258
Appendix J	263
Appendix K	276
Appendix L	277
Appendix M	278
Appendix N	279

List of Figures

Figure 1-1: The process of cardiac embryology in human and mouse.	3
Figure 1-2: Transcription factors involved in region-specific cardiac development.	10
Figure 1-3: The major known structures of the primary cilium.	16
Figure 2-1: Agilent SureSelect XT workflow for Paired-end Sequencing, DNA shearing to exome capture with addition of biotinylated RNA library baits.	37
Figure 2-2: Illumina sequencing protocol, showing cluster amplification on a flow cell.	38
Figure 2-3: Data analysis pipeline used for WES data.	42
Figure 2-4: Gateway cloning methods.	47
Figure 3-1: Bioanalyzer results for samples JT785 and JT380.	59
Figure 3-2: Bioanalyzer results from sample JT584 prepared using the QXT method.	59
Figure 3-3: Bioanalyzer results from sample JT624 prepared using the QXT method.	59
Figure 3-4: Mean depth of coverage across the exome comparing different library preparation, input and sequencing methods.	61
Figure 3-5: Percentage duplication according to sample number and WES method.	62
Figure 3-6: A & B: Quality scores for sample JT380 from Picard's Collect Multiple Metrics Tool.	63
Figure 3-6: C & D: Quality scores for sample JT380 from Picard's Collect Multiple Metrics Tool.	64
Figure 3-7: Pedigree of family CHD14.	65
Figure 3-8: Pedigrees of families A: CHD3 and B: CHD4.	74
Figure 3-9: Pedigree of family CHD5.	75
Figure 3-10: Pedigree of family CHD6.	76
Figure 3-11: Pedigrees of families A: CHD7 and B: CHD8.	77
Figure 3-12: Pedigrees of families A: CHD12, B: CHD13, C: CHD15 and D: CHD16.	78
Figure 3-13: Sanger sequencing electropherograms to demonstrate a frameshift variant, c.133_134delAA, in <i>C2ORF88</i> in family CHD7.	91
Figure 3-14: Sanger sequencing electropherograms to demonstrate the c.3682G>A variant in <i>RREB1</i> in family CHD7.	92
Figure 4-1: Pedigree of family CHD2.	111

Figure 4-2: IGV read-out showing sequencing reads aligned to the reference genome (GRCh37) in family CHD2.....	115
Figure 4-3: Evolutionary conservation of p.P167L in vertebrates.....	117
Figure 4-4: Output from the String database to summarize a query of interacting partners of PPA2.....	117
Figure 4-5: Pedigrees of families with <i>PPA2</i> mutations.....	119
Figure 4-6: Structural modeling of <i>PPA2</i> variants.....	122
Figure 4-7: Diagram showing the complexes of the electron transport chain located at the inner mitochondrial membrane.....	126
Figure 4-8: The formation of Inorganic Phosphate from Inorganic Pyrophosphate by PPA2.....	127
Figure 5-1: Pedigree of Family CHD1.....	135
Figure 5-2: IGV readout showing the variant in <i>DNAAF1</i> , c.571C>T, which is homozygous in both affected individuals in family CHD1.....	139
Figure 5-3: Conservation of variant p.L191F in <i>DNAAF1</i> orthologues.....	140
Figure 5-4: Position of c.571C>T in the <i>DNAAF1</i> gene and position of the missense substitution, L191F, in the <i>DNAAF1</i> protein.....	140
Figure 5-5: Sanger sequencing electropherograms to show the c.571C>T variant in <i>DNAAF1</i> in family CHD1.	141
Figure 5-6: Pedigrees of families with <i>DNAAF1</i> mutations. Mapping of an identified deletion in <i>DNAAF1</i> using WGS and Sanger sequencing.	145
Figure 5-7: A representation of the N-terminal TAP <i>DNAAF1</i> construct.....	146
Figure 5-8: A map of the pCS2+ <i>DNAAF1</i> construct.....	147
Figure 5-9: Agarose gel electrophoresis to show digestion of wildtype (wt) and mutant L191F (mut) PCS2+ constructs with EcoR1 enzyme.	148
Figure 5-10: Phenotypic rescue of <i>dnaaf1</i> ^{+/-} heterozygote-heterozygote intercrossed zebrafish embryos on injection of human wildtype, but not mutant L191F <i>DNAAF1</i> constructs.....	149
Figure 5-11: Co-immunoprecipitation assays to investigate interacting partners of <i>DNAAF1</i>	154
Figure 5-12: Co-localization of endogenous <i>DNAAF1</i> (green) and ciliary markers, RPGRIP1L, CEP290, Centrin-3 and γ -tubulin (red).	156
Figure 5-13: Immunofluorescence of overexpressed wildtype or mutant F191 FLAG- <i>DNAAF1</i> in RPE1 cells.....	157
Figure 5-14: Immunofluorescence of overexpressed wildtype or mutant F191 FLAG- <i>DNAAF1</i> in RPE1 cells and co-localisation with endogenous and overexpressed IFT88.	159
Figure 5-15: Immunofluorescence to show co-localisation of endogenous <i>DNAAF1</i> and Pontin/Reptin in RPE1 cells.	160

Figure 5-16: qPCR and western blot to demonstrate knockdown of RUVBL1/2 using siRNA.	161
Figure 5-17: Results of siRNA knockdown of Pontin (RUVBL1) in RPE1 cells and affect on co-localisation of DNAAF1 to IFT88 and IFT88 to γ -tubulin.	162
Figure 5-18: Light-sheet microscopy images showing immunofluorescence of RUVBL1 and ARL13B in staged mouse embryos.	164
Figure 5-19: Light-sheet microscopy images showing immunofluorescence of RUVBL1 and DNAAF1 in staged mouse embryos.	165
Figure 5-20: Pontin (<i>ruvbl1</i>) whole mount in situ hybridization in zebrafish embryos.	167
Figure 5-21: Transmission Electron Micrograph (TEM) of a human respiratory motile cilium.	170
Figure 5-22: Schematic diagram to show the hypothesized model of an RT2P-like complex involved in dynein arm assembly.	175
Figure 6-1: Pedigree of family CHD9.	178
Figure 6-2: Sanger sequencing electropherograms to demonstrate a mutation in <i>KMT2D</i> , c.10658 G>T, in family CHD9.	182
Figure 6-3: Sanger sequencing electropherograms showing the presence of <i>KMT2D</i> c.10658G>T (p.G3553V) in family CHD9b.	184
Figure 6-4: High-sensitivity bioanalyser traces from sample JT619.	187
Figure 6-5: Demonstration of absence of c.10658G>T in patient JT619 from family CHD11.	188
Figure 6-6: <i>KMT2D</i> protein showing domains and position of the G3553V variant.	195

List of Tables

Table 1-1: Classification of CHD.....	4
Table 2-1: List of primary antibodies.....	28
Table 2-2: Parameters used for hard filtering of variants.....	40
Table 2-3: Antibiotics used in microbiology methods, concentration used and supplier.....	46
Table 3-1: The ten largest regions of homozygosity identified in family CHD14.....	66
Table 3-2: Number of variants remaining following each filtering step in family CHD14.....	66
Table 3-3: Homozygous variants remaining following filtering, assuming a recessive disease model, in family CHD14.....	67
Table 3-4: Clinical features of families recruited to the project with CHD phenotypes, but without mutations identified in known disease genes....	70
Table 3-5: Candidate variants remaining following filtering of WES data in CHD families.....	80
Table 3-6: A summary of CNV calls from patient JT381 which were called by all three CNV analysis programs; CNV-seq, CNVnator and ReadDepth.....	87
Table 3-7: A summary of CNV calls from patient JT443 which were called by all three CNV analysis programs; CNV-seq, CNVnator and ReadDepth.....	88
Table 3-8: Candidate genes from CHD patients and their expression in fetal heart sample SRR643779.....	89
Table 3-9: Clinical features of patients from the DDD project coded as 'laterality disorder'.....	106
Table 3-10: Mutations in known disease-causing genes identified in patients from the DDD project.....	107
Table 4-1: Number of variants remaining at each stage of variant filtration in the analysis of family CHD2.....	113
Table 4-2: Variants remaining following filtering WES data using the data analysis pipeline in family CHD2.....	114
Table 4-3: Output of SNP Viewer program showing large regions of homozygosity shared between affected individuals and not with unaffected individuals in family CHD2.....	116
Table 4-4: Mutations in <i>PPA2</i> identified in families 2b-d alongside <i>in silico</i> pathogenicity prediction scores and MAF observed in the GnomAD database.....	121
Table 5-1: Large regions of homozygosity, identified on SNP array, shared between the affected siblings in family CHD1.....	136

Table 5-2: Number of variants remaining following each step in WES data analysis for family CHD1.	137
Table 5-3: Variants remaining following filtering in family CHD1.	138
Table 5-4: Heterotaxy samples screened for mutations in <i>DNAAF1</i> by Sanger sequencing.....	143
Table 5-5: Proteins associated with peptides pulled-down in a TAP experiment, identified by mass spectrometry, using <i>DNAAF1</i> wildtype or mutant (L191F) constructs as bait.....	151
Table 6-1: Number of variants remaining following each stage of filtering in family CHD9, assuming a new dominant disease model.....	180
Table 6-2: Number of variants remaining following each stage of filtering in family CHD9, assuming a recessive disease model.....	180
Table 6-3: Variants remaining following filtering in Family CHD9	181
Table 6-4: Number of variants remaining following filtering assuming a dominant inheritance pattern in family CHD10.....	185
Table 6-5: Showing the variant in <i>PTCH1</i> identified in family CHD10.....	185
Table 6-6: Summary of clinical features of patients with athelia syndromes from this study and the literature.....	189

Abbreviations

AAV	Adeno-Associated Virus
aCGH	array Comparative Genome Hybridization
ADP	Adenosine Diphosphate
AKAP	Protein Kinase A Anchoring Protein
ALL	Acute Lymphoblastic Leukaemia
ALS	Amyotrophic Lateral Sclerosis
AML	Acute Myelogenous Leukaemia
AR	Autosomal Recessive
ARPKD	Autosomal Recessive Polycystic Kidney Disease
ASD	Atrial Septal Defect
ATP	Adenosine Triphosphate
AV	Atrioventricular
AVC	Atrioventricular Canal
AVSD	Atrio-Ventricular Septal Defects
BAV	Bicuspid Aortic Valve
BCIP	5-bromo-4-chloro-3'-indolyphosphate
BMP	Bone Morphogenetic Proteins
BSA	Bovine Serum Albumin
BT	Blalock-Taussig
CC	Coiled-Coil
CDH	Congenital Diaphragmatic Hernia
cDNA	complementary DNA
CES	Combinational Enhancer Solution
CF	Cystic Fibrosis
CHD	Congenital Heart Disease
CK	Creatinine Kinase
CML	Chronic Myelogenous Leukaemia
CMV	Cytomegalovirus
CNS	Central Nervous System
CNV	Copy Number Variant
coIP	co-Immunoprecipitation
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CS	Citrate Synthase
CSF	Cerebrospinal Fluid
CT	Computerised Tomography
CTCF	Corrected Total Cell Fluorescence
CVS	Chorionic Villus Sampling
DCM	Dilated Cardiomyopathy
DD	Developmental Delay
DDD	Deciphering Developmental Delay

DMD	Duchenne Muscular Dystrophy
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DORV	Double Outlet Right Ventricle
DTT	Dithiothreitol
EBS	Epidermolysis Bullosa Simplex
EDTA	Ethylenediaminetetraacetic Acid
EMT	Epithelial-Mesenchymal Transformation
ETC	Electron Transport Chain
FCCP	Carbonyl-cyanide-4-(trifluoromethoxy)-phenylhydrazone
FCS	Fetal Calf Serum
FFPE	Formalin-Fixed Paraffin-Embedded
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
GATK	Genome Analysis Toolkit
GDP	Guanosine Diphosphate
GEF	Guanine Nucleotide Exchange Factor
GPI	Glycosylphosphatidylinositol
GPI	Glucose-6-Phosphate Isomerase
GU	Genito-Urinary
GWAS	Genome Wide Association Study
HDR	Homology-Directed Repair
HLHS	Hyperplastic Left Heart Syndrome
HPMRS	Hyperphosphatasia and Mental Retardation Syndrome
HPO	Human Phenotype Ontology
HSP	Heat Shock Protein
HUS	Hemolytic-Uremic Syndrome
ICD	Implantable Cardioverter Defibrillator
IDA	Inner Dynein Arm
IFT	Intraflagellar Transport
IP	Immunoprecipitation
iPSC	induced pluripotent stem cells
IUGR	Intrauterine Growth Retardation
IVC	Inferior Vena Cava
IVF	In-vitro Fertilisation
KS	Kabuki Syndrome
KV	Kupffer's Vesicle
LB	Luria Bertani
LHON	Leber Hereditary Optic Neuropathy
lncRNA	Long Non-coding Ribonucleic Acid
LRR	Leucine-Rich Repeat
LVOTO	Left Ventricular Outflow Tract Obstruction
MAF	Minor Allele Frequency

miRNA	Micro Ribonucleic Acid
MNP	Multi-Nucleotide Polymorphism
MO	Morpholino
MRI	Magnetic Resonance Imaging
NBT	Nitro-Blue Tetrazolium
NGS	Next Generation Sequencing
NHEJ	Non-Homologous End Joining
NIPT	Non Invasive Prenatal Testing
NO	Nitric Oxide
OCR	Oxygen Consumption Rate
ODA	Outer Dynein Arm
OXPPOS	Oxidative Phosphorylation
PAM	Protospacer-Associated Motif
PBS	Phosphate Buffered Saline
PCD	Primary Ciliary Dyskinesia
PCR	Polymerase Chain Reaction
PDA	Patent Ductus Arteriosus
PFA	Para-Formaldehyde
PFO	Patent Foramen Ovale
PGD	Pre-implantation Genetic Diagnosis
PH	Pulmonary Hypertension
Pi	Inorganic Phosphate
PPi	Inorganic Pyrophosphate
PRO	Proline-Rich
PS	Pulmonary Stenosis
PVL	Periventricular Leukomalacia
RAI	Right Atrial Isomerism
RNA	Ribonucleic Acid
RRC	Reserve Respiratory Capacity
RT	Room Temperature
SCID	Severe Combined Immunodeficiency Disorder
SDM	Site-Directed Mutagenesis
SDS	Sodium Dodecyl Sulfate
siRNA	Small interfering Ribonucleic Acid
SIT	Situs Inversus Totalis
snoRNA	Small Nucleolar Ribonucleic Acid
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
SOC	Super Optimal broth with Catabolite repression
SSC	Saline Sodium Citrate
SUDIC	Sudden Unexpected Death in Infants and Children
SVT	Supraventricular Tachycardia
TA	Truncus Arteriosus

TAAD	Thoracic Aortic Aneurysm and Dissection
TALeNs	Transcription Activator-Like Effectors
TAP	Tandem Affinity Purification
TBS	Tris Buffered Saline
TEM	Transmission Electron Microscopy
TGA	Transposition of the Great Arteries
TGF	Transforming Growth Factor
TOF	Tetralogy of Fallot
TPR	Tetratricopeptide Repeat
VEGF	Vascular Endothelial Growth Factor
VSD	Ventricular Septal Defect
WBS	William-Beuren Syndrome
WCE	Whole Cell Extract
WES	Whole Exome Sequencing
WGA	Whole Genome Amplification
WGS	Whole Genome Sequencing
YFP	Yellow Fluorescent Protein
ZNF	Zinc Finger Nuclease

1 Introduction

1.1 General Introduction

Congenital Heart Disease (CHD) is the most common congenital defect and affects 0.5-0.8% of all live births (Reller et al., 2008). CHD is a complex group of conditions that encompass disorders of structural cardiac development, including septal, valvular and outflow tract lesions. These disorders demonstrate a wide spectrum of severity, from the complex forms of CHD, to milder lesions, such as bicuspid aortic valve (BAV), that may not cause symptoms apparent from birth.

The complex forms of CHD carry a high burden of morbidity and mortality (Moller and Anderson, 2013), with approximately 25% of patients requiring intervention before the age of 1 year (Ahrens-Nicklas et al., 2016). Advances in the surgical and medical management of CHD have led to improved survival for patients, who are now more likely to survive into adulthood and reproduce (Pierpont et al., 2007), meaning that understanding the underlying genetic aetiologies of CHD are of increasing importance.

The majority of instances of CHD are sporadic occurrences and the cause of these events remains largely unknown (Zaidi and Brueckner, 2017). It has been traditionally suspected that many of these sporadic cases are heterogenic or the result of a genetic-environmental interaction. There is now strong evidence to support the genetic basis of CHD, including the increased risk of familial occurrence. The prevalence of CHD has been shown to be higher in first-degree relatives of an affected individual than second-degree relatives, and higher between monozygotic twins than siblings (Wang et al., 2013). The recurrence risk for CHD is estimated to be between 4-6% for unaffected parents (Fesslova et al., 2011) and as high as 10% when two previous siblings are affected (Chaix et al., 2016). The recurrence risk from affected parents to offspring is 8-10%, with the highest risks associated with affected mothers, rather than fathers, and with left-sided lesions rather than other types of CHD (Chaix et al., 2016, van der Bom et al., 2011). The heritability of some types of CHD has been shown to be as high as 70-90%, indicating a strong genetic contribution (Cripe et al., 2004, McBride et al., 2005, Hinton et al., 2007). However, it is not yet clear whether this genetic contribution is attributable to single loci, to the additive interactions of multiple loci, or to various combinations of these (Zaidi and Brueckner, 2017). Other non-genetic aetiologies of CHD include maternal drug exposures, for example smoking and alcohol use (Jenkins

et al., 2007), maternal diabetes (Wren et al., 2003) and congenital infections, such as the rubella virus (Zhu et al., 2009).

Advancements in cytogenetic analysis, such as the availability of Array Comparative Genome Hybridization (aCGH) and molecular techniques such as Whole Exome Sequencing (WES) have led to a number of exciting recent advances in the genetic aetiology of CHD and have improved our understanding of the complex developmental networks involved in human cardiac development (LaHaye et al., 2016, Al Turki et al., 2014). The number of genes known to be associated with CHD has grown rapidly in recent years, but the genetic aetiology of a large proportion of CHD remains unknown (Andersen et al., 2013). In total, the known genetic causes of CHD, including syndromic causes, account for less than one third of all cases (Pediatric Cardiac Genomics Consortium et al., 2013, Zaidi and Brueckner, 2017).

Improved understanding of the genetic aetiology of CHD is important to allow insight into the biological mechanisms, interactions and pathways involved in these disorders, and to allow disease stratification and categorization of risk. This is crucial in defining prognosis for affected patients and in counseling family members. Furthermore, this may lead to the development of more accurate prognostic testing, new management strategies and treatments (Pierpont et al., 2007). This project aims to identify new genetic causes of CHD through the use of Whole Exome Sequencing in a family-based approach.

1.2 Formation of the Foetal Heart

The formation of the foetal heart commences at 15-16 days gestation in humans. Precursors of the majority of cardiac cells are the splanchnic mesodermal cells, the exception being the cushions of the outflow tract, which are formed, in part, from ectodermal-derivatives of the cardiac neural crest (Sylva et al., 2014). During gastrulation, these cells migrate to the midline myocardial plate and form the cardiac crescent, also known as the first heart field. The cardiac crescent then folds and eventually fuses in the midline to create the heart tube. The heart tube has two venous inlets and one arterial outlet.

Addition of cells from the secondary heart field causes the heart tube to elongate at the either pole. The heart tube then undergoes ventral bending and right heart looping, at gestational day 28. Formation of the cardiac loop is closely related to the establishment of laterality at the embryonic node (see section 1.5.2.3) (Lin et al., 2012). Differential

gene expression in the looped heart causes dilatation of the atrial and ventricular chambers, and the heart tube becomes segmented into the primitive atria, atrioventricular canal (AVC), primitive ventricles, and outflow tract (Lin et al., 2012). This dilatation requires multiple regulators and transcription factors expressed in overlapping gradients, including important cardiac transcription factors, such as *Nkx2.5* and *Gata4* (Boogerd et al., 2009). Cushions form in the inner vessel walls of the AVC and outflow tract, through a process of epithelial-mesenchymal cell transformation (EMT). EMT is regulated by Bone Morphogenetic Proteins (BMPs), TGF-beta and Notch signaling (Sylva et al., 2014).

Following this, during days 50-90 of gestation in humans, the atrial and ventricular chambers septate to form the four-chambered heart. The atrioventricular (AV) cushions fuse at the AVC and the outflow tract is septated by the septal and parietal outflow tract ridges to form the ventricular outlets (Sylva et al., 2014). The AV cushions, having undergone EMT, are remodeled to become the AV valves and the outflow tract cushions give rise to the semilunar valves (Lin et al., 2012). The valves undergo a final remodeling stage to create three distinct valvular layers (Sylva et al., 2014). Figure 1-1 shows a representation of the formation of the embryonic heart in both human and mouse.

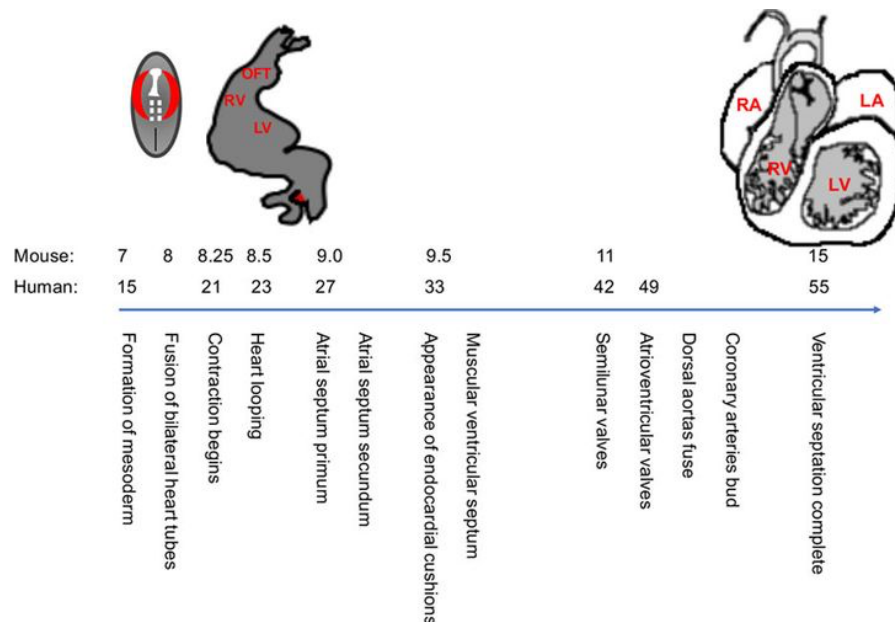


Figure 1-1: The process of cardiac embryology in human and mouse.

Image adapted from Zaidi et al., 2017. Scale shows gestational days 15-55 in humans, and 7-15 in mouse. OFT= outflow tract, RV= right ventricle, LV= left ventricle, RA= right atrium, LA= left atrium.

1.3 Classification of Congenital Heart Disease

Congenital Heart Disease can be defined as a defect in the normal structure of the heart that is present from birth. CHD includes septal defects such as Ventricular Septal Defects (VSD), Atrial Septal Defects (ASD) and Atrio-Ventricular Septal Defects (AVSD), describing an abnormal connection between the right and left sides of the heart. Valvular abnormalities include atresia, describing an absence of the valve orifice, and stenosis, describing narrowing, due to an abnormality in the structure or function of the valve. CHD also encompasses more complex defects such as Tetralogy of Fallot (TOF), a combination of VSD, pulmonary stenosis, right ventricular hypertrophy and an overriding aorta, and hypoplastic left heart syndrome, an abnormality in the formation of the left ventricle, often associated with valvular defects. Congenital cardiomyopathies are not strictly CHDs, but describe diseases of the heart muscle present from birth (van der Bom et al., 2011). Table 1-1 shows a classification of different types of CHD according to the major structural abnormality present.

Table 1-1: Classification of CHD.

Table adapted from (Miyague et al., 2003). Prevalence of different types of CHD is given per million live births, based on data from the CONCOR (Dutch Congenital Corvitia) Database (van der Bom et al., 2012).

Structure affected	Type of CHD	Prevalence per million live births (van der Bom et al., 2012)
Atria	Single atrium	5
Atrioventricular connection	Tricuspid atresia Ebstein anomaly Mitral Valve Lesions	79 61 50
Outflow tract	Transposition of the Great Arteries Double outlet right or left ventricle (DORV/ DOLV) Truncus Arteriosus (TA) Pulmonary Stenosis (PS) Pulmonary Atresia Aortic Stenosis Tetralogy of Fallot (TOF) Coarctation of the Aorta	315 157 107 728 132 401 421 409
Septum	Atrioventricular Septal Defect Atrial Septal Defect (ASD) Ventricular Septal Defect (VSD)	348 941 3570
Ventricle	Hypoplastic left ventricle Single ventricle	266 106
Cardiac symmetry	Dextrocardia Heterotaxy Right Isomerism Left Isomerism Ivemark's Syndrome	333 186

1.4 Management of CHD

CHDs can now be detected using prenatal echocardiography at 17-18 weeks gestation and many CHDs are well-tolerated during foetal life, due to the support of the maternal circulation. However, following birth, many will manifest clinically. Before the advent of cardiac surgery, CHD was one of the leading causes of death in neonates and infants. Since the introduction of surgery and intensive care in the 1950s, there has been a huge change in the life expectancy of these patients, with >90% now surviving to adulthood (Triedman and Newburger, 2016).

Paediatric cardiac surgery is now commonly performed at a number of specialist cardiac centers throughout the UK, providing a range of medical and surgical services, allowing for specialist commissioning (<https://www.engage.england.nhs.uk/>) and combined multidisciplinary expertise. Furthermore, many lesions can now be managed by cardiac catheter, for example ASD and pulmonary valve stenosis, without the need for open cardiac surgery (Abu-Tair et al., 2016, Singh et al., 2015, Moustafa et al., 2016).

The required surgery for CHD will depend on the nature of the lesion present. In the case of VSD, an abnormal connection between the two ventricles, the patient may require surgery to close the hole in the septum if the lesion is large and unlikely to close spontaneously. This involves the insertion of a synthetic patch to close the connection, either by cardiac catheter or open-heart surgery (Holzer et al., 2016). In transposition of the great arteries (TGA), where the aorta is abnormally attached to the right side of the heart and the pulmonary artery is on the left, an arterial switch repair is required shortly after birth that transposes the orientation of these vessels (Villafane et al., 2014). As well as corrective procedures, some patients require a cardiac shunt - the insertion of a short synthetic tube - to relocate the flow of blood in the cardiac circulation. This is performed either as a palliative procedure or prior to later planned palliative or corrective surgery. A Blalock-Taussig (BT) shunt is used in cyanotic heart defects, such as pulmonary atresia. In this procedure, one branch of the subclavian or carotid artery is separated and connected to the pulmonary artery to relieve the symptoms of cyanosis. A bidirectional cavopulmonary anastomosis (Glenn shunt) involves disconnecting the superior vena cava from the right atrium and connecting it to the right pulmonary artery. This allows deoxygenated blood to return directly to the lungs without going through the heart, reducing the workload on the heart. A Glenn

shunt is often one of a series of operations in the treatment of uni-ventricular disease (Mainwaring et al., 1994).

The increase in life expectancy for CHD patients has brought new challenges, with adult CHD patients being at risk of long-term complications, for example arrhythmia, heart failure, endocarditis and aneurysm (Postma et al., 2016). The main causes of death in this population are heart failure and sudden cardiac death. Current management guidelines focusing on adults with CHD recommend surveillance, advise on the risk of complications and offer advice for pregnancy (Warnes et al., 2008, Baumgartner et al., 2010).

1.5 The Known Genetic Aetiology of Congenital Heart Disease

The known causes of CHD are varied and include cytogenetic abnormalities, single gene disorders, epigenetic alteration or multifactorial causes. CHD can occur as an isolated feature, or part of a known or undiagnosed syndrome of congenital anomalies (Landis and Ware, 2016). Approximately 25% of children with CHD are thought to have a syndromic cause, due to the presence of other congenital anomalies or neurodevelopmental delay (Landis and Ware, 2016). However, this definition is not always clear-cut, and sometimes the gene responsible for a syndromic form of CHD can cause non-syndromic disease. For example, *JAG1* mutations can cause Alagille syndrome, but are also seen in non-syndromic TOF (Eldadah et al., 2001). This section describes some of the known genetic causes of congenital heart disease.

1.5.1 Copy Number Variation

Copy Number Variation is classed as a structural variation within the genome, specifically a duplication or deletion event (McCarroll and Altshuler, 2007). This type of variation can affect whole chromosomes (aneuploidy, for example Trisomy 21), down to the level of a single base-pair deletion or duplication. Different technologies are able to identify CNVs of differing size; a karyotype will identify deletions or duplications down to the resolution of approximately 5Mb (Geiersbach et al., 2014), an array CGH those down to approximately 50-100kb (although this is dependent upon the array platform and differs across the genome) (Coe et al., 2007) and whole genome Single Nucleotide Polymorphism (SNP) array those of sub-100kb (dependent upon the platform and markers used) (Peiffer et al., 2006).

Chromosomal aberrations and aneuploidies are known to be a common cause of syndromic CHD, with 22q11 deletion being a common cause, occurring in 1 in 4000 live births and accounting for around 15% of TOF cases (Agergaard et al., 2012). 22q11 deletion encompasses the *TBX1* gene, haploinsufficiency of which is reported to underlie the cardiac phenotype (Jerome and Papaioannou, 2001). *TBX1* has recently been shown to regulate monomethylation of H3K4me1 (Histone H3 monomethylated at lysine 4) through the recruitment of histone methyltransferases (Fulcoli et al., 2016) and H3K4 methylation is known to be important in cardiac development (see section 1.5.3).

Since the introduction of array CGH technology, new microdeletion and duplication syndromes associated with syndromic CHD have been identified, such as the 16p11.2 deletion syndrome (Ghebranious et al., 2007) and 8p23 deletion, which includes the *GATA4* gene (Pehlivan et al., 1999).

Rare copy number variants (CNVs) (<1% population frequency) have increasingly become a focus for investigation in many complex diseases, including CHD (Soemedi et al., 2012). Patients with non-syndromic CHD have been found to have a higher than expected number of submicroscopic rare CNVs (Erdogan et al., 2008). There are a number of recent studies of whole-genome CNV screening in CHD cohorts, many of which have identified an increase in rare CNVs in CHD patients when compared to controls (Silversides et al., 2012, Soemedi et al., 2012, Hitz et al., 2012, Fakhro et al., 2011, lascone et al., 2012). CNVs identified by this methodology provide a method for identifying potential candidate genes. However, there is little consistency between these studies, perhaps due to variations in study size and cardiac phenotype investigated (Andersen et al., 2013).

In a recent paper, Glessner et al., 2014, used high-density SNP arrays and WES to detect small *de novo* CNVs in a trio study of 538 cases of conotruncal defects, heterotaxy and left ventricular outflow tract defects. The results were compared to those of 1301 healthy controls. There was a significant increase in *de novo* CNVs in cases compared to controls. A number of the CNVs identified encompassed genes that interact with *NKX2-5* and *GATA4*, known to play important roles in cardiac development (Glessner et al., 2014).

Overall the incidence of pathogenic CNVs, identified by array CGH or SNP array, in isolated CHD is reported to be between 4-14%, although the genomic architecture of

these CNVs is often unknown; some appear to be single causative alleles, but others are additive at two or more different loci (Soemedi et al., 2012, Breckpot et al., 2011). This is lower than in syndromic CHD where the incidence may be as high as 15-20% (Breckpot et al., 2011, Tomita-Mitchell et al., 2012). CNVs identified in non-syndromic CHD more often have a moderate effect size and reduced penetrance, as compared to those causative of syndromic CHD (Breckpot et al., 2011).

1.5.2 Single Gene Variants

1.5.2.1 Transcription Factors

The importance of key transcription factors in cardiac development has been highlighted by the discovery of mutations in genes such as *NKX2-5* (Schott et al., 1998), *GATA4* (Tomita-Mitchell et al., 2007), *GATA5*, *GATA6* (Zheng et al., 2012) and *TBX5* (Reamon-Buettner and Borlak, 2004) in human CHD. These core transcription factors interact in a complex pathway, along with other key transcription factors in the developing heart. Some of these interactions are illustrated in Figure 1-2.

GATA4, 5 and 6 are zinc-finger transcription factors expressed in overlapping regions in the developing heart (Durocher and Nemer, 1998). Mouse models have shown that *Gata4* works together with *Smad4* to regulate EMT, and therefore plays a key role in cardiac septation and valve formation (Moskowitz et al., 2011). *Gata4* also interacts with other crucial transcription factors, including *Hand2*, *Nkx2.5* and *Tbx5* (Garg et al., 2003). Mice lacking endocardial *Gata4* have halted EMT at the AV cushions and develop hypoplastic ventricles, whereas mice with heterozygous *Gata4* mutations develop AVSD and DORV (Rivera-Feliciano et al., 2006). *GATA4* mutations are found to be the cause of non-syndromic septal, valvular and outflow tract defects in humans, including Pulmonary Stenosis (PS) and TOF, as well as dextrocardia (Moskowitz et al., 2011). Similarly, mutations in both *GATA5* and *GATA6* have been found to cause disorders of the aortic valve and outflow tract in humans (Kodo et al., 2009, Wei et al., 2013).

In mice, *Nkx2-5* is expressed in both the first and second heart field and is involved in the specification of cardiac precursors and in the development of the outflow tract (McCulley and Black, 2012). *Nkx2-5* is regulated by *Smad* proteins and interacts with both *Gata4* and *Tbx5* to regulate septation in the atria and ventricles (Lien et al., 1999, Hiroi et al., 2001). *Nkx2.5* null mice do not develop a primary heart tube, but do not undergo the processes of cardiac looping or septation. Mutations in *NKX2-5* in humans

are responsible for a broad range of CHDs including DORV, TOF, VSD and ASD (Schott et al., 1998). Furthermore, *NKX2-5* mutations are also associated with cardiac conduction defects (Benson et al., 1999).

The T-box transcription factors play important roles in the developing heart and regulate cardiomyocyte identity. *TBX5*, a member of the T-box transcription factors, is expressed in a gradient from the venous pole to the right ventricle and activates chamber-specific gene expression, as well as playing a role in cardiac looping and septation (Sylva et al., 2014). *Tbx5* null mice suffer early lethality and do not undergo cardiac looping or form cardiac chambers. Heterozygous mutations of *Tbx5* in mice cause ASD and VSD (Bruneau et al., 2001). Mutations in the *TBX5* gene in humans cause Holt-Oram syndrome, a condition associated with a spectrum of congenital heart defects and limb abnormalities (Gall et al., 1966). Somatic *TBX5* mutations have also been associated with CHD in patients without Holt-Oram Syndrome (Reamon-Buettner and Borlak, 2004). Furthermore, gain-of-function mutations in *TBX5* have recently been reported in TOF patients (Baban et al., 2014). Mutations in other T-box genes have also been associated with CHD, for example *TBX20* mutations are associated with septal defects and DCM and *TBX1* with DiGeorge syndrome (Yagi et al., 2003). *Tbx5* and *NKX2-5* synergistically bind the promoter of the *Nppa* gene (cardiac specific natriuretic peptide precursor A) and cause activation during cardiac development (Hiroi et al., 2001).

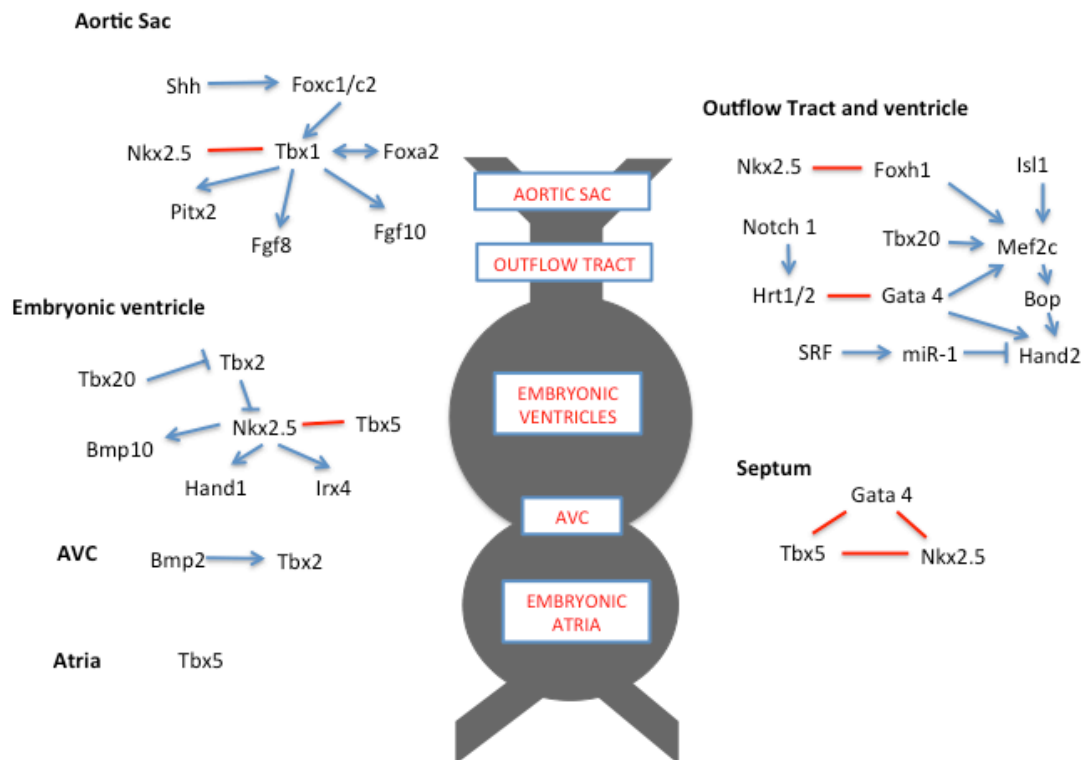


Figure 1-2: Transcription factors involved in region-specific cardiac development.

Schematic that is a representation of the major transcription factors involved in cardiac development and their interactions. Arrowheads indicate positive regulatory effects, whereas bars represent negative effects. Red lines indicate known protein-protein interactions. Adapted from (Srivastava, 2006) and (Mitchell et al., 2007).

1.5.2.2 Signaling Pathways

The NOTCH signaling pathway is an excellent example of the importance of signaling pathways in human cardiac development. This pathway regulates cell-fate decisions in heart morphogenesis, and is involved in myocardial development and left-right partitioning. The Notch pathway also regulates EMT in development of the cardiac valves (Niessen and Karsan, 2008). Alagille syndrome is caused by mutations in genes in this pathway (*JAG1* and *NOTCH2*) and is associated with pulmonary stenosis. Mutations in *Notch1* in mice cause hypoplastic endocardial cushions, reported to be due to impaired EMT (Timmerman et al., 2004). Mutations in genes encoding proteins involved in the Notch signaling pathway have been associated with human aortic valve disease and BAV with thoracic aortic aneurysm (McKellar et al., 2007), as well as left ventricular outflow tract obstruction (McBride et al., 2008). In a recent study using custom NGS panels, Blue et al., 2014, identified five variants in four genes in 16 families with CHD, including *NOTCH1*, which were associated with disease (Blue et al., 2014). Variants in *NOTCH1* have also been associated with non-syndromic LVOTO in the study by Jia et al., 2015, however the pathogenicity of these variants was not

certain, as the variants did not consistently segregate with disease in the identified families and the causality of these variants was not further investigated by functional studies (Jia et al., 2015). In a further study, mutations in genes of the NOTCH pathway (*NOTCH1*, *MAML1* and *JAG1*) were enriched in 51 families affected by CHD affecting the left ventricular outflow tract (McBride et al., 2008). *NOTCH1* haploinsufficiency was recently identified to cause Adams-Oliver Syndrome, characterized by limb defects, cutis aplasia and, in around half of cases, CHD (Southgate et al., 2015).

1.5.2.3 Laterality in the Early Embryo

Left-right patterning in the developing embryo is principally determined by nodal cilia in the embryonic node of the primitive streak (see section 1.6), although some earlier influences on left-right patterning may occur by processes such as planar cell polarity (Vandenberg and Levin, 2013). The clockwise rotational movement of the nodal cilia produces a leftward fluid flow. This nodal flow is reported to be either detected by mechanosensory (primary) cilia surrounding the node, or to initiate the movement of small vesicles containing morphogens such as Sonic Hedgehog or retinoic acid which move towards the left side of the node (Shinohara et al., 2012). This morphogenic signal may cause an influx of intracellular calcium in the cells to the left of the node, which activates a left-sided signalling pathway, involving genes such as *NODAL*, *LEFTY2*, *ACVR2B*, *CFC1*, *CITED2*, *PITX2*, *FOXH1* and *ZIC3* (Ramsdell, 2005, Sutherland and Ware, 2009). Creation of an artificial rightward flow at the node in mouse embryos causes inverted cardiac looping and contralateral expression of *Pitx2* and *Nodal* (Nonaka et al., 2002). Human mutations in genes involved in the establishment of embryonic laterality, such as *ZIC3*, *ACVR2B*, *CFC1* and *LEFTYA* cause a broad range of CHD, often associated with laterality defects (Sutherland and Ware, 2009). Mutations in these genes can also cause CHD without a laterality defect, for example a cohort of patients with TGA, but no other laterality defect, were found to have mutations in *CFC1* (Goldmuntz et al., 2002).

1.5.2.4 Variants in Structural Genes

Mutations in genes encoding components of the cardiac sarcomere have been linked to CHD. Mutations in myosin heavy chain 6 (*MYH6*) have been linked to ASD (Ching et al., 2005). Furthermore, mutations in *MYH7* have been identified to cause Ebstein's anomaly (Postma et al., 2011, Bettinelli et al., 2013). Mice with homozygous mutations in *Myh11* showed delayed closure of the arterial duct, and mutations in the human orthologue of this gene can cause Patent Ductus Arteriosus (PDA) associated with thoracic aortic aneurysm (Zhu et al., 2006). In two large families with autosomal dominant ASD, linkage analysis and candidate gene sequencing revealed the

causative mutation to lie in alpha-cardiac actin (*ACTC1*), a major component of the sarcomeric thin filaments, essential for heart muscle contraction (Matsson et al., 2008).

1.5.3 Epigenetics

Epigenetic modulation of the cardiovascular system plays a key role in cardiac embryology and defective regulation has recently been identified as an important cause of CHD. Knockdown experiments in animal models of chromatin factors, such as Brg1 and Chd7, and histone modifiers, such as Jumonji and Ptdsr, have demonstrated a variety of CHDs (Chang and Bruneau, 2012, Zhang and Liu, 2015).

Brg1 is a chromatin remodeling protein that has been identified to play an important role in cardiovascular development. Brg1 plays a dose-dependent role in the activation of cardiac specific target genes to control tissue growth and differentiation. Cardiac transcription factors Nkx2.5, Tbx4 and Gata4 have all been shown to physically interact with Brg1, via Baf60c. Brg1 has also been shown to genetically interact with Tbx5 (Bevilacqua et al., 2013). These interactions are known to be complex and there are sets of genes that appear to rely on the balance between Brg1 and Nkx2.5 or Tbx5 for their regulation (Hang et al., 2010). Deletion of *Brg1* in the early embryo causes lethality with cardiac underdevelopment, and deletion at a later stage produces mice with thin ventricular walls and an absent ventricular septum (Hang et al., 2010). Heterozygous *Brg1* mutations in mice cause multiple heart defects including VSD, PFO and dilated cardiomyopathy (Lickert et al., 2004). Morpholino oligonucleotide knockdown of Brg1 in zebrafish embryos causes abnormal cardiac looping with a hypoplastic myocardium (Chang and Bruneau, 2012). *BRG1* is re-activated in some human patients with hypertrophic cardiomyopathy and its level seems to correlate with disease severity (Hang et al., 2010).

H3K4 methylation is associated with transcriptionally active genes in the developing embryo. H3K4 methylation is maintained by methyltransferases, for example the MLL/KMT family of proteins. Mutations in a number of these methyltransferases cause human disease associated with CHD (Zhang and Liu, 2015). Mutations in *CHD7*, encoding a chromatin remodeling helicase protein, cause CHARGE syndrome (Coloboma, Heart defects, Atresia choanae, Retarded growth and development, Genital hypoplasia and Ear abnormalities/deafness) (Vissers et al., 2004). The associated cardiac defects are usually outflow tract malformations. Mice with heterozygous *Chd7* mutations have a similar combination of defects to human patients with CHARGE syndrome. This is replicated by *chd7* knockdown in frog embryos, which

also demonstrate similar features (Randall et al., 2009, Bajpai et al., 2010). *Chd7* was shown to interact with Brg1 and Pba1 to control neural crest activation by immunoprecipitation experiments in human neural crest cells and this may be the reason for the propensity to defects in outflow tract development in this condition, as the neural crest cells contribute to the formation of the outflow tract (Bajpai et al., 2010). Mutations in an additional H3K4 methyltransferase, *KMT2D*, cause Kabuki syndrome, which is also associated with a number of CHDs such as septal defects (Ng et al., 2010). Furthermore, heart-specific knockout of *Kmt2d* results in outflow tract septation defects in mice (Ang et al., 2016).

The *Jumonji (Jarid2)* gene codes for a histone demethylase. Homozygous mutations in *Jarid2* in mice cause lethal cardiac defects, including VSD, non-compaction cardiomyopathy, DORV and aortic dilatation. There is a reduction in atrial and ventricular chamber-specific gene regulation in these mice (Lee et al., 2000). *Jarid2* is known to inhibit expression of cardiac transcription factors such as *Nkx2.5* and *Gata4*, affecting the regulation of early cardiac differentiation (Kim et al., 2004).

The importance of histone modifying genes in human CHD was highlighted in an influential paper by Zaidi et al., in 2013. This group studied 362 parent-offspring trios consisting of a child affected with CHD and their unaffected parents. Exome sequencing techniques were used to identify *de novo* variants in the affected children. Control subjects were 264 trios consisting of unaffected siblings of autism cases and their parents. On initial review there was not a significant difference in the number of *de novo* variants identified between cases and controls. However, RNA sequencing was used to identify genes expressed at embryonic day E14.5 in the mouse heart and the group identified an increase in deleterious *de novo* mutations in these genes in cases compared to controls (Odds Ratio = 7.5). When the genes mutated in the case group were examined, eight were found to be involved in H3K4 methylation production, removal or reading. *SMAD2* was found to be mutated in two cases, both with dextrocardia. Interestingly, *SMAD2* is known to be involved in demethylation of H3K27 at the left-right organizer of the embryonic node. The group concluded that genes within the H3K4 methylation and H3K27 methylation pathway must be important in cardiac development (Zaidi et al., 2013). This would be consistent with prior evidence that genes in the H3K4me-H3K27me pathway are important in regulating genes involved in embryonic development, including those involved in cardiac development (Paige et al., 2012).

1.5.4 MicroRNAs

The importance of the non-coding parts of the genome is becoming apparent in many disease processes, including cardiovascular diseases. MicroRNAs (miRNAs) are usually around 22 nucleotides in length and bind to complementary sequences on target mRNA, often resulting in gene silencing or repression. They act either by reducing mRNA stability or inhibiting translation (Li et al., 2013). It is thought that there are perhaps as many as 1000 miRNA genes in the human genome. These are highly conserved across species and have been implicated in numerous disease processes including cardiac disease and cancers (Salmena et al., 2011). There is evidence that other non-coding components of the genome, such as long non-coding RNAs (lncRNAs) and small nucleolar RNAs (snoRNAs) are also crucial in organismal development and may also be implicated in CHD (O'Brien et al., 2012).

Research into miRNAs in cardiovascular pathologies such as heart failure and cardiomyopathies has recently rapidly expanded. This is, in part, due to the identification of the importance of miRNAs in cardiac development in animal models. This importance has been demonstrated by disruption of the miRNA processing enzyme Dicer, an RNase III endonuclease, required in the maturation of most miRNAs. In both mice and zebrafish, cardiac specific disruption of Dicer causes underdevelopment of the ventricular myocardium, dilated cardiomyopathy and cardiac failure (Ono et al., 2011).

miR-1 and miR-133 are involved in cardiomyocyte cell differentiation and cardiac muscle cell differentiation respectively. These families of miRNAs are generated as pairs and their expression is specific to cardiac and skeletal muscle (Chen and Wang, 2012). Overexpression of miR-1 in mice leads to thinning of the ventricular wall and early lethality (Zhao et al., 2005). Conversely, miR-1 knockout mice were found to have hyperplastic ventricles (O'Brien et al., 2012). One of the targets of miR-1 in the heart is *Hand2*, a transcription factor that plays a key role in cardiac myocyte development. *Hand2* mutant mice have similar cardiac defects to miR-1 mutants (Zhao et al., 2005).

Compound deletion of miR-133a-1 and miR-133a-2 causes early embryonic lethality, VSDs, dilated cardiomyopathy and cardiac failure in mice (Liu et al., 2008). Overexpression of miR-133a causes thinning of the ventricular wall and septal defects resulting in lethality (Liu et al., 2008).

miR-138 is expressed in very specific domains in the zebrafish heart and is reported to be involved in cardiac patterning. Morpholino antisense oligonucleotide knockdown of miR-138 expression in zebrafish caused a lack of cardiac maturation (Morton et al., 2008), although the effect of off-target effects in these experiments cannot be discounted. There are several miRNAs known to be linked to septal defects, including miRNA126 (Lin et al., 2012) and miRNA-1-1 and miRNA-181c (Li et al., 2013).

O'Brien *et al.*, 2012, studied expression of miRNA and snoRNA in cardiac tissue of 16 non-syndromic patients with Tetralogy of Fallot, 3 fetal samples and 8 control samples. They carried out microarrays for 847 human miRNAs and 922 snoRNAs. They found 61 miRNAs and 135 snoRNAs to have significantly different expression between the cases and control samples. Of those miRNAs with differential expression, there was an increased incidence of those known to regulate genes with direct links to cardiac development. The difference in expression of miRNAs and snoRNAs between the control and case samples suggests the importance of these non-coding RNAs in cardiac development and CHD due to their role in the regulation of key cardiac genes (O'Brien et al., 2012).

Long non-coding RNAs are >200 nucleotides in length and are transcribed throughout the genome. In general they have low evolutionary conservation. lncRNAs can interact with proteins, such as transcription factors and chromatin modifying complexes, or other RNAs such as miRNAs, to affect gene regulation and expression (Scheuermann and Boyer, 2013). It is known that lncRNAs play a role in important cell processes such as imprinting and cell cycle regulation. Recently it has been identified that lncRNAs play a role in cardiac development. One example of this is Braveheart, a lncRNA that is expressed in the mouse heart. Using multiple embryonic stem cells, Klattenhoff *et al.*, 2013, demonstrated that removal of Braveheart impaired formation of cardiomyocytes and was important in embryonic cardiac cell fate (Klattenhoff et al., 2013). Braveheart is required for the activation of MesP1, an upstream regulator of core cardiac transcription factors such as Gata4, Gata6, Tbx2 and Nkx2.5. lncRNAs have also been implicated in adult-onset cardiac diseases such as myocardial infarction and dilated cardiomyopathy. These non-coding RNAs are likely to be associated with congenital heart diseases due to their role in cardiac development and interaction with cardiac transcription factors, although this has not, as of yet, been demonstrated (Scheuermann and Boyer, 2013).

1.6 Cilia and CHD

1.6.1 Cilia

Cilia are small microtubule-based organelles, formed from a mother centriole, that play a sensory role in response to chemical, light, morphogenic and mechanical stimuli. The majority of mammalian cell types express an immotile primary cilium. The structure of this organelle is complex and it contains an axoneme structure with nine microtubule doublets in a 9+0 arrangement (see Fig. 1-3). At the base of the axoneme is the basal body and between the basal body and the cilium is a specialised area called the 'ciliary gate', a barrier between the cilium and the cytoplasm. This ciliary gate consists of two regions; the distal appendages of the centriole and the transition zone. Cargo molecules are transported along the axoneme by intraflagellar transport (IFT) molecules, which consist of subcomplexes A and B (Malicki and Johnson, 2017). The importance of the primary cilium has been highlighted by a group of diseases called ciliopathies, which cause neurodevelopmental, skeletal, renal and retinal disease, and are linked to CHD (Hildebrandt et al., 2011).

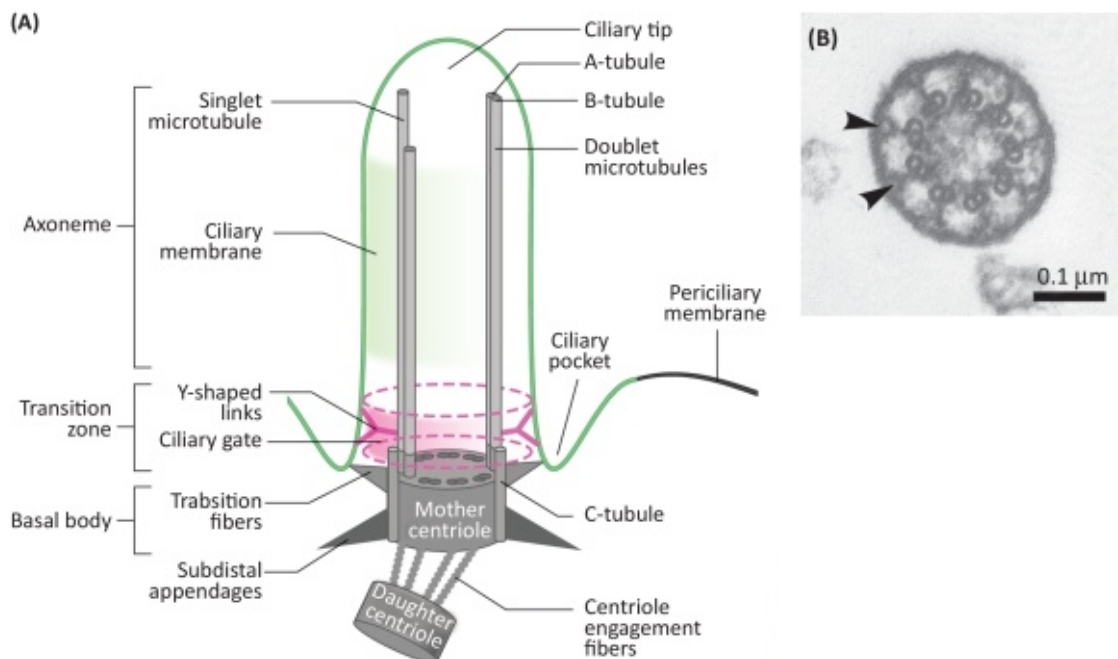


Figure 1-3: The major known structures of the primary cilium.

(A) Diagram of the primary cilium showing the major known structures, (B) Electron-microscopy to show the nine microtubular doublets in a 9+0 arrangement in the primary cilium. Arrows indicate the Y-shaped links of the transition zone. Figure adapted from Malicki et al., 2017.

Specialized motile cilia are present on a smaller number of cells types that, in humans, include those of the respiratory tract, Fallopian tubes and inner ear. Motile cilia differ from primary cilia in that they have an outer ring of nine microtubular doublets around a central microtubule pair (9+2 arrangement). In motile cilia, inner and outer dynein arms (IDA and ODA) connect the microtubules and act as molecular motors to provide the force required for ciliary motility (Nicastrò et al., 2006). A further type of specialized motile cilia are present at the embryonic node, with a 9+0 microtubule structure. These nodal cilia are involved in the creation of the leftward fluid flow at the embryonic node, a crucial function in the development of left-right asymmetry in the early embryo (see section 1.5.2.3).

1.6.2 Defects in Primary Cilia are Associated with CHD

Mutations in genes affecting the structure and function of primary cilia have been identified as an important cause of CHD. A number of ciliopathies have CHD as associated clinical features, for example those disorders associated with mutations in *NPHP4*, *KIF7* and *MKS1* (Braun and Hildebrandt, 2017). In E9.5-12.5 *Kif3a*-mutant mouse hearts, which totally lack cilia, there was thinning of the myocardium and abnormal development of the endocardial cushions, whereas mouse mutants for the left-right dynein gene (*Ird1*^{-/-}), which have static cilia, had relatively normal hearts (Slough et al., 2008). This suggests a sensory role for primary cilia in cardiac development. Furthermore, in a forward genetic screen in mice, where severe CHD was detected by fetal ultrasound scan, 61 genes were identified related to recessively-inherited CHD. Notably, of these, 34 were cilia-related genes. These cilia-related genes contributed not only to laterality-related CHD, but also non-laterality related disease. This demonstrates the important role of primary cilia in cardiac development, separate from the role in breaking laterality in the developing embryo (Li et al., 2015).

1.6.3 Defects in Motile Cilia are Associated with CHD

Defects in genes associated with motile cilia cause a condition called 'Primary Ciliary Dyskinesia' (PCD). PCD is characterized by neonatal respiratory distress, chronic respiratory tract infections, bronchiectasis and sinusitis (Horani et al., 2014). *Situs solitus* is the term used to describe a normal arrangement of body anatomy, where *situs inversus* is the mirror image of this arrangement. A failure of asymmetry of the unpaired organs is called *situs ambiguus* or heterotaxy, and is associated with major CHD. Around 50% of PCD patients have *situs inversus* due to defects in the nodal cilia and therefore dysfunction in development of asymmetry in the early embryo. Both

heterotaxy and complex CHD are more common in PCD than in the general population, with an incidence of around 6.3% (Kennedy et al., 2007).

1.7 Gene Identification in Congenital Heart Disease

In the past 10 years more than 60 genes associated with human CHD have been identified. CHD gene identification has characteristically been difficult, with the majority of cases being sporadic occurrences and the rarer familial cases being often complicated with variable expression, differences in age of presentation and variable penetrance (Blue et al., 2017). Mutations may also involve non-exonic DNA, such as regulatory regions, which can be missed by modern technologies and confirmation of their pathogenicity is more problematic. CHD gene discovery is further limited as many identified mutations are family-specific or private mutations, making genotype-phenotype correlation difficult (Chaix et al., 2016).

Original techniques for gene discovery focused either on positional cloning or candidate gene analysis. Positional cloning using linkage analysis was used successfully in CHD to identify mutations in genes such as *NKX2-5* (Schott et al., 1998) and *ACTC1* (Matsson et al., 2008). Similarly, the candidate gene approach uses information from animal models, known protein function, expression patterns or developmental pathways to identify likely functional candidates as the focus of analysis in affected individuals or families. In some cases, CNVs, such as microdeletions or translocations, were used to identify functional candidates. This was the case in the identification of the role of *GATA4* in human CHD (Pehlivan et al., 1999).

These approaches were often undertaken in families where there was an apparent Mendelian pattern of inheritance. However, large families segregating CHD are rare, limiting the possibilities for gene discovery with these techniques.

1.7.1 Next Generation Sequencing Technologies

The development of massively-parallel sequencing (or Next-Generation Sequencing, NGS) has significantly reduced the cost and increased the speed of DNA sequencing. The original sequencing of the complete human genome took around 10 years and cost nearly \$3 billion (Lander et al., 2001). NGS technologies have reduced the cost of Whole Genome Sequencing (WGS) to around \$1000-\$1500 per genome and the process takes a matter of days. NGS describes new technologies that, instead of undertaking sequencing using the classic Sanger chain-termination method, allow for large-scale multiplex sequencing-by-synthesis. The Leeds Sequencing Facility use

Illumina technologies (<https://www.illumina.com/technology/next-generation-sequencing.html>) on an Illumina HiSeq2500 or HiSeq3000 machine to undertake NGS, which uses a proprietary reversible ‘Sequencing-by-Synthesis’ technology (see chapter 2, section 2.2.14). NGS technologies can be used to undertake sequencing of custom gene panels, Whole Exome Sequencing and Whole Genome Sequencing.

1.7.2 Whole Exome Sequencing

The advent of NGS has altered the landscape of genetic research. Now genes can be sequenced at a faster speed, at reduced cost and a hypothesis-free approach to gene discovery can be undertaken. WES is the use of NGS technologies to sequence the entire coding sequence of DNA (around 1.6% of the whole genome), following enrichment with a capture approach. This has been used successfully in gene discovery for many Mendelian conditions, including familial CHD (Greenway et al., 2014). This technique often involves variant comparison between affected and unaffected family members or affected and unaffected cohorts to identify putative pathogenic variants.

Autosomal recessive disorders, particularly in consanguineous families, have traditionally been a focus in some WES studies, as the analysis of variants is more directed when searching for shared homozygous or biallelic variants in affected individuals. In CHD, examples of successful identification of causative mutations in recessive disease using WES include *SHROOM3* (Tariq et al., 2011), and *WDR16* (Ta-Shma et al., 2015). However, the investigation of recessive causes of CHD has, to date, been relatively understudied (Zaidi and Brueckner, 2017).

WES in the setting of autosomal dominant inheritance commonly uses linkage analysis of a large pedigree to indicate a likely disease locus (Boycott et al., 2013). LaHaye et al., 2016, performed WES in six families with autosomal dominant inheritance of CHD, four with ASD and two with PDA. They identified likely pathogenic mutations in *GATA4*, *TLL1* and *MYH11*, demonstrating the clinical utility of WES in this setting (LaHaye et al., 2016). Some difficulties arise in the presence of non-penetrance, which has been observed in CHD families, for example in well-characterized genes such as *NOTCH1* (Kerstjens-Frederikse et al., 2016). This may indicate unidentified modifier alleles are present that affect disease penetrance in these families, and suggests that epistasis between alleles is an important determinant of phenotypic variability.

WES has been used successfully in the setting of *de novo* dominant mutations, particularly with the use of parent-child trios (Gibson et al., 2012). Due to the large number of sporadic cases of non-syndromic CHD, it is supposed that *de novo* mutations may play a major contribution (Zaidi and Brueckner, 2017). In a large cohort of 1213 trios, which included CHD with extra-cardiac malformations and neurodevelopmental abnormalities, *de novo* mutations were found to be present in a high number of CHD cases associated with syndromes and other abnormalities (20% of syndromic cases, but only 2% of isolated CHD cases) (Homsy et al., 2015).

In a further example, Sifrim *et al.*, 2016 undertook a large exome sequencing study of patients with both syndromic and non-syndromic CHD, examining 1891 probands. In the syndromic cases there was an increase in *de novo* protein truncating variants, however, in non-syndromic CHD there was an enrichment of truncating variants inherited from unaffected parents, compared to controls. This finding suggests a level of non-penetrance in non-syndromic CHD and implies the effect of modifier alleles, for example intronic SNVs that affect splicing. This large exome study was able to identify new disease genes including *CDK13* and *PRKD1* (Sifrim et al., 2016).

It is likely that NGS technology will continue to become increasingly less costly and quicker and easier to perform, meaning that small families and those with more complex inheritance patterns may be examined.

1.7.2.1 Whole Exome Sequencing in Consanguineous Unions

In the ethnically-diverse region of Bradford in Yorkshire, there is a large community of residents of Pakistani origin. Around 40% of births in this community are to first cousin unions (Sheridan et al., 2013). For medical purposes, consanguinity is defined as a union between second degree cousins or closer relatives. Populations where consanguineous marriage is common have high rates of autosomal recessive disease, due to the increased likelihood of homozygosity. In the Bradford community, it has been shown that consanguineous union doubles the risk of congenital anomaly (multivariate risk ratio [RR] 2.19, 95% CI 1.67–2.85), including CHD (Sheridan et al., 2013). There are now 200 distinct recessive disorders recorded in the Bradford population. In other communities, consanguinity, particularly first cousin unions, has been shown to be associated with an increased incidence of CHD (Becker and Al Halees, 1999, Sadiq et al., 1995, Hoffman and Kaplan, 2002, Yunis et al., 2006).

The increased prevalence of autosomal recessive disorders in the Bradford community offers a unique research opportunity. Previously, gene identification in this community used the application of autozygosity mapping in larger consanguineous families to identify shared homozygous regions between affected individuals. However, more recent work has allowed the identification of novel disease genes without the need for prior autozygosity mapping, even in trios and sporadic cases (Parry et al., 2013). It is therefore now feasible to study smaller families, even in the presence of genetic heterogeneity.

1.7.3 Whole Genome Sequencing

Whole Genome Sequencing is the sequencing of an entire organism's DNA sequence. In variant discovery, WGS still has some limitations over WES, although it has been used successfully to identify candidate disease genes (C. Yuen et al., 2017). Current difficulties include a lack of comprehensive databases of normal variation, standardized pipelines for data analysis and availability of storage for large data files. These factors are likely to be overcome in the near future, with decreasing costs and increasing understanding of the non-coding regions. This will lead to an improved understanding of the effect of mutations in enhancers and promoter regions on disease. This may be very relevant in CHD disease-gene research, for example a homozygous variant in a *TBX5* enhancer was recently shown to be the cause of disease in a patient with septal defects (Smemo et al., 2012). WGS, as it is lacking a capture step, can also provide more complete coverage of the exomic regions than WES (Belkadi et al., 2015) and it can lead to detection of intronic Single Nucleotide Variants (SNVs), CNVs and translocations, which is not possible with WES.

1.7.4 GWAS

The complex nature of non-syndromic CHD suggests it may often be a multifactorial disorder and behave like a complex trait. Genome Wide Association Studies (GWAS) have become increasingly utilised in many common disorders to identify the association between the disorder and common low-penetrance variants. This methodology has only recently been applied in the setting of CHD.

Goodship *et al.*, 2012, studied 207 Single Nucleotide Polymorphisms (SNPs) in 22 candidate genes in their cohort of 362 non-syndromic patients with TOF, 717 unaffected parents and 183 unrelated controls. Candidate genes were selected on evidence of involvement in outflow tract development from transcriptional studies or mouse models. They found a significant association between a SNP in the *PTPN11*

gene (rs11066320) and non-syndromic TOF and this was replicated in a further large cohort. Gain of function mutations in *PTPN11* are responsible for Noonan Syndrome and the authors hypothesised that this common SNP may up-regulate the RAS-MAPK pathway to a lesser extent to that seen in Noonan syndrome, leading to a raised susceptibility to TOF (Goodship et al., 2012).

In 2013, Cordell *et al.* undertook a large international GWAS study involving over 5000 patients, divided into three major categories of CHD comprising septal, obstructive and cyanotic defects. Analysis of over 500,000 SNPs was undertaken in these cases and over a thousand controls. In the initial analysis, no SNP was found to show reliably significant association. However, when patients were divided into diagnostic groups, three SNPs on chromosome 4p16 were found to be significantly associated with ASD. The most significant of these, rs870142, conferred an Odds Ratio of 1.46 ($p = 9.5 \times 10^{-7}$). This SNP lies in the region between the *STX18* and *MSX1* genes. Large deletions of this region cause Wolf-Hirschhorn syndrome, known to be associated with cardiac defects. *MSX1* is expressed in the atrial septum in the mouse and is known to interact with *TBX5* (Cordell et al., 2013).

In another recent paper, Hu *et al.*, 2013, carried out a GWAS study looking at septal defects, including ASD, VSD and AVSD in the Han Chinese population. They assessed over 700,000 SNPs in 957 cases and 1,246 controls. Two SNPs (rs2474937 at 1p12 and rs1531070 at 4q31.1) were found to have significant association in this group and in a validation sample. rs2474937 is located near to the *TBX15* gene and rs1531070 is within the *MAML3* gene, which are both expressed in high levels in human cardiac tissue. *MAML3* plays an important role in the Notch signalling pathway (see section 1.5.2.2) (Hu et al., 2013).

1.8 Clinical Utility of Gene Identification

The identification of new genetic aetiologies in CHD are likely to assist in patient management, both in the fields of preventative care and in therapeutics. Identification of monogenic causes of CHD allows for genetic investigation and testing, providing knowledge for patients on inheritance patterns, risk assessment, screening for relatives, and allows clinical teams to counsel families on recurrence, prenatal diagnosis and prenatal intervention. Furthermore, gene identification allows for pathophysiological insights into the familial cases of CHD, which may help to delineate causes in sporadic cases (Chaix et al., 2016).

Genetic testing in the context of CHD in a clinical setting may give information on potential other non-cardiac organ involvement, or other associated cardiac complications. For example, in *NKX2-5*-related ASD, atrioventricular block and progressive ventricular dysfunction are a further risk (Pashmforoush et al., 2004). Also, mutations in *MYH6* can cause structural heart disease as well as cardiomyopathy (Granados-Riveron et al., 2010). Genetic testing may also be used to identify CHD patients who are more at risk of poor neurodevelopmental outcomes, allowing for earlier supportive intervention (Zaidi and Brueckner, 2017).

A genetic basis for disease may allow for prognostic information for affected patients, as disease stratification according to genetic aetiology becomes more available, and factors such as post-operative survival or long-term outcome may be assessed according to underlying cause. As an example of this, *endothelin-1* c.5665G>T (p.K198N, rs5370) has been associated with improved transplant-free survival in patients with Hyperplastic Left Heart Syndrome (HLHS) (Kirshbom et al., 2008).

Ultimately, the goal of genetic diagnosis in CHD is therapeutic potential, and targeted curative therapy. This is currently beyond our reach, but will theoretically be possible in the future. An encouraging example of this type of work is the knockout *Wnt2* mouse model, where AVSD could be rescued on activation of Wnt signaling (Tian et al., 2010).

1.9 Aims

The overall objective of this research project was the identification of novel CHD genes by Whole Exome Sequencing. The methods of this study were to identify families where there was a likely monogenic cause of CHD, so where the family structure was consistent with a Mendelian pattern of inheritance, and/or there was parental consanguinity. Both syndromic and non-syndromic forms of CHD were included and 'CHD' was broadly defined to include both structural defects and congenital or early onset myopathies. As well as the formation of partnerships with local clinicians in the Leeds-Bradford area, collaborations were formed with other genetics and cardiac centers nationally for the recruitment of patients. The full inclusion criteria are given in appendix A.

The following were specific objectives of the project:

- 1) To identify families with a likely Mendelian inheritance of CHD, according to inclusion criteria, obtain informed consent for research and to gain appropriate DNA samples from consented family members.

- 2) To perform WES for selected individuals from recruited families and to use a bioinformatic workflow of filtering sequence variant data to identify novel CHD genes.
- 4) To confirm the pathogenicity of putative mutations by comparison with exome sequencing data, both within local cohorts and/or by screening genes in other cohorts available by collaboration.
- 5) To perform functional characterization of new disease proteins using established biochemical and cell biology techniques. The pathogenic potential of putative missense mutations could then be investigated using *in vitro* cell and zebrafish morphant models.

2 Materials and Methods

2.1 Materials

Room Temperature (RT) is defined as 20-22°C.

2.1.1 General Reagents

Water	Milipore
Ethanol	Sigma-Aldrich
Methanol	Sigma-Aldrich
Phosphate Buffered Saline (PBS) 1x	Sigma-Aldrich

2.1.2 Solutions

2.1.2.1 PBST

1x	PBS
0.1% [v/v]	Tween20 (Sigma-Aldrich)

2.1.2.2 Tris-EDTA (TE) Buffer 1x

10mM	Tris HCl (pH 8.0) (Sigma-Aldrich)
1mM	EDTA (pH 8.0) (Ambion®, Thermo Fisher Scientific, Waltham, MA, USA)

2.1.2.3 Tris-Acetate-EDTA (TAE) Buffer 50x

2M	Tris HCl
50mM	EDTA (pH 8.0)
0.97M	Glacial acetic acid (Sigma-Aldrich)

2.1.2.4 Tris-Borate-EDTA (TBE) Buffer 10x

890mM	Tris HCl
890mM	Boric Acid (Sigma-Aldrich)
20mM	EDTA

2.1.2.5 Gel Loading Dye 2x

50%[v/v]	Glycerol (Sigma-Aldrich)
10%	TBE 10x
0.1%[w/v]	Orange G (Sigma-Aldrich)
0.1%[w/v]	Xylene Cyanol (Sigma-Aldrich)

2.1.2.6 NP40 Lysis Buffer

1%[v/v]	NP40 (IPEGAL CA-630) (Sigma-Aldrich)
50mM	Tris-HCl (pH 8.0)
150mM	NaCl (Sigma-Aldrich)
1x	Protease Inhibitor Cocktail (Thermo Fisher Scientific)

2.1.2.7 IP Wash Buffer

150mM	NaCl
50mM	Tris-HCl (pH 8.0)
0.5mM	EDTA
0.1% [v/v]	NP40

2.1.2.8 IP Incubation Buffer

150mM	NaCl
50mM	Tris-HCl (pH 8.0)
1x	Protease Inhibitor Cocktail

2.1.2.9 SDS Loading Buffer

100 mM	Tris – HCl pH 6.8
4%[w/v]	SDS (BDH, Poole, UK)
20%[v/v]	Glycerol
20 mM	β – mercaptoethanol (BDH, Poole, UK)
0.004% [v/v]	Bromophenol Blue (Sigma-Aldrich)

2.1.2.10 Lysis Buffer for Cellular DNA Extraction

10mM	Tris pH 8.0
100mM	EDTA
20 μ g/ μ l	RNase (Sigma-Aldrich)
0.25%[w/v]	SDS

2.1.2.11 TAP Wash Buffer

1x	TBS
1x	Protease/Phosphatase Inhibitor Cocktail
0.1% [v/v]	NP40

2.1.2.12 D-desthiobiotin Elution Buffer

100mM	Tris HCl pH8.0
150mM	NaCl
1mM	EDTA

5mM D-desthiobiotin.

2.1.3 Cell Lines

hTERT- RPE1 cells	Source ATCC (Manassas, Virginia, USA)
HEK293 cells	Source ATCC (Manassas, Virginia, USA)
IMCD3 cells	Source ATCC (Manassas, Virginia, USA)

2.1.4 Primary Antibodies

Primary antibodies were used as detailed in table 2-1.

Table 2-1: List of primary antibodies.

Table includes the species the antibody was raised in, dilutions of antibodies used in IF, whole mount in situ hybridization and western blot, as well as the antibody supplier.

Antigen	Species raised in	Polyclonal/ monoclonal; Clone name; Isotype	IF dilution	Whole Mount Embryo dilution	Western blot dilution	Supplier
DNAAF1	Rabbit	Polyclonal	1/50	1/50		Santa-Cruz
DNAAF1	Rabbit	Polyclonal	1/100			Abcam
RPGRIP1L	Guinea Pig	Polyclonal	1/500			Proteintech
γ -tubulin	Mouse	Monoclonal; GTU-88; IgG1	1/1000			Sigma-Aldrich
Centrin 3	Mouse	Polyclonal	1/100			Merck
Acetylated alpha-tubulin	Mouse	Monoclonal; 6-11B-1; IgG2b	1/1000			Sigma-Aldrich
Poly glutamylated tubulin	Mouse	Monoclonal; GT335; IgG1k	1/100			AdipoGen, San Diego, CA, USA
Pontin	Mouse	Monoclonal; 5G3-11; IgG	1/200	1/200	1/1000	Sigma-Aldrich
Gamma-tubulin	Goat	Polyclonal	1/50	1/100		Santa Cruz
Reptin	Mouse	Monoclonal; 2E9-5; IgG	1/200	1/200	1/1000	Sigma-Aldrich
FLAG	Mouse	Monoclonal; M2; IgG1	1/500		1/1000	Sigma-Aldrich
ARL13B	Rabbit	Polyclonal	1/500	1/1000		Proteintech
ARL13B	Mouse	Monoclonal; N295B/66 IgG2a	1/500			Antibodies Inc, Davis, CA, USA
IFT88	Rabbit	Polyclonal	1/200			Proteintech
“Living Colours”	Mouse	A mixture of monoclonal antibodies against native fluorescent proteins, that recognizes GFP, YFP and CFP			1/1000	Clontech

2.2 Methods

2.2.1 Patient Identification

Families were identified through collaborators in Departments of Clinical Genetics and Departments of Paediatric & Adult Cardiology across the UK, including Yorkshire, Newcastle-Upon-Tyne and North West Thames (Kennedy-Galton Centre). To raise awareness of the research study, a presentation was given at the Yorkshire Regional Cardiac Network meeting (June 2013), at the Leeds Teaching Hospitals Paediatric Cardiac Surgery meeting (August 2013) and at the Yorkshire Regional Genetics weekly multidisciplinary meeting (August 2013). Verbal and written information was distributed to Clinical Geneticists, Genetic Counselors, Consultant Cardiologists, Cardiac Nurses and Cardiac Radiologists within Leeds Teaching Hospitals Trust, containing contact details for the research team and patient inclusion criteria.

Patients were ascertained according to inclusion criteria as detailed in Appendix A. Families identified in the clinical setting were provided with verbal and written information regarding the research project (see Appendix B). The lead consultant then referred to the family to the research team. On referral an appointment was made for the research team to meet the family, undertake a full clinical examination and obtain informed consent and DNA sampling as detailed below.

Furthermore, a systematic search of medical records was undertaken in the Department of Clinical Genetics, Chapel Allerton Hospital, Leeds. Medical records for patients coded under 'Congenital Heart Disease' were identified and assessed. Once families meeting the inclusion criteria were identified, a letter was sent from their lead consultant. Patients were only further contacted if they expressed interest and returned the consent form.

2.2.2 Consent

Interested families were contacted by telephone and a home visit or clinical appointment was arranged with Dr Hartill and Mrs Shabana Khan, a clinical research nurse. At the time of home visit, families were given verbal information about the study and were presented with the written information sheet. They were then asked to give written informed consent using the approved consent form (see Appendix C).

2.2.3 Ethical Approval

The study title associated with this project is 'Molecular genetic investigations of autosomal recessive conditions', granted ethical approval for 5 years by South Yorkshire LREC on 18th Feb 2011, ref. number 11/H1310/1. There is also local Leeds Teaching Hospitals Trust R&D approval (see Appendix D).

2.2.4 Clinical Assessment

A full detailed family pedigree, medical history and clinical examination, including cardiac examination and assessment of dysmorphic features was performed for all patients. The medical and clinical genetics notes, plus any previously performed radiological imaging was reviewed for all consented patients. Previously performed genetic testing, for example Array CGH, was assessed for all consented patients.

2.2.5 Clinical Investigation for PCD

Clinical investigations for PCD were performed in one family. These investigations were performed by a representative of a specialist PCD center (UCL, Institute for Child Health, London, UK). Nasal Nitric Oxide (NO) levels were assayed using a chemiluminescence analyser. Three NO measurements were taken per nostril, with the mean of these being taken as the NO level. Ambient NO level was recorded before each test. Respiratory epithelial cells were obtained by nasal brush biopsy and suspended in cell culture medium, before transfer to the London laboratory. Light, electron, and immunofluorescence microscopy were used to analyse ciliary motility and structure. Human Transmission Electron Microscopy (TEM) and high-speed video microscopy were carried out as previously reported (Shoemark et al., 2012).

2.2.6 DNA Sampling

Blood or saliva samples were obtained from affected individuals, parents and, where possible, unaffected siblings. These were collected by standard methods. Oragene saliva collection kits (DNA Genotek, Orasure Technologies Inc., Ottawa, Canada) were used for saliva collection.

2.2.7 DNA Extraction

2.2.7.1 Peripheral Blood Samples

DNA extraction from blood samples was performed by the Yorkshire Molecular Genetics Laboratory (St James' University Hospital, Leeds, UK). DNA was extracted from peripheral blood leukocytes and stored in Tris-EDTA buffer (pH 8.0) using a standard salt precipitation protocol.

2.2.7.2 Saliva Samples

DNA extraction from saliva samples was performed using prepIT L2P DNA extraction kits (DNA Genotek, Orasure Technologies Inc.), according to the manufacturer's protocol. DNA pellets were dissolved in 1x TE buffer for long-term storage.

2.2.7.3 Cell Culture

DNA extraction from cell cultures was performed using a standard phenol/chloroform extraction. In brief, cells were resuspended in a total volume of 200µl of lysis buffer (see 2.1.2.10) following one 1xPBS wash. They were incubated at 37°C for one hour. Proteinase K (Sigma-Aldrich) was added to a final concentration of 100µg/µl and incubated at 55°C for one hour. An equal volume of phenol and chloroform was added, vortexed and then centrifuged at 16,000xg for 5 minutes. DNA was precipitated from the aqueous phase in 100% ethanol with 0.2M NaCl. The extracted DNA was washed in 75% ethanol twice and dissolved in 1X TE buffer. The DNA concentration was determined using a Qubit BR dsDNA assay (Life Technologies, Thermo Fisher Scientific), see 2.2.14.

2.2.7.4 Paraffin-Embedded Slides

In one case, paraffin-embedded slides of renal tissue were available for DNA extraction. DNA extraction was performed twice from these slides, initially by Dr Gemma Hemmings, Leeds Institute of Cancer & Pathology (LICAP), University of Leeds, and the second extraction was performed by Dr Hartill, using identical methods. Initially samples were dewaxed in a xylene/ ethanol series. This involved immersing slides in xylene (Sigma-Aldrich) for 4 x 5 mins, 100% ethanol for 3 x 4 mins and then in decreasing ethanol concentrations for four minutes each. Microdissection was used to collect the tissue of interest from the slides. DNA extraction used the Formalin-Fixed Paraffin-Embedded (FFPE) kit from Qiagen (Germantown, MD, USA) and extraction was performed according to the kit protocol.

2.2.8 Whole Genome Amplification

Where a very low quantity of DNA was available, for example in the case of DNA extracted from FFPE slides, Whole Genome Amplification (WGA) was used to increase DNA quantity for Polymerase Chain Reaction (PCR). This was performed using the Illustra GenomiPhi V2 DNA amplification kit (GE Healthcare). 1µl (10ng) of template DNA was combined with 9µl of sample buffer. The DNA was denatured at 95°C for three minutes and then cooled to 4°C on ice. The amplification master mix was created from 9µl reaction buffer and 1µl enzyme mix (Phi29 DNA polymerase). This was transferred to the cooled sample. DNA amplification was carried out at 30°C for 18

hours (longer than suggested in the protocol) and then the enzymes were inactivated at 65°C for 10 minutes. The sample was cooled to 4°C and stored at -20°C. It was then used as usual in downstream applications.

2.2.9 RNA extraction

RNA was extracted from cells in culture. All solutions were prepared with RNase-free water. Growth media was removed from the culture dish. An appropriate volume of 1x Phosphate Buffered Saline (PBS) was added and cells were scraped off into the PBS and centrifuged for 5 minutes at 200xg at RT. The cell pellet then underwent a further 1xPBS wash and was spun down. Trizol® reagent (Life Technologies) was added to the cell pellet at a volume of 1ml of Trizol per 10cm² culture dish surface area. The cells were pipetted up and down several times to lyse. The homogenized sample was incubated for 5 minutes at RT. 0.2ml of chloroform was added per 1ml Trizol reagent, shaken and incubated for 2-3 minutes at RT. The sample was centrifuged at 12,000xg for 15 minutes at 4°C. The aqueous phase of the sample was removed. 0.5ml of 100% isopropanol per 1ml Trizol was added to the aqueous phase and the sample was incubated at RT for 10 minutes. The sample was then centrifuged at 12,000xg for 10 minutes at 4°C. The RNA pellet was washed in 75% ethanol and left to air dry. The RNA pellet was resuspended in RNase free water.

2.2.10 Polymerase Chain Reaction

2.2.10.1 Primer Design

Primers were designed using the Primer 3 program (<http://bioinfo.ut.ee/primer3/>) or Auto Primer 3 (<https://github.com/gantzgraf/autoprimers3>), which retrieves gene information from UCSC (the University of California Santa Cruz Genome Browser) and uses Primer 3 to design primers. Primers were designed to avoid common SNPs and parameters were set to produce primers with an optimum annealing temperature of around 60°C and with products of between 250-600bps. Primer sequences were checked using a BLAST tool (<http://blast.ncbi.nlm.nih.gov/Blast>) to ensure specific binding and avoidance of SNPs. Primers for genomic sequencing were designed to cover all exons and at least 15bp of flanking intronic regions. Overlapping primers were used in cases where exons were too large for the product to be amplified in a single experiment. Primer sequences used in this project are listed in appendix E.

2.2.10.2 PCR Reaction

PCR reactions were set up in a total volume of 10µl. 20ng of genomic DNA was combined with 2µM of forward and reverse primers (Sigma-Aldrich), 1 unit of Taq DNA

polymerase (ABgene), 0.25 μ M dNTPs (Thermo Fisher Scientific) and PCR buffer (75mM Tris-HCl buffer pH8.8, 20 mM (NH₄)₂SO₄, 0.01% Tween 20 and 1.5 mM MgCl₂). Alternatively 20ng of genomic DNA was combined with 2 μ M forward and reverse primers and 1 x HotShot Diamond PCR Master Mix (Clontech Life Science, Stourbridge, UK).

Reactions were performed in a Veriti Thermal Cycler (Thermo Fisher Scientific). Denaturation was at 95°C for 5min, followed by 45 cycles of 94°C for 15 seconds, the appropriate annealing temperature for 15 seconds, 72°C for 30secs and then finally extension at 94°C for 30secs. PCR products were visualized by agarose gel electrophoresis (see section 2.2.10.3).

Where necessary, for GC rich regions, reactions were supplemented with 5 x Combinational Enhancer Solution (5 x CES); 2.7 M betaine, 6.7 mM DTT, 6.7% DMSO and 55 μ g/ml BSA.

2.2.10.3 Agarose Gel Electrophoresis

Samples were mixed in a 1:1 [v/v] ratio with 2x loading buffer. The gel was composed of molecular biology grade agarose powder (Bioline, London, UK) dissolved in 1 x TAE to a concentration of 2% [w/v]. Ethidium bromide (Sigma-Aldrich) was added to a final concentration of 0.5 μ g/ml or Midori Green (Geneflow, Staffordshire, UK) at 5 μ l per 100ml gel. A DNA size standard ladder was used (Easy Ladder I, Bioline). An electric field was applied at 120V for 40min. The gel was visualized on a UV transilluminator and displayed using Image Lab (v. 4.0) analysis software (Bio-Rad, Hemel Hempstead, UK).

2.2.11 Purification with ExoSAP

Prior to sequencing, samples were purified with ExoSAP-IT® (Affymetrix, Thermo Fisher Scientific) to remove unused primer and dNTPs. For 2.5 μ l of PCR product, 1 μ l of ExoSAP-IT was added. Samples were incubated for 45min at 37°C and then inactivated for 15min at 80°C.

2.2.12 Sanger Sequencing

ExoSAP-IT purified PCR products were sequencing using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific). 1 μ l of product was combined with 1 μ l of BigDye®, 1 μ l of forward or reverse primer at a final concentration of 0.2 μ M, 1.5 μ l of BigDye® buffer and 5.5 μ l of dH₂O.

Reactions were performed in a Veriti Thermal Cycler (Thermo Fisher Scientific). Denaturation was at 96°C for 1min, followed by 45 cycles of 96°C for 10 seconds, 50°C for 15 seconds, and 60°C for 4 minutes. Temperatures were ramped at 1°C/second.

Precipitation was performed by adding 1µl of 0.25M EDTA to the sample, followed by 31µl of 100% ethanol. This was centrifuged at 3900xg for 30 minutes at 4°C. The plate was then inverted onto tissue and spun upside down at 200xg for 5 seconds. Next, 60µl of 70% [v/v] ethanol was added and the sample was centrifuged at 3900xg for 15 minutes. Then the plate was inverted and spun at 200xg for 5 seconds. The sample was then left to air dry for 15min at RT under foil or on a hot block at 37°C for a few minutes. To resuspend the samples, 10ul HiDi Formamide (water-free and deionized, Applied Biosystems, Thermo Fisher Scientific) was added. The plate was sequenced on an ABI3130xl Genetic Analyzer (Applied Biosystems) using polymer POP7 (Applied Biosystems). Resulting data was analysed using Seqscape v2.5 (Applied Biosystems) or 4Peaks (Mek&Tosj.com).

2.2.13 Genotyping with SNP array

DNA for Affymetrix Genome-Wide Human SNP Array 6.0 was sent from selected family members to AROS Applied Biotechnology, Denmark, for commercial processing. Received data was analyzed using SnpViewer (<https://sourceforge.net/projects/snpviewer/files/>). This program allows homozygous regions of the genome to be identified using colour-coded bars for homozygous and heterozygous SNP calls. Affected and unaffected family members can then be assessed together by eye or by using the program's 'find regions' tool.

2.2.14 Whole Exome Sequencing

DNA samples for WES were quantified using Quant-iT dsDNA assay (Lifetechnologies, Waltham, MA, USA), to ensure an appropriate amount of DNA was used for shearing. The Quant-iT Reagent was diluted to 1:200 in Quant-IT buffer. 1µl of sample DNA or 10µl of control standard were added to the assay tube and vortexed for 2-3 seconds. The tubes were incubated for 2 minutes at room temperature and read in the Qubit fluorometer. The concentration was calculated by multiplication with the dilution factor.

Two methods for WES were undertaken, the SureSelect XT method (Agilent, Santa Clara, CA, USA), which uses 1-3µg of DNA, and the SureSelect QXT method (Agilent, Santa Clara, CA, USA), which uses 25ng DNA.

2.2.14.1 Sure Select XT method

The Sure Select XT method was performed using either the SureSelect XT Human All Exon V4 or V5 kits (Agilent Technologies). DNA shearing was carried out for 1-3µg of DNA using a Covaris S2 Adaptive Focused Acoustics Sonicator (Covaris, Woburn, MA, USA). This method uses focused bursts of ultrasonic acoustic energy to create 200-300bp fragments of DNA. Standard settings of 10% Duty Factor, Peak Incident Power 175, 200 Cycles Per Burst, treatment time of 360s and bath temperature 4-8°C were used. Following shearing, the quality and quantity of the prepared library was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), as per the manufacturer's protocol. Quantification was performed using the 2100 Bioanalyser software (Agilent) by interrogating the area under the curve. The SureSelect XT library preparation was carried out according to the kit instructions and involves the following steps: repairing the ends of sheared DNA, adding 'A' bases to the 3' end of the DNA fragments, ligating index-specific paired-end adaptors and amplification of the adaptor-ligated library using a 6-cycle enrichment PCR. Before each step, the sample is purified using Agencourt AMPure XP beads. Figure 2-1 shows a summary of the Agilent SureSelect workflow for paired-end sequencing.

Following amplification, the quality and quantity of the prepared library was further assessed using the Agilent 2100 Bioanalyzer. A Savant SpeedVac 110 vacuum concentrator (ThermoFisher Scientific, Waltham, MA, USA) was used where necessary to achieve a prepared library concentration of 221ng/µl in 3.4µl (750ng) before hybridisation. The prepared genomic sample was hybridized to SureSelect biotinylated RNA library baits and hybrid capture was carried out using streptavidin coated magnetic beads. For hybridization, 5.6µl of SureSelect Block mix was added to 3.4µl of 221ng/µl prepared library. This was heated to 95°C for 5 minutes and then held at 65°C. This stops adapted-ligated DNA hybridising end-to-end. 13µl of Hybridisation buffer was added to 7µl of the SureSelect capture library mix and 13µl of the resulting reagent was added to the prepared library. This was maintained at 65°C for 16 hours.

Capture of the resulting hybridisation was performed using Dynabeads® MyOne™ Streptavidin T1 (Life Technologies) following incubation at room temperature for 30 minutes. SureSelect index tags were added to each sample by post-hybridization amplification, with a total of 12 cycles, and the quality was finally assessed, using a high-sensitivity protocol, on an Agilent 2100 Bioanalyzer. Equal molar quantities of each captured sample were pooled and diluted to 10nM in nuclease-free water. Six

samples were pooled per lane for multiplex sequencing on an Illumina HiSeq 2500 DNA sequencing system (Illumina, San Diego, CA) in 100bp paired end mode.

Cluster amplification and generation, as well as subsequent sequencing was performed by the Leeds Next Generation Sequencing Facility (Leeds, UK, <http://dna.leeds.ac.uk/genomics/>) (see Fig. 2-2). The

Next Generation sequencing facility uses an Illumina HiSeq 2500 (and more recently a HiSeq 3000), which use 'clonal bridge amplification' and a 'sequencing by synthesis' reversible dye-termination method (Bentley et al., 2008) (Figure 2-2). This process involves washing the prepared DNA fragments onto a flow cell where oligonucleotides complementary to the adaptors are situated. Each DNA fragment forms a bridge between two adaptors and is then PCR amplified, the so-called 'bridging amplification', which forms clusters of identical DNA sequences on the flow-cell. Fluorescently labeled reversible terminators are washed across the flow cell and a single terminator is added to each strand. As each terminator is incorporated and is laser-excited, all of the strands in a cluster fluoresce with the same colour, which is detected by the sequencer. Following this, the terminator is chemically inactivated so that the next cycle can commence. This process is termed 'sequencing by synthesis' The resulting recorded sequence is called a 'read'. In paired-end sequencing, once the first strand reaction is complete, the process is repeated for the opposite strand (see fig 2-2).

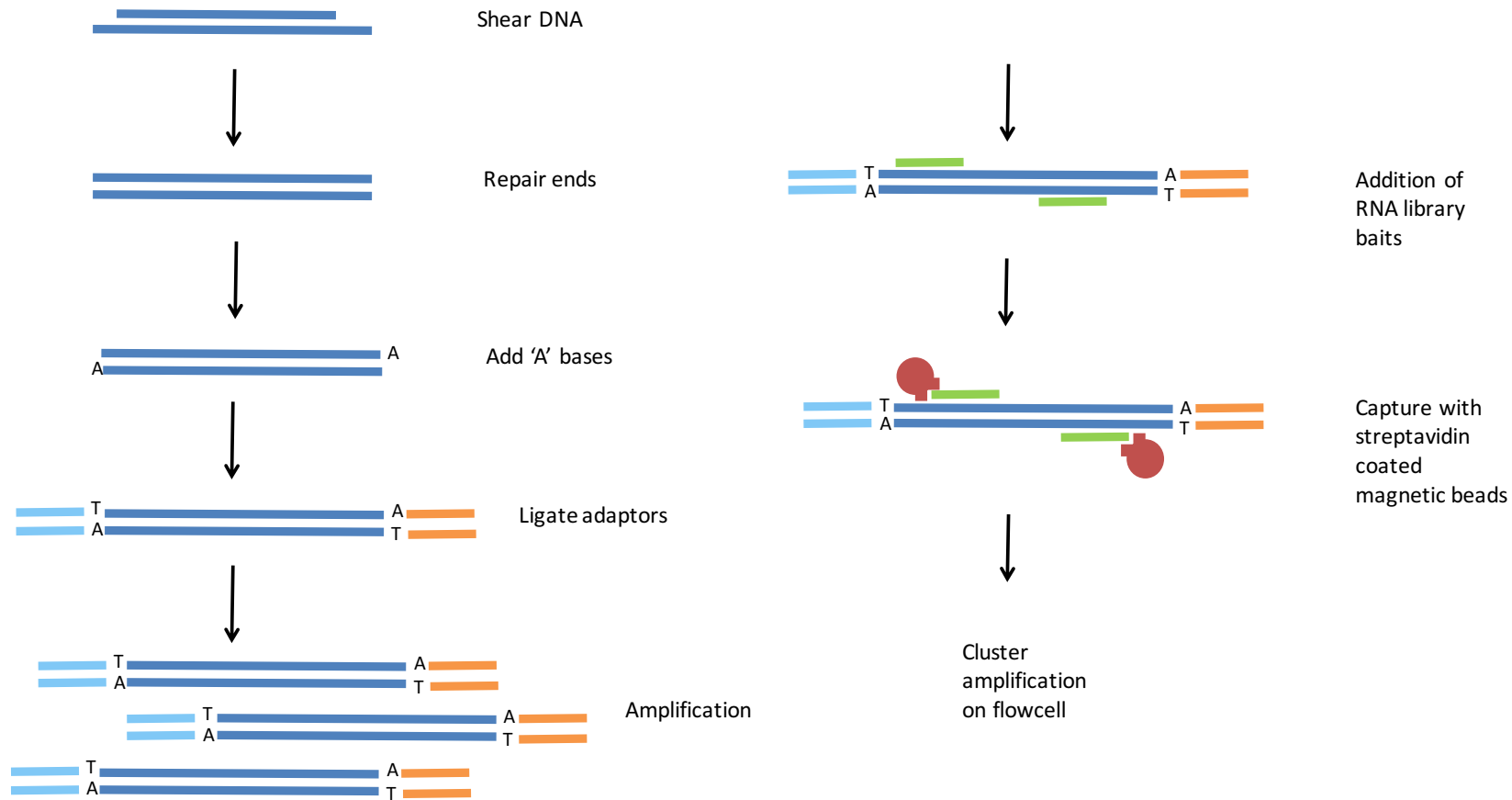


Figure 2-1: Agilent SureSelect XT workflow for Paired-end Sequencing, DNA shearing to exome capture with addition of biotinylated RNA library baits.

In this protocol, DNA is sheared using sonication, the ends of sheared DNA are repaired, 'A' bases are added to the 3' end of the DNA fragments, index-specific paired-end adaptors are ligated and the adaptor-ligated library is amplified using a 6-cycle enrichment PCR. Library capture involves the addition of RNA library baits, which are captured with streptavidin-coated magnetic beads before addition of the prepared DNA fragments to the flow-cell.

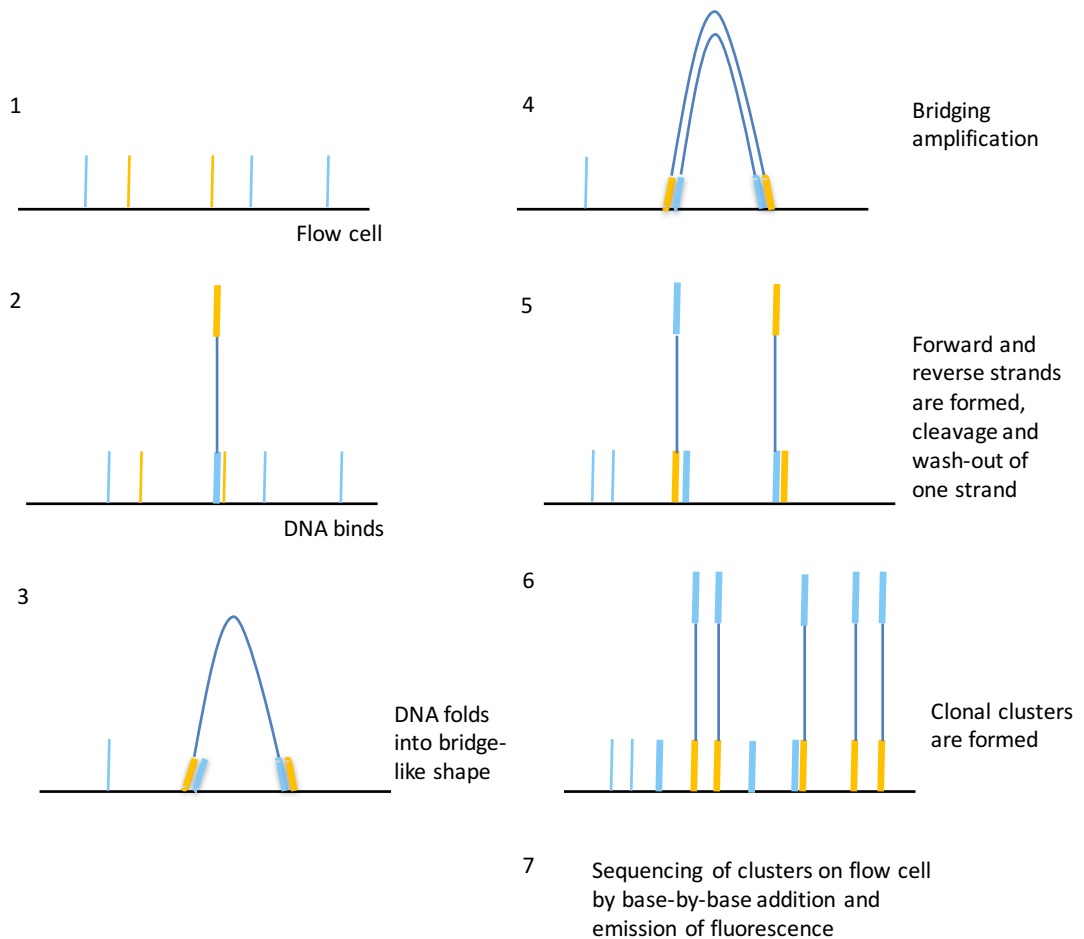


Figure 2-2: Illumina sequencing protocol, showing cluster amplification on a flow cell.

Prepared DNA fragments are washed onto the flow cell where complementary oligonucleotides are situated. Each DNA fragment forms a bridge between two adaptors and is then PCR amplified ('bridging amplification'). This forms clusters of identical DNA strands on the flow-cell. Fluorescently labeled reversible terminators are washed across the flow cell and a single terminator is added to each strand ('base-by-base addition'). Each base is laser-excited, causing all of the strands in a cluster to fluoresce with the same colour, detected by the sequencer. Following this, the terminator is chemically inactivated so that the next cycle can commence.

2.2.14.2 Sure Select QXT Method

The Sure Select QXT method was performed using either the Sure Select QXT Human All Exon V5 or V6 kits (Agilent Technologies). The kit protocol was followed exactly. Each DNA sample was prepared to a dilution of 25ng/μl using the Qubit ds BR assay. Fragmentation of DNA samples and addition of adaptors was performed in a single enzymatic step, using SureSelect QXT Enzyme Mix ILM, with an incubation time of 10 min at 45°C. Purification following each step was carried out using AMPure XP beads. The adaptor-tagged library was amplified using Herculase II fusion DNA polymerase in an 8 cycle PCR reaction. DNA library quantity and quality was assessed on the Agilent 2100 Bioanalyzer, as before.

For hybridization of the DNA samples to the capture library, 1500ng of the prepared DNA library was made up to 12µl in nuclease-free water. 5µl of SureSelect QXT Fast Blocker Mix was added and the samples were held at 65°C for 10 mins. 25% RNase block and the capture library, along with a buffer, were added and a thermal cycler was used to perform 60 cycles of 65°C for 1 min and 37°C for 3s. Dynabeads® MyOne™ Streptavidin T1 (Life Technologies) magnetic beads were used for DNA hybrid capture following a 30min incubation at 65°C. A magnetic separator was used for bead capture. Index tags were added using Herculase II Fusion DNA polymerase in a 10 cycle PCR. A further assessment of DNA quality and quantity was performed on an Agilent 2100 Bioanalyzer, using a high sensitivity chip, following purification with AMPure XP beads. Equal molar quantities of each captured sample were pooled and diluted to 10nM in nuclease-free water. Nine or ten samples were pooled per lane and sequencing was performed on an Illumina HiSeq 3000 DNA sequencing system by the Leeds Next Generation Sequencing Facility (Leeds, UK, <http://dna.leeds.ac.uk/genomics/>).

2.2.14.3 WES Data Analysis

For each sample, a standardized data analysis workflow was followed. Data was initially collected from the Hiseq2500 or Hiseq3000 in .fastq format. Reads were aligned to the reference sequence (GRCh37, Genome Reference Consortium human reference assembly build 37) using Novoalign (Novocraft Technologies, Selangor, Malaysia) or BWA (Burrows-Wheeler Aligner (Li and Durbin, 2009)) to create a .sam (sequence alignment map) file. Each .sam file was sorted into chromosome order with Picard v1.85 (<http://picard.sourceforge.net>) and converted to .bam (binary alignment map) format. Picard tools were used to mark and remove duplicate reads. The Indel Realigner function of the Genome Analysis Toolkit (GATK, Broad Institute, USA) was used to perform realignment around indels (insertions and deletions). The BaseRecalibrator (GATK) was used to create a recalibration model, which was assessed by creating recalibration plots. Base quality score recalibration was then applied. This recalibrates quality scores for all bases in a read, taking into account the reported quality score, position of the base within a read, and the preceding and concurrent nucleotide.

Variants were called using HaplotypeCaller (GATK). This was predominantly used over the Unified Genotyper (GATK) due to the reported superior calling of indels (Dr David Parry, personal communication). This program can be multithreaded for multiple samples. This created a .vcf (variant call format) file containing variants from all family

members in a single file. Familial samples were called together to increase confidence in the calls.

Variants were hard filtered using the VariantFiltration programme (GATK). Indels and SNPs were separated before hard filtering. This program labels variants falling below certain specified parameters. Parameters used for indels and SNP/MNPs are shown in Table 2-2.

Table 2-2: Parameters used for hard filtering of variants.

Variation	Abbreviation	Character	Parameter to be marked
Indel	QD	Quality by Depth	<2.0 reads
	FS	Fisher Strand PHRED scaled p-value to detect strand bias	>200.00
	Readposranksum	Mann-Whitney Test for distance from end of read for alternate allele in het call	<-20
SNP/MNP	QD	Quality by Depth	<2.0 reads
	FS	Fisher Strand	>60.00
	ReadPosRankSum	Distance from end of read for alternate allele	<-8
	mappingqualityranksum	Mann-Whitney Rank Sum Test for Mapping Qualities (reference vs alternate allele)	<-12.5

Variants were then further filtered to remove those with a mean allele frequency (MAF) of above a user specified cut-off in The National Center for Biotechnology Information's (NCBI's) database of single nucleotide polymorphisms (dbSNP) version 129 and previous (Sherry et al., 2001), the NHLBI Exome Sequencing Project (ESP) (<http://evs.gs.washington.edu/EVS/>), the Broad Institute's Exome Aggregation Consortium (ExAC) version 0.3 database (<http://exac.broadinstitute.org>) and 3000 ethnically-matched in-house control exomes. Cut-off values for MAF were $\geq 1\%$ for suspected rare recessive disease and $\geq 2\%$ for suspected rare dominant disease. Variants were then annotated using the Variant Effect Predictor (Ensembl, (McLaren et al., 2010)) and likely 'functional' variants were identified (indels within coding regions, non-synonymous SNPs or alterations of splice consensus sequences). Samples were filtered on family structure using bespoke perl scripts 'vcf hacks' (Dr. David Parry, <https://github.com/gantzgraf/vcfhacks>) and annotated to give gene information in Microsoft Excel format. These scripts allow for a family pedigree to be utilized in .ped format or for certain characteristics, for example to filter for biallelic variants present in

affected individuals and absent in unaffected family members, to be identified. The pathogenicity of putative mutations was assessed using PolyPhen2 (<http://genetics.bwh.harvard.edu/pph2/>) (Adzhubei et al., 2010), SIFT (Sorting Intolerant From Tolerant, <http://sift.jcvi.org>) (Ng and Henikoff, 2003) and Condel (<http://bg.upf.edu/fannssdb/>). Variants were ranked according to Combined Annotation Dependent Depletion (CADD) score version 1.3 (<http://cadd.gs.washington.edu>). Remaining variants were annotated to give a summary of known gene function and mouse model information. For all remaining candidate variants a literature review was undertaken including assessment of interacting partners, expression pattern and animal models. Further to filtering in this manner, variants falling within known disease-causing genes were checked using *in silico* pathogenicity prediction tools (PolyPhen2, SIFT, Condel and CADD score) and segregation in affected family members, even if they were not present following filtering steps.

The workflow for WES data analysis is summarized in Figure 2-3. Before using this pipeline on the study samples, it was trialed on data from a patient with a known mutation to demonstrate that this as an appropriate variant filtering strategy.

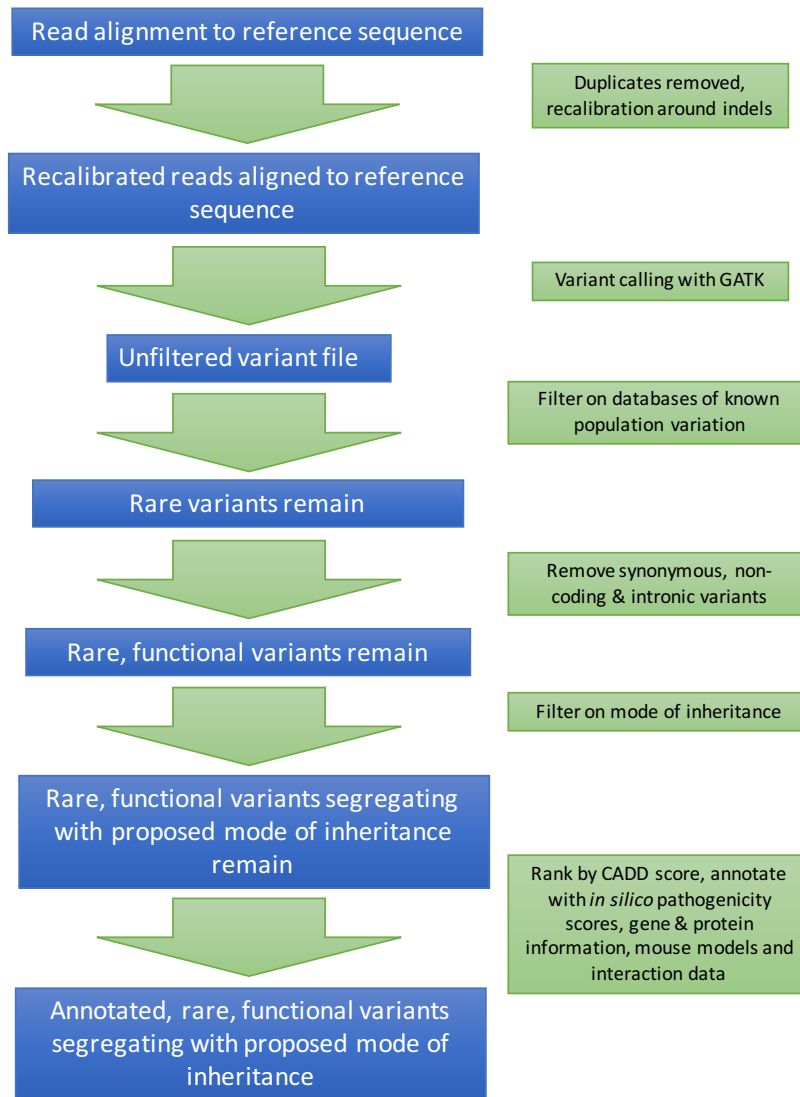


Figure 2-3: Data analysis pipeline used for WES data.

2.2.14.4 Assessment of Mutation Pathogenicity

Variants identified following WES filtering methods were further assessed for likely pathogenicity by the following methods:

- (a) Selection of potential pathogenic variants in candidate genes on the basis of putative function in cardiac development and/or members of known developmental pathways from literature review and known protein interaction data (String Database; <http://string-db.org>).
- (b) Collaborators from The Netherlands provided the group with a list of genes obtained from co-expression modeling with known ciliary proteins. These were used to screen potential candidate variants (unpublished communication; Dr Dorus Mans, UMC Utrecht, The Netherlands).
- (c) Comparison to a list of candidate ciliary genes from an Austrian collaborator. These were identified in *Drosophila* by tissue-specific RNAi-mediated depletion in sperm (motile cilia that form in an IFT-independent manner), neurons (non-

motile cilia that form in the canonical, IFT-dependent manner), muscle (non-ciliated) and throughout the animal. Mutants were then examined in *C. elegans* where possible. This gave a list of 72 novel genes, which presented clear cilia phenotypes in the fly and/or worm. This list was used to screen against potential candidates (unpublished communication; Prof. Alex Dammermann, University of Vienna, Austria).

- (d) Comparison to RNA expression data. RNA expression data from fetal heart tissue was downloaded from publically available resources (NCBI, Sequence Read Archive data) (<http://www.ncbi.nlm.nih.gov/sra>). This was analysed using Cufflinks (Trapnell et al., 2009) after alignment with the “splice aware” aligner TopHat2 (Trapnell et al., 2009). This list of highly expressed genes was compared to candidates from WES.
- (e) Structural modeling of the position of candidate variants within a protein or specific protein domain. This was performed using tools such as Cn3D (<http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml>) and Swiss Pdb-viewer (<http://www.expasy.org/spdbv/>).
- (f) Where possible, screening further cohorts of patients with the same phenotype for mutations in a candidate gene. This was often through collaboration.

2.2.14.5 WES Assessment of Quality of Data

To assess the quality of data obtained from WES, a number of methods were used. Firstly, GATK’s Depth Of Coverage tool was used to assess average read depth across the exome (specified using a .bed file) and percentage of positions with coverage >5 reads. This tool can also give depth across a certain gene. Secondly, Picard tools CollectMultipleMetrics command was used (<https://broadinstitute.github.io/picard/command-line-overview.html#CollectMultipleMetrics>). This runs a number of alignment and quality metrics including quality score distribution, insert size metrics and sequencing artifact metrics. Output is in a graphical form that can be compared across samples/runs. Percentage of duplicate reads was also assessed using Picard’s Mark Duplicates tool (<https://broadinstitute.github.io/picard/command-line-overview.html#MarkDuplicates>).

2.2.14.6 Autozygosity Mapping from WES Data

Where SNP microarray analysis was not available and the family history suggested consanguinity, autozygosity mapping was performed from WES data. This was performed using VCF Hacks script ‘VCFtoSnpViewer’ which selects common SNPs from the data, identifies them as homozygous or heterozygous calls and converts them into files compatible with SnpViewer v0.9.2

(<https://sourceforge.net/projects/snpviewer/>). Samples could then be viewed as described previously, in order to identify large regions of homozygosity (see section 2.2.13).

2.2.14.7 Copy Number Variation from WES Data

In the majority of families, Array CGH was performed as part of Clinical Genetics input prior to referral to the study. The result of this test was assessed for all consented patients. For families with and without an array CGH result, further analysis for CNVs was performed using WES data.

The 'Exome Depth' program was utilized to perform this function (Plagnol et al., 2012). This program compares read depth between the test sample and 5-10 control samples. Control samples were selected to be as close as possible to the test sample, for example those run on the same sequencing run as the test sample, with libraries prepared using identical WES methodology. Only samples from unrelated individuals were used as controls. The output from the ExomeDepth program was in the form of a .csv file. This was annotated to indicate common CNVs from control populations (Conrad et al., 2010). CNV calls were then ranked by the Bayes Factor, defined as the log₁₀ likelihood ratio of data for the CNV call divided by normal copy number. CNV calls were manually inspected to assess read ratio and involved genes. These genes were cross-referenced against the cardiac literature, cilia-related gene lists and exome call data for each sample. For example, if a putative heterozygous deletion of a certain gene was identified in an affected individual, this gene was checked for the presence of a pathogenic SNP in WES data. A graphical representation of the results was created using Broad Institute's Integrative Genomics Viewer (IGV) (Robinson et al., 2011).

2.2.15 Whole Genome Sequencing at Low Read Depth

WGS at low read depth was utilized in some cases in order to identify potential pathogenic CNVs.

DNA samples were quantified by Qubit BR dsDNA assay (see section 2.2.14) and were sent to the Leeds Next Generation Sequencing Facility who prepared the libraries for sequencing using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA). Library preparation was carried out according to the manufacturer's directions. DNA shearing was performed using a Covaris S2 Adaptive Focused Acoustics sonicator to create fragments of a peak size of 185bp.

PCR amplification was carried out in a 12 cycle PCR. Ten libraries were pooled per lane and sequencing performed on the Illumina HiSeq2500 platform (Illumina) in 50bp single end read rapid mode.

Data analysis was performed with support from Dr David Parry (Post Doctoral Research Fellow, Leeds Institute of Molecular Medicine). Alignment to the GRCh37 reference sequence was performed using the Burrow-Wheeler Aligner (BWA) (Li and Durbin, 2009). Duplicate reads were removed using Picard (Broad Institute; <http://picard.sourceforge.net>).

Three different analysis programs were used: CNVnator (<http://sv.gersteinlab.org/>), readdepth (<http://code.google.com/p/readdepth/>) and CNV-seq (<http://tiger.dbs.nus.edu.sg/cnv-seq/>). Each was utilized according to the product protocol. CNVnator and readdepth infer CNVs from average read depth across a single sample, while CNV-seq compares read depths from a test sample against a selected reference sample. For CNV-seq, CNV calls were calculated for all possible pairwise combinations of samples available. Plots were created for each chromosome with a CNV call. The output files from each of the three programs were compared. All called CNVs were analyzed in detail and particular attention was paid to those called by more than one program. These calls were also compared to databases of known population CNVs, as well as CNV calls from exome data.

2.2.16 Whole Genome Sequencing and Breakpoint Analysis

WGS was used in one instance to correctly map a deletion identified in WES data. This analysis was performed by Dr Chris Watson, Leeds Sequencing Facility, Leeds Teaching Hospitals NHS Trust, including the described Sanger sequencing. WGS in this instance was performed with 1 μ g of DNA, which was sheared using a Covaris S2 machine (Covaris Inc, Woburn, MA, USA). A WGS library was created using the NEBNext[®] Ultra[™] reagents, following the manufacturer's protocols (New England Biolabs, Ipswich, MA, USA). The library insert size was 300-400bp and 6 cycles of enrichment PCR were used. The Agilent Bioanalyser was used to check the quality of the library, as previously described. Sequencing was performed on an Illumina HiSeq3000 DNA sequencer (Illumina, San Diego, CA, USA), which generated paired-end 151bp reads. From .fastq files, adapter sequences and low quality bases (QC<10) were trimmed using Cutadapt v.1.9.1 (<https://github.com/marcelm/cutadapt>). Alignment to the reference sequence was with BWA MEM v.0.7.13 (<http://bio-bwa.sourceforge.net>). Picard v.2.1.1 was used to mark duplicates and aligned

sequence reads were visualized with IGV Integrative Genome Viewer v.2.3.80 (<http://software.broadinstitute.org/software/igv/>). Reads aligning to the locus of interest were interrogated using samtools v.0.1.18 (Li et al., 2009) to identify potential breakpoint spanning reads and a BLAT search ([http:// genome.ucsc.edu](http://genome.ucsc.edu)) was used to interrogate the sequence of these.

Sanger sequencing was used to identify the exact breakpoint (see section 2.2.12). A PCR amplicon was optimized to cover the identified breakpoint using a forward and reverse primer (for primer sequences see appendix E) to generate a 377bp product. A further reverse primer was used to give a larger (551bp) product, specific to the normal allele.

2.2.17 Microbiology

Microbiology methods were performed in a dedicated microbiology lab. Luria Bertani (LB) broth was made up with 20g LB powder (Sigma-Aldrich) in 1 litre of dH₂O and autoclaved in a bench top autoclave (Prestige Medical, Coventry UK). Agar was made up from 10g of agar powder (Merck) plus 10g LB powder (Sigma-Aldrich) in 500ml dH₂O. This was autoclaved in a bench top autoclave (Prestige Medical, Coventry UK) and cooled to around 50°C. An appropriate concentration of antibiotic was added (see Table 2-3) and 25ml of agar was plated per 100 x 15mm plate, left to cool and stored at 4°C. SOC medium (Super Optimal broth with Catabolite repression) was a commercial product obtained from New England Biolabs. For bacterial colony growth, plates were incubated at 37°C in a dedicated incubator.

Table 2-3: Antibiotics used in microbiology methods, concentration used and supplier.

Antibiotic	Concentration	Supplier
Ampicillin	100mg/ml	Melforth labs (Suffolk, UK)
Kanamycin	100mg/ml	Sigma-Aldrich

2.2.17.1 Gateway Cloning

Gateway technology is a method of cloning which takes advantage of the specific recombination properties of bacteriophage lambda I. It is fast and highly efficient. The method of gateway cloning is shown in figure 2-4.

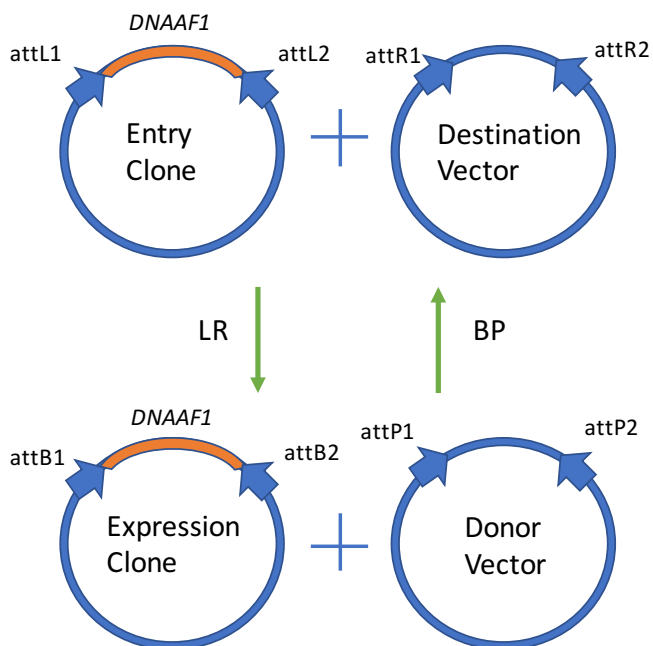


Figure 2-4: Gateway cloning methods.

Gateway cloning is a method of transfer of DNA fragments between plasmids, making use of the site-specific recombination system used by Phage λ , and using recombination sites labeled 'att' and specific recombination enzymes. The 'LR' reaction allows the gene of interest within an entry clone to be combined with the destination vector to create the desired expression clone. This process is reversible, using the 'BP' reaction to convert an expression clone of interest plus donor vector into an entry clone.

BP Clonase II reaction (Thermo Fisher Scientific) was performed as per manufacturer's protocol. 150ng of expression clone was combined with 150ng of entry clone vector. This was made up to 9 μ l with dH₂O. The BP Clonase II Enzyme Mix was thawed on ice for 2 minutes and vortexed before addition of 1 μ l to the reaction. Following a further two vortexes, the reaction was incubated at 25°C for one hour in a water bath. 1 μ l of Proteinase K was added to terminate the reaction. Following a brief vortex, samples were incubated at 37°C for 10mins. Samples were then transformed in *E. coli* DH5-alpha competent cells (New England Biolabs) (see section 2.2.17.2) and DNA extracted from bacterial cultures by mini and maxi preps (see section 2.2.17.4 and 2.2.17.5).

The LR Clonase gateway reaction (Thermo Fisher Scientific) was performed to subclone cDNA into destination vectors pCS2+ and GW331 (N-terminal SF-TAP). 150ng of entry clone was combined with 150ng of destination vector, 2 μ l 5xLR Clonase Reaction Buffer and 4 μ l TE buffer. LR Clonase enzyme mix was thawed on ice for 2 minutes and vortexed briefly twice. 1 μ l of enzyme mix was added to each reaction and vortexed briefly twice. The reaction was incubated at 25°C for 60mins. 2 μ g of Proteinase K was added to terminate the reaction and incubated at 37°C for 10mins.

The resulting plasmid was transformed into competent cells as described above (see section 2.2.17.2 for details).

All cDNA clone sequences were confirmed by Sanger sequencing. For primer sequences see appendix E.

2.2.17.2 Bacterial Transformation

Bacterial transformation was performed in a dedicated microbiology room. 1µl of the desired plasmid was added to an aliquot of *E. coli* DH5-alpha competent cells (New England Biolabs), freshly thawed at 4°C and maintained at this temperature. This mixture was flicked briefly to combine and left on ice for 20mins. A heat shock at 45°C in a water bath for 45s was performed. The reaction was then immediately transferred into ice for 5 min. 500µl of SOC medium (New England Biolabs), which was prewarmed to RT, was added to the tube. This was left in the orbital shaker (Excella E25, Eppendorf, Hamburg, Germany) at 300 rpm at 37°C for at least 1 hour. 150µl of the bacterial sample was then spread on an LB agar plate with the appropriate antibiotic (Table 2-3). The plate was incubated for no more than 16hr (usually overnight) at 37°C, to reduce the likelihood of satellite colonies growing.

The following morning, individual colonies were selected from the plate, using a loop, for growth in a 5ml culture. The 5ml culture was made up in a universal container, with 5ml of LB and the appropriate antibiotic (Table 2-3). Each 5 ml culture was grown in an orbital shaker (Excella E25, Eppendorf, Hamburg, Germany) at 300 rpm at 37°C overnight, to ensure sufficient aeration of the culture. DNA was extracted from these samples by mini-prep (Qiagen) (see 2.2.19) and Sanger sequencing was performed. For samples with the desired sequence, bacterial colonies were further grown in a 200ml LB culture, in microbiology flasks with baffles to ensure adequate aeration of the culture. Maxi-prep (Qiagen) for DNA extraction was then performed (see 2.2.20).

2.2.17.3 Site-Directed Mutagenesis

Site-directed mutagenesis (SDM) was performed using the Q5 site-directed mutagenesis kit protocol (New England Biolabs), following the manufacturer's directions. Primers were designed on the NEBaseChanger website (<http://nebasechanger.neb.com>). A reaction consisted of 6.25µl Hot Start High-Fidelity Master mix (2x), 1.25µl Forward and Reverse primer (10µM), 0.5µl plasmid DNA (25ng/µl), and 4.5µl dH₂O. The reaction was vortexed and transferred to a thermocycler to perform a PCR with the following conditions: 98°C for 30s, 25x cycles of 98°C for 10s, 'X'°C for 30s and 72°C for 'Y's, then 72°C for 2mins and hold at 4°C. Here 'X' is

the appropriate temperature for primer annealing, calculated using NEBaseChanger, and 'Y' is 30s/kb of plasmid DNA.

Following the PCR the 'KLD' step was performed, which stands for Kinase, Ligase and DpnI. For this reaction, 0.5µl of PCR product was combined with 2.5µl 2 x KLD reaction buffer, 0.5µl 10x KLD enzyme mix and 1.5µl dH₂O. This was pipette-mixed and incubated at RT for 5mins. Finally, a transformation was performed (see section 2.2.17.2).

2.2.17.4 Mini Preps of Plasmid DNA

Bacterial colonies were grown in a 5ml LB culture at 37°C overnight in an orbital shaker. Samples were pelleted by centrifugation at 10,000xg for 3 mins. Mini-prep (Qiagen) was performed according to the manufacturer's protocol. Pelleted bacterial cells were resuspended in 250µl buffer P1. 250µl of buffer P2 was added and the reaction was mixed thoroughly by inverting the tube 4-6 times. Then, 350µl of Buffer N3 was added and mixed by inverting the tube 4-6 times. The sample was centrifuged for 10 mins at 17,900xg. The supernatants were added to a QIA prep spin column and centrifuged for 60s. The spin column was washed with 0.5ml buffer PB and centrifuged for 60s, then further washed with 0.75ml buffer PE and centrifuged for 60s. To elute the DNA, 50µl buffer EB was added to the spin column, left to stand for 1 min and centrifuged for 1 min. DNA concentration was measured on a Nanodrop instrument (Thermo Fisher Scientific) prior to Sanger sequencing.

2.2.17.5 Maxi Preps of Plasmid DNA

For large-scale plasmid growth and DNA extraction, Maxi-preps (Qiagen) were utilized. These were performed following a 200ml LB broth growth at 37°C for 12-16 hours. The kit protocol was followed. The bacterial culture was harvested by centrifuging at 4000xg for 25 mins at 4°C. The pellet was completely resuspended in buffer P1. 10mls of buffer P2 was added and the tube was mixed by inverting 4-6 times. The reaction was incubated at RT for 5min. Following this, 10ml of buffer P3 was added and mixed by inverting 4-6 times. The lysate was poured into a QIAfilter Cartridge and incubated at RT for 10min. 10ml of Buffer QBT was used to equilibrate the Qiagen-tip and allowed to drip through by gravity flow. The cell lysate was discharged through the QIAfilter Cartridge into the Qiagen-tip and allowed to enter the resin by gravity flow. The tip was washed with two washes of 30ml of buffer QC. DNA was eluted with 15ml of buffer QF. DNA was precipitated with 10.5ml RT isopropanol and centrifuged at 15,000xg for 30 mins at 4°C. The supernatant was decanted and 5ml RT 70% ethanol was added and centrifuged at 15,000xg for 10 min. The supernatant was removed and

the DNA pellet was air-dried for 10 min under foil. The DNA pellet was redissolved in 500µl of sterile TE buffer. DNA concentration was measured on a Nanodrop instrument (Thermo Fisher Scientific).

2.2.18 Restriction Enzyme Digest

Restriction enzyme mapping was performed to confirm the correct organization of the plasmid before Sanger sequencing. A restriction enzyme was selected with appropriate sites within the construct using the SNAP gene viewer (<http://www.snapgene.com>). Digestion was performed using an appropriate concentration of enzyme and buffer and was incubated on a thermocycler as per manufacturer's directions. The resulting fragments were run on 2% agarose gel stained with ethidium bromide (see 2.2.9.3).

2.2.19 Cell Culture

hTERT-RPE1, mIMCD3 and HEK293 cells were grown in culture according to standard aseptic techniques. All cell culture was performed in a dedicated Tissue Culture room. Media was DMEM/F12 + GlutaMAX™ (Life Technologies, Thermo Fisher Scientific) supplemented with 10% Fetal Calf Serum (FCS) (Life Technologies) with cells grown at 37°C/5% CO₂ in a humidified tissue culture incubator. Patient fibroblasts were grown in DMEM/F12 + GlutaMAX™ + 10% FCS in an incubator dedicated for primary cell culture.

2.2.20 Cell Passage and Harvesting

Cell-lines were grown in T75 flasks. When cultures reached near confluency they were passaged. Initially, growth media was removed and cells were washed in 5mls 1xPBS. 2ml 1x trypsin/EDTA was then added and the cells left for five minutes to lift into suspension. 8ml of growth media was then added and the cells were spun down at 200xg for 5mins in a 15ml Falcon tube. The cell pellet was resuspended in 10ml media. An appropriate volume of cells was then transferred to a new flask with warmed media. All cell-lines were split according to the supplier's recommendations: RPE1 cells 1:10 twice weekly, HEK293 cells 1:10 weekly, IMCD3 1:8 twice a week. When plating for further experiments, the cell number was determined using a Trypan Blue stain (Invitrogen) and Countess® cell counter (Invitrogen) and an appropriate number of cells were plated per well. All cell lines were passaged to a maximum number of 35 passages before discard.

2.2.21 Transfection

2.2.21.1 Over-expression

Cell cultures were grown in 6 well plates to 70-80% confluency. If immunofluorescence was to be performed, glass cover slips treated with ethanol and acetone were used to line the base of the plate. For hTERT-RPE1 and HEK293 cells, 80% confluency was achieved by plating 1×10^5 cells per well, 48 hours prior to transfection. 1 μ g of plasmid DNA was added to Opti-MEMTM (Thermo Fisher Scientific) to a total volume of 250 μ l. If two different plasmids were to be transfected, 0.5 μ g of each was used. 6 μ l of Lipofectamine[®]2000 (Life Technologies) was separately combined with 246 μ l of Opti-MEMTM. This was incubated for 5 minutes. The Lipofectamine[®]2000/Opti-MEMTM mix was added to the DNA/Opti-MEMTM and incubated for 20 minutes. Cell growing media was exchanged for Opti-MEMTM and the transfection complexes were added. These were left for 4-6 hours at 37°C. Finally, Opti-MEMTM was exchanged for normal growing media (DMEM F12 plus 10% FCS) or starvation media (DMEM F12 plus 0.2% FCS) if required, to stimulate cilia formation. Assays were performed after 48-72 hours.

If required, for example in the case of co-immunoprecipitation, experiments were scaled-up and performed in a T75 flask. Volumes were scaled accordingly; so 1.5ml of Opti-MEMTM was combined with 5 μ g DNA and then 1.5ml Opti-MEMTM and 45 μ l Lipofectamine[®]2000 were combined and incubated for 5 min. These two pools were combined for 20 minutes. This gave a total of 3ml of complexes to transfect one T75 flask of cells, where normal growing media was exchanged for 9mls of Opti-MEMTM and complexes were added. Otherwise the protocol was followed as for a 6-well plate.

2.2.21.2 siRNA Knockdown

siRNA knockdown was performed alongside construct transfection and over-expression studies, so Lipofectamine[®]2000 (Life Technologies) was used for both simultaneously. Commercial siRNAs were On-TargetPlus siRNA Smartpool oligos (Dharmacon) used at a final concentration of 50nM.

hTERT-RPE1 cells were grown in 6 well plates to 70-80% confluency. 1 μ g of plasmid DNA for overexpression and 50nM siRNA was added to Opti-MEMTM (Thermo Fisher Scientific) to a total volume of 250 μ l. 6 μ l of Lipofectamine[®]2000 (Life Technologies) was separately combined with 246 μ l of Opti-MEMTM. This was incubated for 5 minutes. The Lipofectamine[®]2000/Opti-MEMTM mix was added to the siRNA/Opti-MEMTM pool and incubated for 20 minutes. Cell growing media was exchanged for Opti-MEMTM and

the complexes were added. These were left for 4-6 hours at 37°C. Finally, Opti-MEM™ was exchanged for normal growing media (DMEM F12 plus 10% FCS) or starvation media (DMEM F12 plus 0.2% FCS) if required to stimulate cilia formation. The assays were performed after 48 hours. Scrambled siRNA (Dharmacon) was used as a control.

The efficiency of siRNA knockdown was assessed using Western blotting, using methods as described in section 2.2.23. Statistical tests to compare results were performed using SPSS version 22 (IBM) (see section 2.2.29).

2.2.22 Immunofluorescence and Confocal Microscopy

At 48-72 hours following transfection or at 80% confluency for non-transfected cells, fixation to coverslips was performed. Fixation was either with ice-cold methanol for 5min at -20°C or with 2% [w/v] para-formaldehyde (PFA), where 2% PFA was added for 20 minutes at RT, then permeabilised with 0.01% [v/v] Triton X-100 (Sigma-Aldrich) for 5 min. Coverslips were blocked in 1% [w/v] Marvel milk solution for at least 5 minutes (occasionally overnight at 4°C). Marvel milk solution was clarified by centrifugation immediately before use to remove any particulates. Primary antibodies (see Table 2-1) were made up in 1% Marvel milk solution and incubated with coverslips in a humid chamber for a time period dependent on the primary antibody, usually 1 hour. Coverslips were then washed three times in 1x PBS. Incubation with the appropriate AlexaFluor® conjugated secondary antibodies (Life Technologies), plus DAPI for nuclear staining, was in 1% Marvel milk solution was carried out for 1 hour. Five 1x PBS washes were then performed, followed by 1x dH₂O wash. Coverslips were set to slides in 30µl Mowiol® (Sigma-Aldrich).

Confocal imaging was carried out using a Nikon A1R confocal microscope, processed by NIS-Elements Confocal 4.5 (Nikon) software. Image J (<https://imagej.nih.gov/ij/>) or Fiji (<https://fiji.sc>) was used for post-capture image processing. Co-localisation was assessed using Fiji's Coloc2 tool, following best practice guidelines, or was assessed manually by eye by a scientist blinded to the experimental conditions of each slide. Corrected Total Cell Fluorescence (CTCF) was calculated by designating a region of interest and using the following formula: $CTCF = \text{Integrated Density} - (\text{Area of selected cell} \times \text{Mean fluorescence of background readings})$.

2.2.23 Western Blotting

At 48-72hr post-transfection, cells were collected for preparation of Whole Cell Extract (WCE). All procedures were done at 4°C. Cells were washed once with ice-cold

1xPBS. A further wash of 1xPBS was added and cells were scraped into the PBS using a cell scraper. Cells were centrifuged at 200xg for 15mins at 4°C to create a cell pellet.

Cell lysis was performed by addition of NP40 cell lysis buffer (see 2.1.2.5), at a volume of 100µl for a 6-well plate or 500µl for a T75 flask. The samples were incubated on ice for 30 minutes. The lysate was centrifuged at 14000xg for 15 minutes. Each sample was sonicated using a Sanyo Soniprep 150 Sonicator for 10 secs. Protein concentration was determined using a Bradford Assay (Bio-rad, Life Science, Berkley, California, USA), following the manufacturer's guidelines. Absorbance was measured on a spectrophotometer at 690 nm and compared to a range of BSA standards (Sigma).

20µg of protein was used per sample. Samples were mixed with 2xSDS loading buffer and 2.5% of 2-mercaptoethanol was added to each sample. Samples were then heated to 95°C for 5 minutes on a hot block. Lysates were electrophoresed in NuPAGE 4%–12% Bis-Tris gel (Life Technologies), alongside All Blue Standards Precision Plus Protein Ladder (Bio-rad) for 1.5hours at 120V. NuPAGE® MES-SDS Running Buffer (Thermo Fisher Scientific) was used. Transfer to PVDF membranes (Life Technologies) was performed at 30V for 1.5 hours. NuPAGE® Transfer Buffer (Thermo Fisher Scientific) was used in the transfer. Membranes were blocked in 5% Marvel milk in 1xPBST (see section 2.1.2.1) for 2 hours. Membranes were washed four times in 1xPBST (10 minute incubations) and incubation in the primary antibody, made up in 5%[w/v] Marvel milk, clarified by centrifugation, was performed for 30 minutes (anti-FLAG antibody) or 1 hour (all other antibodies) at 4°C in a Falcon tube on an electric roller. Membranes were washed four times with 1xPBST (10 minute incubations) and incubation with the appropriate HRP-tagged secondary antibody (Dako, Agilent Technologies) at a concentration of 1:10000 in 5% Marvel milk was performed for 1 hour at 4°C. Membranes were developed using SuperSignal West Femto kit (Thermo Fisher Scientific), following a final four 1xPBST washes. Images were acquired using a Bio-Rad molecular image ChemiDoc™ MP Imaging System with a UV transilluminator and displayed and analyzed using Image Lab (v. 4.0) analysis software (Bio-Rad, Hemel Hempstead, UK). Band intensity was calculated using the Image Lab software and compared to a beta-actin loading control for quantification.

For membrane stripping, the membrane was washed in 1xPBST and then 5ml of stripping buffer (Restore™ Plus, Thermo Fisher Scientific) was used to wash the

membrane for 10min. A further two 1xPBST washes were performed prior to further staining.

2.2.24 Co-Immunoprecipitation

Whole Cell Extract (WCE) was prepared by following the protocol for western blotting (see section 2.2.23). Protein A agarose beads (Sigma) were prepared with a total of 3 washes. For each wash a 50% solution [v/v] with incubation buffer (see 2.1.2.7) was created and then spun at 1000xg at 4°C for 1 minute. The beads were finally made up to a 50% [v/v] solution in incubation buffer.

500-1000µg of soluble protein was made up to 1000µl in incubation buffer. The samples were pre-cleared with end-to-end mixing with 30µl of washed Protein A agarose beads for 30min at 4°C on a rotating wheel. The beads were extracted by centrifugation at 1000xg at 4°C for 1min. The supernatant was incubated with 1µg of the appropriate antibody overnight at 4°C on an orbital shaker. The capture of immunocomplexes was performed by adding 70µl of washed Protein A beads to each sample and incubating for 3 hours at 4°C. The beads were centrifuged at 1000xg for 1 minute to collect and then washed 3 times with 500µl ice-cold IP wash buffer (see 2.1.2.8) with a spin at 1000xg for 1 minute between each wash. Protein elution was performed by the addition of 20µl 2%SDS and incubation for 30 minutes at room temperature. The sample was then centrifuged at 14,000xg for 2 min and the supernatant collected in a new tube.

2.2.25 Tandem Affinity Purification

This experiment was performed by Dr Kasia Szymanska, Post Doctoral Research Fellow, Leeds Institute of Molecular Medicine. In brief, at 48 hours following transfection, cells were collected in 1xPBS and lysed in NP40 lysis buffer (see section 2.1.2.6), supplemented with protease inhibitors (Thermo Fisher Scientific). The cell lysates were collected by centrifugation for 5min at 200xg, at a temperature of 4°C. Approximately 2mg of lysate was incubated with 100µl of washed streptactin "Superflow" beads (IBA, Goettingen, Germany). End-to-end mixing was performed for 1.5 hours at 4°C. The sample was centrifuged at 5000xg to pellet the beads and the beads were transferred to a new microspin column. Three washes with 500µl of TAP wash buffer (see section 2.1.2.11) were performed with centrifugation for 10s at 1000xg following each wash. 200µl ice-cold D-desthiobiotin elution buffer (see section 2.1.2.12) was used to elute bound proteins with a 15 min incubation at 4°C. The reaction was transferred to a new microspin column and centrifuged at 1000xg for 10s.

The elute was incubated with 50µl of washed anti-FLAG M2 affinity gel beads (Sigma Aldrich) for 1.5 hours on a rotating wheel. A 1000xg spin was used to separate the beads. The beads were washed three times with 500µl of ice-cold wash buffer and once with 500µl 1xPBS. Proteins were eluted with 200µl of ice-cold FLAG peptide elution buffer (200 µg/ml FLAG peptide (Sigma Aldrich) in 1xPBS) for 15 minutes at 4°C. The samples were centrifuged at 1000xg and the elute transferred to a fresh tube.

Precipitated proteins were sent for mass spectrometry at a collaborating centre (Centre for Ophthalmic Research, Tubingen, Germany). Analysis of mass spectrometry data was performed using DAVID software (<https://david.ncifcrf.gov>) and displayed in an Excel spreadsheet. Results of the three biological replicates were pooled together. The pulled down peptides were matched to corresponding proteins. Those hits that were observed in only one replicate, as well as those that were an effect of contamination (present in BSA washes) were removed. The results were ordered to show the number of pulled down peptides per protein.

2.2.26 Mouse Embryo Dissection

Mouse embryo dissection and staining was performed in close collaboration with Dina Abdelmottaleb, Visting Researcher, Leeds Institute of Molecular Medicine, Leeds, UK. Black six mouse embryos at embryonic stages E7.0-E8.5 were euthanized and a small midline lateral incision was made with surgical scissors. The two uterine horns were identified and dissected into 1xPBS + 1% BSA (Sigma-Aldrich). Under a dissecting microscope, each embryonic sack was isolated by cutting between the implantation sites along the uterine horn. The surrounding tissues were removed from the embryonic sack and it was dissected longitudinally. As it was opened the embryo was identified and collected. Intact embryos were transferred to a 12-well plate containing 1xPBS + 1% BSA. Further careful dissection was performed to remove the Reichart's membrane from each embryo and intact embryos were taken forward for fixation and staining.

2.2.27 Mouse Embryo Whole Mount Immunofluorescence

Fixation of dissected mouse embryos was performed in 4% *para*-formaldehyde for 1 hour at RT. Embryos were then rehydrated in a series of methanol dilutions in PBS for 15 minutes at each decreasing concentration; 100% for 15 mins, 75% for 15 mins, 50% for 15 mins, 25% for 15 mins. Embryos were washed 3 times in PBS for 5 mins each. Embryos were blocked in 1% BSA for 1 hour on the rocker. The primary antibody was diluted in 1% BSA and incubated with the sample overnight at 4°C. Three 1xPBS

washes were then performed. Incubation with secondary antibody and DAPI, at 1:1000 concentration, was in 1% BSA for one hour. Finally, samples underwent three 1xPBS washes. Samples were imaged using either light-sheet microscopy (see section 2.2.29) or confocal microscopy, where the sample was mounted in Mowiol (Sigma-Aldrich) and imaged using a Nikon A1R confocal microscope, processed by NIS-Elements Confocal 4.5 (Nikon) software.

2.2.28 Light Sheet Microscopy

A Zeiss Z1 light sheet fluorescence microscope, based at the University of Sheffield, was used for imaging mouse embryo specimens. This technology illuminates a thin section of tissue at a time, detected by a perpendicular digital camera and so generates an optical section of the sample. This is a good method for imaging within an organism and photobleaching is reduced compared to confocal microscopy (Keller et al., 2010). It also allows imaging to be performed from multiple viewpoints, allowing the anatomy to be visualised during imaging and a 360° image to be created. Each embryo was suspended in 1% low melting point agarose gel in a 1.5ml glass capillary before loading into the microscope for imaging. Imaging was performed as per best practice guidelines (<http://applications.zeiss.com>). Post-image processing was performed using Image J software.

2.2.29 Statistical Tests

Statistical tests were performed using SPSS version 22 (IBM). Normal distribution was tested using the Shapiro-Wilk test and by plotting a normal Q-Q Plot. Levene's test was used to test equality of covariance. Results are depicted in the form of box plots or bar charts, and error bars show the Standard Error of the Mean. Statistically significant differences between the means were calculated using the Independent Samples T test (two groups, parametric), Mann-Whitney-U test (two groups, non-parametric) or one-way ANOVA (three or more groups). Significance is described as $P < 0.05$.

In this study 'biological replicates' describe a repeat of the entire experiment, using different cell samples, tissues and plates. 'Technical replicates' would describe different fields of view from the same sample well or plate (Blainey et al., 2014).

3 Whole Exome Sequencing can Identify Known and Candidate Genes in Families with Congenital Heart Disease

3.1 Introduction

In this chapter, the overall results of the library preparation for Whole Exome Sequencing (WES) are presented. Further to this, the results of WES data analysis from 11 families recruited to the study are described. Five further recruited families are discussed in detail in chapters 4, 5 and 6.

The results of clinical phenotyping and variant filtering are presented according to the following categories: families with mutations in known genes; families affected with Congenital Heart Disease phenotypes; and finally patients with laterality disorders recruited to the Deciphering Developmental Delay (DDD) project. The methods of variant filtering and a discussion of the candidate variants for these patient groups are summarized at the end of each sub-section.

The results show that the achieved sequencing depth and coverage were consistent across the sequenced samples, and that the different preparation methods gave results consistent with the published literature. WES is shown to be a successful method for identification of mutations in known disease genes and potential novel candidates in CHD. In the sub-section of families described with independent CHD phenotypes, in which 10 families are assessed, no clear functional candidates were identified, but a number of possible disease-causing variants were discovered and their potential pathogenicity is discussed.

3.2 Overall Results

3.2.1 Patient Recruitment and Phenotyping

In total, 16 families were recruited to the CHD study. Family members gave informed consent for research studies and had detailed phenotyping undertaken, followed by DNA extraction from blood or saliva samples. In total, 57 DNA samples were collected.

3.2.2 Whole Exome Sequencing

53 DNA samples from 16 families were prepared for WES. For three samples (JT583, JT785 and JT619), WES library preparation was attempted but was unsuccessful. All of these samples had low concentration and quality of DNA. Following initial shearing, Agilent Bioanalyzer traces were poor with either severe under-fragmentation or low DNA concentration and so the analyses were discontinued (see Figure 3-1). One sample (JT475) was not taken forward for WES library preparation, as this sample was from an unaffected sibling and was not deemed necessary for the analysis.

3.2.2.1 Assessment of WES Library Preparation

WES library preparation was undertaken with either the Illumina XT or QXT methods (see Chapter 2, section 2.2.14). For XT preparation, library quality was checked at three stages during the preparation, comprising the post-shearing, post-amplification and post-hybridization steps. Figure 3-1:B-D shows an example of Bioanalyzer traces at the three stages of analysis from one sample (JT380), as an example, which was prepared using the XT method. DNA concentration was calculated from these traces by measuring the area under the curve.

For QXT library preparation, Bioanalyzer traces were recorded at two stages: post-shearing and post-hybridization. Figure 3-2 shows an example of results from sample JT584, which was prepared using the QXT kit. Occasionally, when using the QXT kit, under-shearing of fragments occurred. If the DNA quantity at the appropriate fragment length (around 200bp) was acceptable, then samples were still taken forward to hybridization and post-hybridization traces always normalized (see example fig. 3-3, sample JT624). This did not appear to affect data quality in downstream applications.

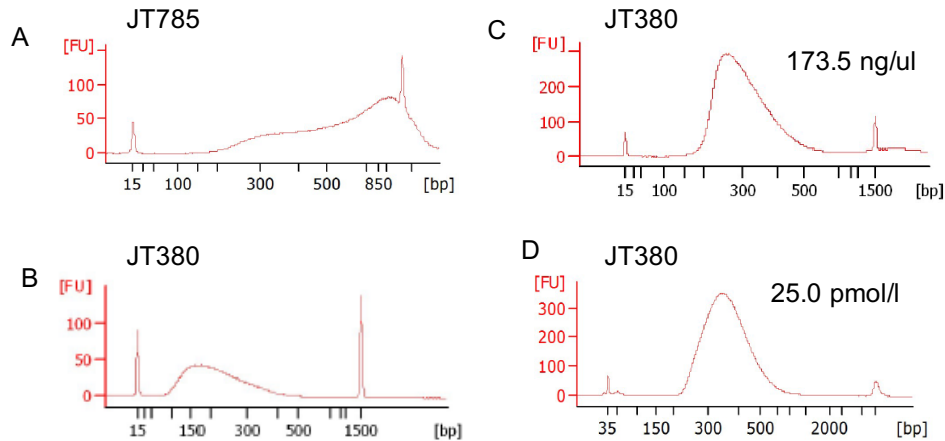


Figure 3-1: Bioanalyzer results for samples JT785 and JT380.

A: Bioanalyzer results from sample JT785 following post-shearing, discontinued due to poor shearing. B-D: Bioanalyzer results from sample JT380 prepared using the XT method showing B:post-shearing, C:post-amplification and D:post-hybridization analysis. Library concentrations are given by interrogating area under the curve.

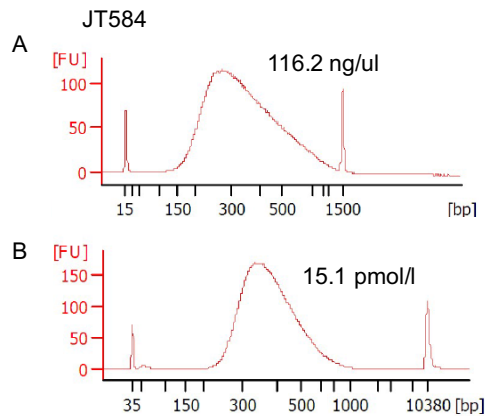


Figure 3-2: Bioanalyzer results from sample JT584 prepared using the QXT method.

A:post-shearing and B:post-hybridization analysis. Library concentrations are given by interrogating area under the curve.

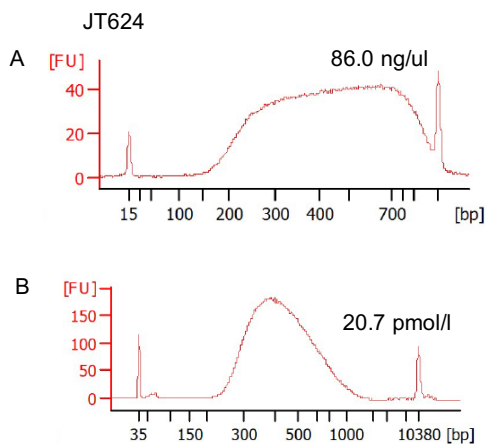


Figure 3-3: Bioanalyzer results from sample JT624 prepared using the QXT method.

A: post-shearing, showing under-fragmentation and B: post-hybridization, showing a correction of the previous under-fragmentation.

3.2.2.2 WES Data Quality Scores

3.2.2.2.1 Depth of Coverage

Data from WES was assessed for depth of coverage from all analyzed samples, using GATK, as described in Chapter 2, section 2.2.14.5. The full results of this analysis are given in appendix F. These results show the mean coverage (number of reads aligned) across the exome, the percentage of bases covered by > 5 reads, the total number of reads, mapped reads and percentage of reads mapped. The data shows a consistently high level of coverage across the exome, with a minimum mean read depth of 57 reads, and with greater than 95% of bases covered by at least 5 reads, in all samples processed. The mean read depth was further assessed to identify whether library preparation method or sequencing method affected the coverage. Figure 3-4 shows that the mean depth of coverage was not affected by the library preparation method (XT or QXT) or the input DNA method (blood, saliva or fibroblasts). However, the type of sequencing instrument did affect the depth of coverage, with an Illumina HiSeq 3000 giving a higher mean depth of coverage than an Illumina HiSeq 2500. The version of the capture library also appeared to affect depth of coverage, with version 6 giving a higher coverage than version 4 or 5. However, this increase in coverage is probably due to the fact that version 6 of the library was only sequenced on the HiSeq3000.

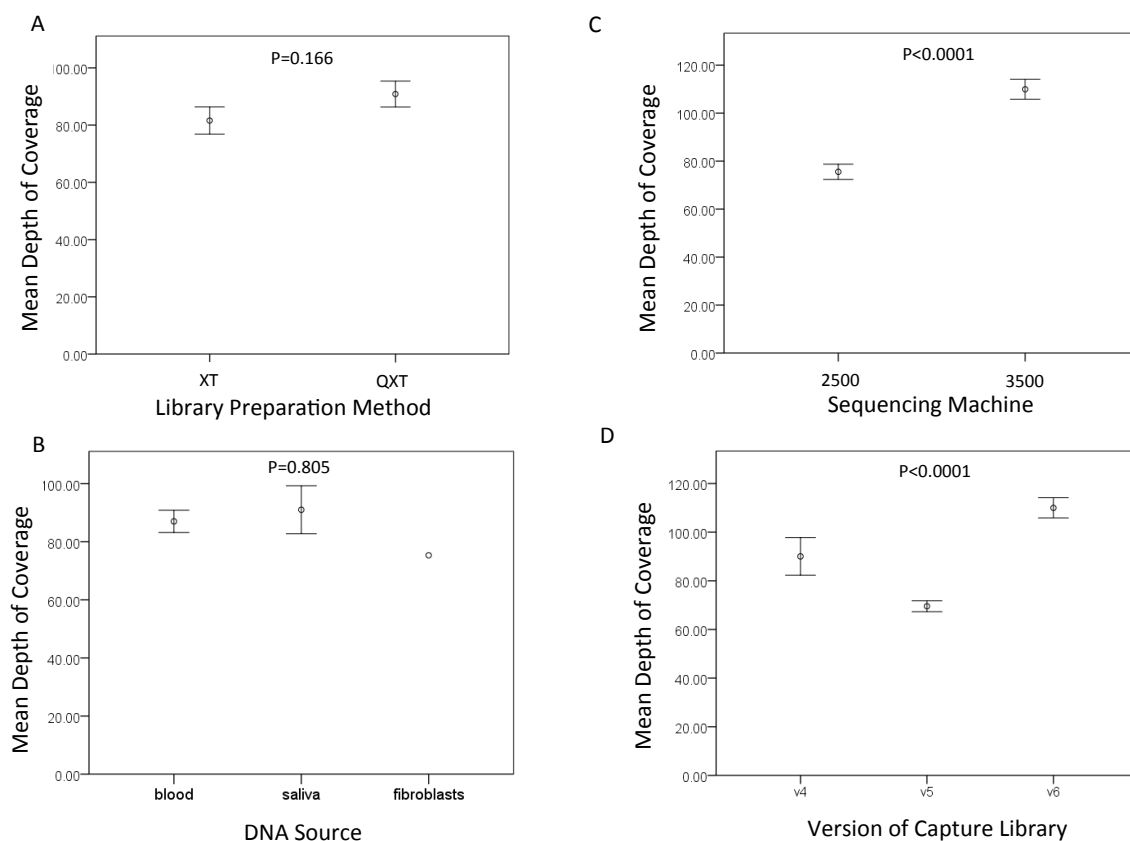


Figure 3-4: Mean depth of coverage across the exome comparing different library preparation, input and sequencing methods.

A: Mean depth of coverage across the exome comparing library preparation method, XT or QXT method. Statistical test is Mann-Whitney-U test. B: Mean depth of coverage across the exome comparing DNA source. Statistical test is one-way ANOVA. C: Mean depth of coverage across the exome comparing sequencing machine, Illumina HiSeq 2500 or 3000, statistical test is Mann-Whitney U test, $P = 0.000005$. D: Mean depth of coverage across the exome comparing the version of capture library used. Statistical test is one-way ANOVA, $P = 0.000000002$. Error bars are standard error of the mean.

The percentage of duplicate reads per sample was assessed using PICARD's "Mark Duplicates" tool. Duplicates are often remnants from the PCR processes inherent to library preparation, but can also represent optical duplicates (sequences from one cluster which are identified by the software as from multiple adjacent clusters). Figure 3-5:A shows the percentage of duplicates recognized by the software (number of duplicates/ total number of reads *100) for each sample. As is clear from the data, there is a difference in percentage duplication from sample JT538 onwards. This is when library preparation was changed from the XT method to the QXT method. Figure 3-5:B shows a comparison of the mean percentage duplication for XT versus QXT method. As can be seen from Figure 3-5:B, the QXT method gives a significantly higher percentage duplication ($p < 0.05$, Student's t-test) than the XT method. However, percentage duplication was still within acceptable limits, even when using the

QXT method, at consistently less than 35% (Shigemizu et al., 2015) and the percentage duplication did not affect the mean coverage or downstream processing.

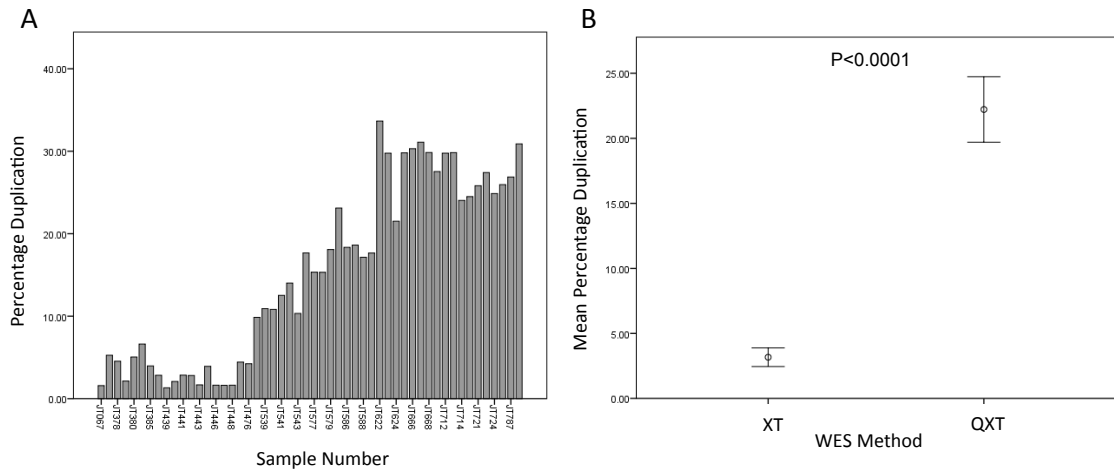


Figure 3-5: Percentage duplication according to sample number and WES method.

A: Percentage duplication per sample number. A higher percentage duplication is identified from sample JT538 and above, consistent with a move to the QXT method. B: Mean percentage duplication comparing library preparation method, XT or QXT method. Statistical test is Mann-Whitney-U test.

3.2.2.2.2 PICARD's Collect Multiple Metrics

Figure 3-6 shows an example of the summary of quality scores from sample JT380 using PICARD's "Collect Multiple Metrics" tool. The output includes 'base distribution by cycle'. In this context a 'cycle' is a single sequenced base, as the sequence runs through one procedure (cycle) per base. As can be seen from figure 3-6:A, the percentage of bases of each type, A, T, C and G, is consistent across the read, with the relative proportions representing the AT:GC content of the exome. Figure 3-6:B is a histogram showing insert size distribution, demonstrating a mean insert size of around 200 bases, as expected. Figure 3-6:C shows the mean PHRED-scaled quality score by cycle, demonstrating a consistently high quality score across the read. The distribution of PHRED-scaled quality scores by observation, showing that the majority of quality scores fall between 30 and 45, is shown in Figure 3-6:D. Graphical representations of quality scores were produced for all samples prior to analysis as a quality measure. Results were consistent across all samples, with no sample failing on quality measures.

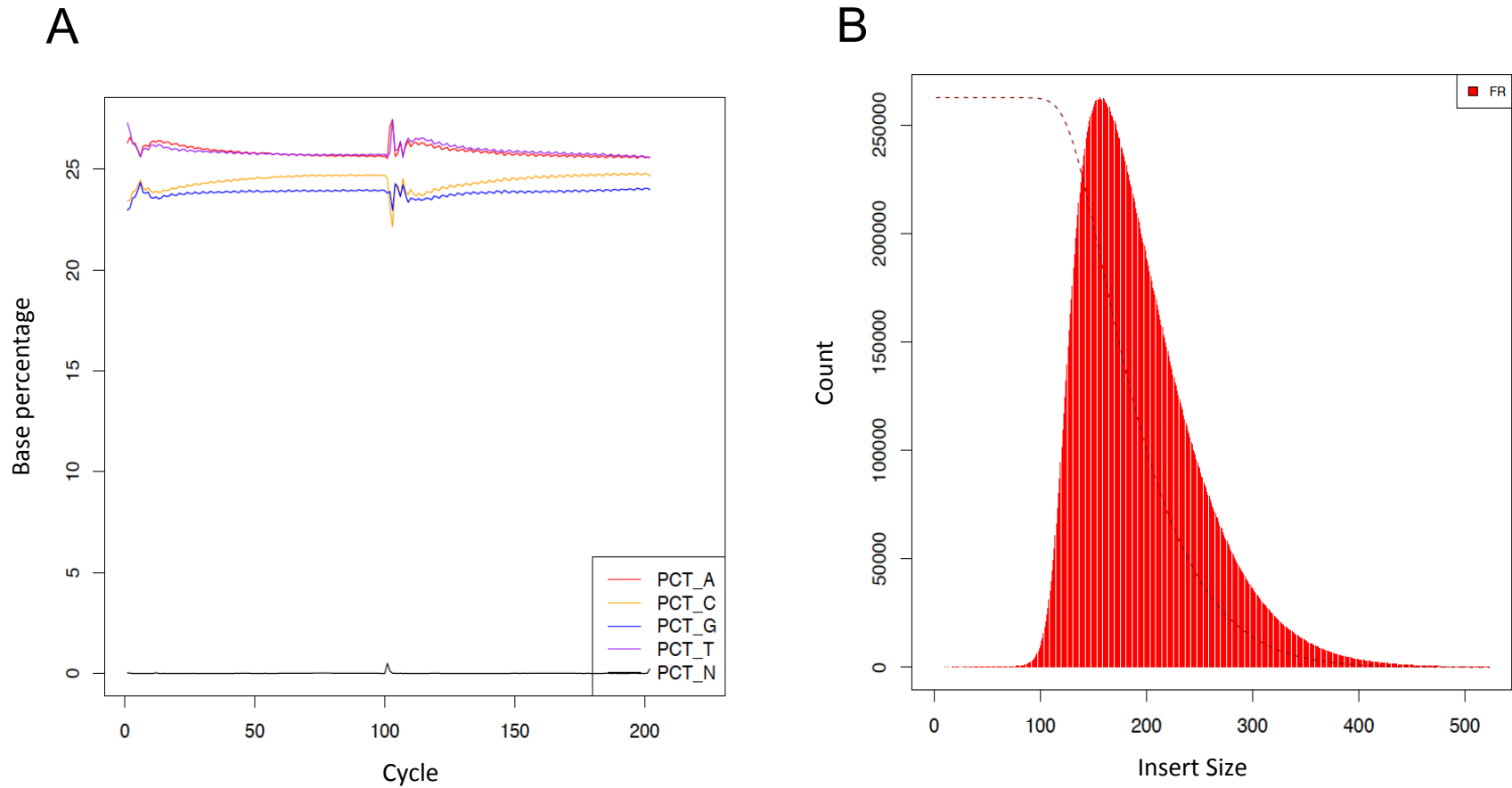


Figure 3-6: A & B: Quality scores for sample JT380 from Picard's Collect Multiple Metrics Tool.

A: Base distribution per cycle; the percentage of base of each type, A (red), T (purple), C (yellow) and G (blue), is shown across the read, with a cycle representing one base. Percentage of each base is consistent across the read and the anomaly at base 101 represents the change-over for paired-end reads. B: Insert size histogram, showing insert size distribution versus count, demonstrating a mean insert size of around 200 bases.

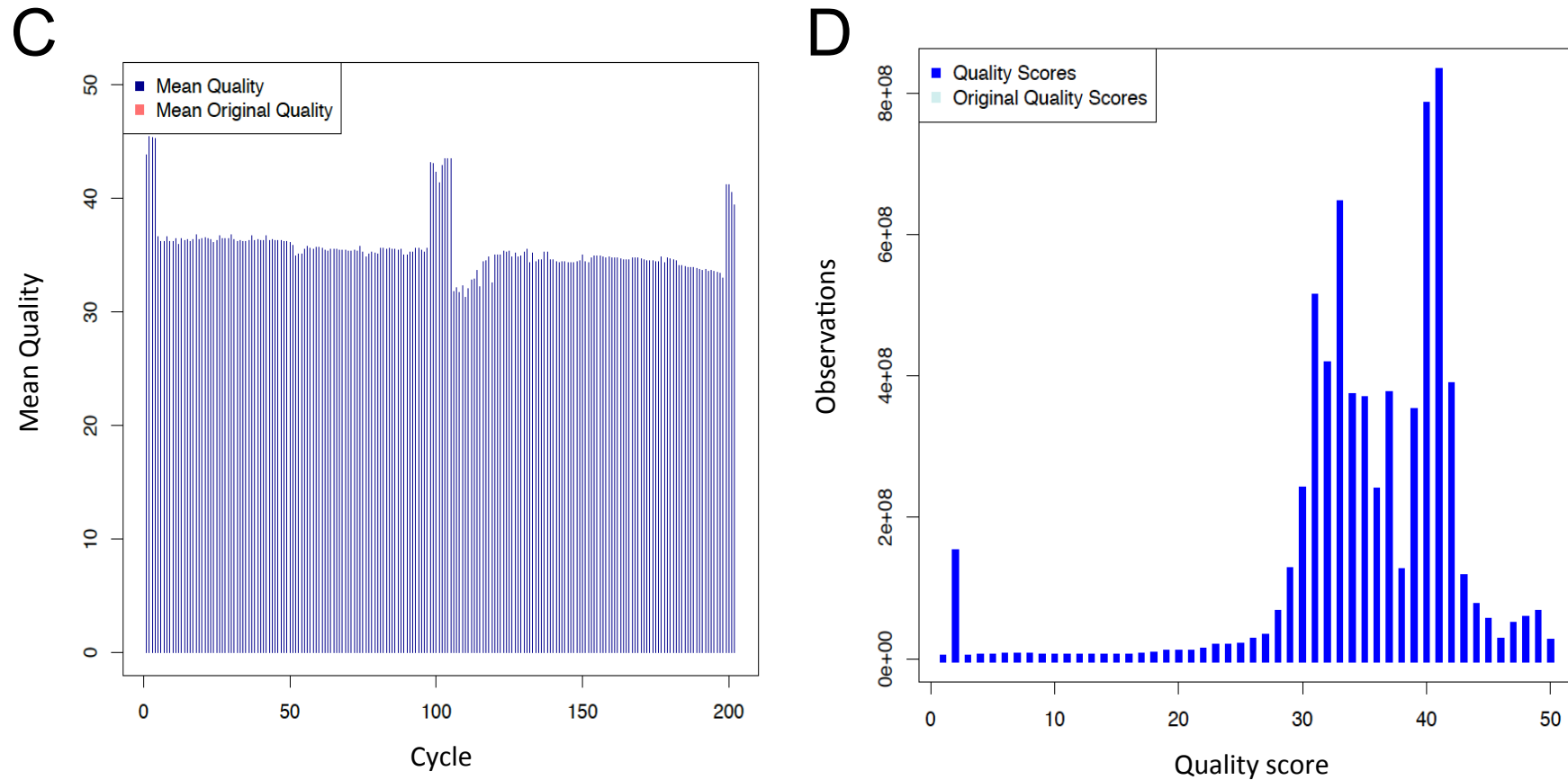


Figure 3-7: C & D: Quality scores for sample JT380 from Picard's Collect Multiple Metrics Tool.

C: Mean PHRED-scaled quality score by cycle, with each cycle representing the addition of a base, demonstrating a consistently high quality score across the read, the anomaly at base 101 represents the change-over for paired-end reads. D: The distribution of PHRED-scaled quality scores by observation count, showing that the majority of quality scores fall between 30 and 45.

3.3 Family with a Mutation in a Known Gene

3.3.1 Family CHD14

3.3.1.1 Clinical Phenotype

Family CHD14 is a consanguineous family of Pakistani origin who have two children affected with a similar phenotype. Their first affected child had coarctation of the aorta, hydronephrosis and a congenital diaphragmatic hernia. This individual died in the newborn period. The second affected child had ASD, episodes of Supraventricular Tachycardia (SVT), renal dilatation, absent bladder, neonatal encephalopathy, absent nails on the 3rd, 4th and 5th toes bilaterally and small 5th nails on the hands. She died at the age of 3 months. The family was referred with a possible diagnosis of Fryns syndrome. DNA was available from the second affected child (JT724), two unaffected siblings (JT722, JT723) and both parents (JT720 and JT721).

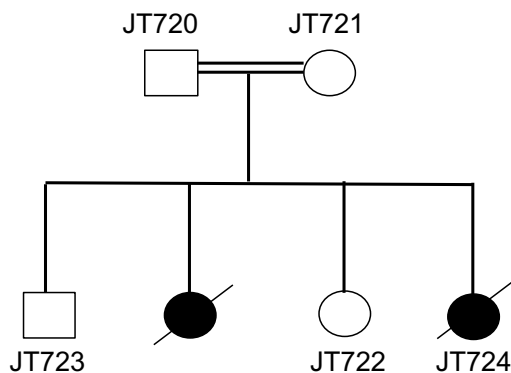


Figure 3-8: Pedigree of family CHD14.

Shading indicates an affected individual.

3.3.1.1.1 Autozygosity Mapping

Autozygosity mapping was performed on WES data from the available family members. Large regions of homozygosity present in the affected individual, JT724, but absent in her unaffected siblings and parents, are shown in Table 3-1. For brevity only the 10 largest homozygous regions are shown in this table, but all regions of >1000000 bp were interrogated.

Table 3-1: The ten largest regions of homozygosity identified in family CHD14.

Regions shown are present in the affected individual and absent in her unaffected siblings and parents.

Region	Size (bp)
chr13:46937074-71347587	24,410,513
chr1:23134218-46158768	23,024,550
chr11:20805066-33374613	12,569,547
chr4:109576843-119414828	9,837,985
chr1:65335323-75005776	9,670,453
chr4:28372261-38029729	9,657,468
chr1:103573517-113033459	9,459,942
chr14:84271985-92157528	7,885,543
chr4:13616778-21305653	7,688,875
chr1:47181952-54252939	7,070,987

3.3.1.1.2 Whole Exome Sequencing

WES libraries were set up using the Illumina QXT method and sequencing was performed on the Illumina Hiseq 3000 with nine samples pooled per lane. DNA from the affected individual, the two unaffected siblings and both parents were sequenced.

Data analysis assumed a recessive disease model. Table 3-2 shows the number of variants remaining following each stage of filtering. Table 3-3 shows the candidate homozygous variants remaining following filtering in this family.

Table 3-2: Number of variants remaining following each filtering step in family CHD14

Stage in Analysis	Number of variants
Called with GATK	688027
<1% dbSNP	125310
<1% EVS	124240
<1% ExAC	120234
Biallelic, filtered on family structure	73
Pass hard filters	41
CADD >15	21
Homozygous	5

Table 3-3: Homozygous variants remaining following filtering, assuming a recessive disease model, in family CHD14.

Gene	Variant	Protein change	Polyphen2	SIFT	CADD score	MAF (GnomAD)	Protein function or description	OMIM disease	In region of homozygosity
<i>TIE1</i>	1:43783657G>T	G946W; NP_005415	Deleterious (0)	Probably damaging (0.987)	31	19/248784	Tyrosine kinase with Ig and EGF homology domains		yes
<i>RIMBP2</i>	12:130927184C>T	R221K; NP_056162	Deleterious (0.02)	Probably damaging (0.999)	26.8	51/252108	Rims binding protein 2		no
<i>PIGV</i>	1:27121430T>C	L302P; NP_001189483	Deleterious (0)	Probably damaging (0.985)	26.1	5/252432	Phosphatidylinositol glycan anchor biosynthesis.	Hyperphosphatasia with mental retardation type 1.	yes
<i>WDR65</i>	1:43665167C>T	T512I; NP_001161437	Deleterious (0.01)	Possibly damaging (0.808)	24.3	23/252324	WD repeat containing protein	? Van der Woude syndrome	yes
<i>SORT1</i>	1:109870219C>T	S459N; NP_002950	Tolerated (0.32)	Benign (0.084)	22.3	3/227372	Sortilin is a VPS10-containing receptor that binds neuropeptides		yes

3.3.1.2 Sanger Sequencing

Sanger sequencing confirmed that the variant in *PIGV* segregated with disease in the family and was performed on a diagnostic basis by the Yorkshire Molecular Laboratory. These results were reported back to the consenting clinician.

3.3.1.3 Discussion of Candidate Variants

The variant in *PIGV* was the most notable of the variants that remained after filtering. Although it did not carry the highest CADD score, the variant occurred in a gene associated with a known disease phenotype. The other variants with higher CADD scores were in genes that were less convincing as functional candidates, and the variant in *RIMBP2* did not fall within a large (>1,000,000 bp) region of homozygosity.

PIGV is known to cause Hyperphosphatasia and Mental Retardation Syndrome (HPMRS) or Mabry syndrome (MIM: 239300). This is an autosomal recessive condition with the features of dysmorphism (hypertelorism, downturned mouth, long palpebral fissures and thin upper lip), mental retardation, seizures, and raised levels of alkaline phosphatase, which can be up to 20 times the normal range (Krawitz et al., 2010). Other associated features include hypoplastic nails of the terminal phalanges, anorectal abnormalities, Hirschsprung's disease, cleft lip and palate, vesicoureteric abnormalities, CHDs such as VSD and PFO, and deafness (Horn et al., 2011, Horn et al., 2014).

PIGV encodes GPI mannosyltransferase 2, an enzyme that acts in the Glycosylphosphatidylinositol (GPI) anchor biosynthesis pathway. Many different proteins require GPI anchors to function, and the steps involved in synthesizing the GPI anchor involve at least 18 different proteins. GPI-anchored proteins fall into many functional classes, for example receptors and adhesion molecules (Krawitz et al., 2010). It is thought that the phenotype observed in HPMRS is due to the dysfunction of proteins that require a GPI anchor. The reason that alkaline phosphatase is raised in these patients is that alkaline phosphatase is a GPI-anchored protein that is normally expressed on the cell surface. In normal cells, secretion of alkaline phosphatase requires GPI transamidase, which cleaves the GPI attachment signal, and attaches a mature GPI anchor, so that the protein is expressed on the cell surface and a small amount only is secreted in the serum. In *PIGV* mutant cells, GPI transamidase cleaves the attachment signal, but GPI anchoring cannot be completed correctly and so a larger amount of alkaline phosphatase is secreted (Murakami et al., 2012). HPMRS can also be caused by mutations in other genes of the same pathway, including

PGAP2 (Krawitz et al., 2013), *PIGO* (Krawitz et al., 2012), *PIGL* (Fujiwara et al., 2015), *PGAP3* (Howard et al., 2014), *PIGW* (Chiyonobu et al., 2014) and *PIGY* (Ilkovski et al., 2015).

The phenotype of HPMRS fits well with that of family CHD14, and the identified variant has previously been reported as a pathogenic mutation in the literature (Horn et al., 2014). The bladder absence noted in the second affected child has not previously been reported in this syndrome and neither has congenital diaphragmatic hernia, making this an interesting observation. On referral, this family was initially suspected to have Fryns syndrome (MIM: 229850), a condition consisting of congenital diaphragmatic hernia, dysmorphic facial features, pulmonary hypoplasia and distal limb hypoplasia (Slavotinek et al., 2005). Interestingly, the only gene known to cause Fryns syndrome is *PIGN*, a further member of the GPI anchor biosynthesis pathway (McInerney-Leo et al., 2016). Our findings suggest that there may be molecular overlap between Fryns syndrome and HPMRS, and screening other genes in the GPI anchoring pathway in cases of Fryns syndrome may help to further understand the genetic aetiology of this condition. Following the finding of this mutation, previous biochemical testing in both affected children was assessed. No results were available from the first child in this family, but the second child was found to have had an alkaline phosphatase level of 2573 IU/L (normal range 44 to 147 IU/L). This clinical finding was strongly supportive of the molecular diagnosis in this family. This finding and the subsequent analysis have been reported back to the referring clinician.

3.4 Families Affected with CHD

3.4.1 Clinical Phenotypes

The results from a total of 10 families are summarized in this section. All were recruited under the CHD project. Table 3-4 shows a summary of the clinical features of the patients in these families. The pedigrees for these recruited families are shown in Figures 3-8 to 3-12. Patients identified by 'JT' numbers (e.g. JT384) are those individuals for whom DNA samples were available.

Table 3-4: Clinical features of families recruited to the project with CHD phenotypes, but without mutations identified in known disease genes.

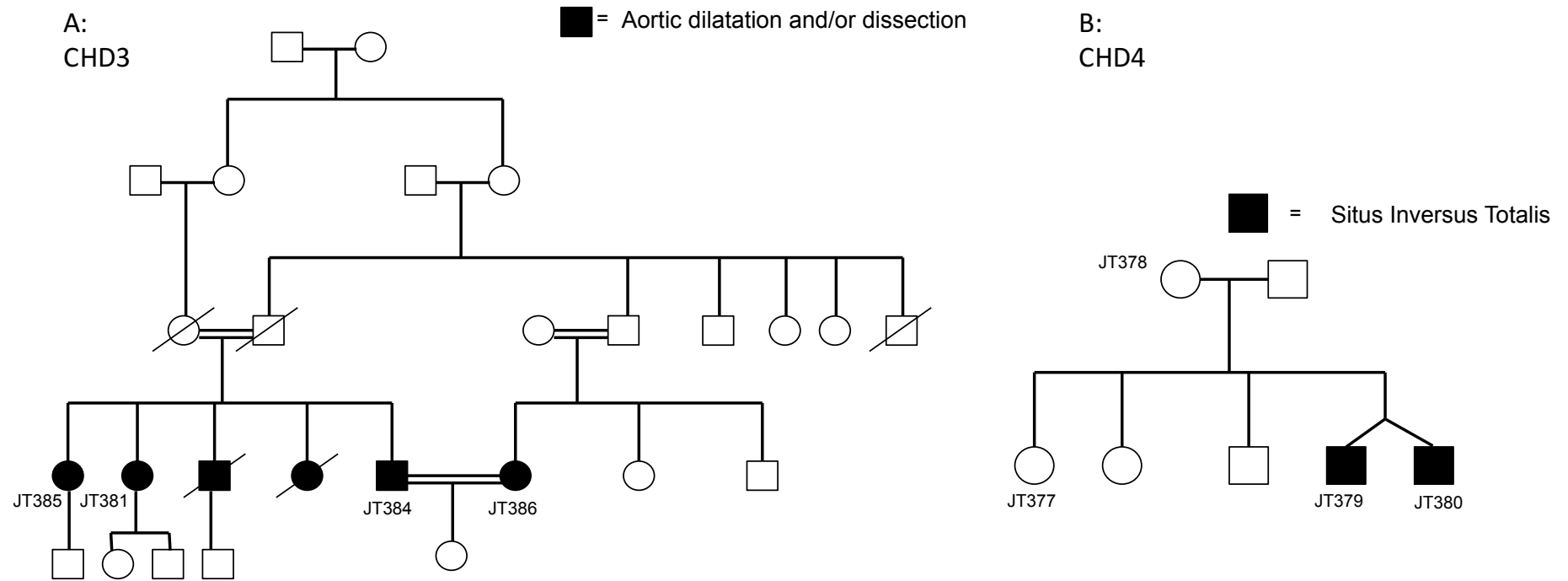
M= male, F= female, NC= non-consanguineous, C= consanguineous, AR= autosomal recessive, DN= de novo. Age of affected individual is at last clinical assessment, *=deceased at this age. TAAD= Thoracic aortic aneurysm and dissection, SIT= Situs inversus totalis, DCM= Dilated Cardiomyopathy, TOF= Tetralogy of Fallot, PH= Pulmonary Hypertension, HLHS= Hypoplastic left heart syndrome, TA= Truncus Arteriosus.

Family no.	JT no.	Ethnic origin	Consanguinity	Likely inheritance	Affected individual no. (JT no.)	Sex	Age	CHD	Other anomalies	Dysmorphism	Clinical test results
CHD3	JT381, JT384, JT385, JT386	Pakistani	C	AR	1 (JT385)	F	49yrs	TAAD		None	
					2 (JT381)	M	52yrs	TAAD	Subarachnoid haemorrhage, fusiform posterior circulation aneurysms	None	Normal clinical exome for TAAD
					3	F	43yrs*	TAAD		None	
					4	M	36yrs*	TAAD		None	
					5 (JT384)	F	46yrs	TAAD		None	
					6 (JT386)	F	46yrs	TAAD		None	
CHD4	JT377- JT380	White British	NC	AR	1 (JT379)	M	4yrs	SIT	Recurrent chest infections, chronic cough	None	Normal ciliary beat and frequency. Array CGH normal. ZIC3 sequencing normal.
					2 (JT380)	M	4yrs	SIT	Strabismus	None	

Family no.	JT no.	Ethnic origin	Consanguinity	Likely inheritance	Affected individual no.	Sex	Age	CHD	Other anomalies	Dysmorphism	Clinical test results
CHD5	JT439- JT443, JT445, JT474, JT475	Pakistani	C	AR	1	M	Stillbirth 28/40			None	
					2	M	18wks*	DCM		None	
					3 (JT443)	F	7yrs	DCM, VSD		None	Karyotype and FISH 22q11 normal.
					4	M	Stillbirth 33/40	Ebstein anomaly, DCM		None	
					5 (JT445)	M	2 months	DCM		None	
CHD6	JT541- JT543	White British	NC	AR	1 (JT543)	F	10yrs	TOF		Tall stature	
					2 (JT542)	M	12yrs	PA, VSD, PH		Long slender thumbs	Array CGH normal
CHD7	JT446- JT448, JT476	Pakistani	C	AR	1	F	2 days*	TA and critical AS			
					2 (JT448)	M	3½ yrs	TA, VSD, ASD, TS	Nephro- calcinosis, FTT, GORD	Low set ears	Array CGH normal
CHD8	JT583- JT588, JT616	White British	NC	AR	1 (JT616)	F	4ds*	HLHS		Wide- spaced nipples, short neck	Array CGH normal.
					2 (JT588)	F	4½ yrs,	VSD, dextro- cardia	Failure to thrive, feeding difficulties, GORD.		Array CGH, renal and abdominal USS normal
					3 (JT587)	F	6mths	VSD			

Family no.	JT no.	Ethnic origin	Consanguinity	Likely inheritance	Affected individual no.	Sex	Age	CHD	Other anomalies	Dysmorphism	Clinical test results
CHD 12	JT622-JT625	White British	NC	AR	1 (JT624)	F	23yrs	Secundum ASD		None	
					2 (JT625)	F	20 yrs	Secundum ASD		None	
CHD 13	JT711-714	Pakistani	C	AR/DN	1 (JT714)	F	2mths	ASD/VSD	Retinal colobomas, horseshoe kidney, segmentation abnormality, cerebellar vermis hypoplasia, hypoplasia of corpus callosum, elongation of superior cerebellar peduncles.	Tall forehead, deep set eyes, down-turned mouth. Depigmentation abnormality arm & leg.	Array CGH normal. Joubert/Meckel syndrome panel normal. Dosage analysis failed.
CHD 15	JT666-JT668	Afghani-stani	C	AR/DN	1 (JT668)	M	TOP at 20 weeks	AVSD	Ventriculomegaly.	-	

Family no.	JT no.	Ethnic origin	Consanguinity	Likely inheritance	Affected individual no.	Sex	Age	CHD	Other anomalies	Dysmorphism	Clinical test results
CHD 16	JT785-788	Pakistani	C	AR	1 (JT788)	M	5yrs	TOF, VSD	Renal agenesis, deafness, talipes, hypospadias, developmental delay	Low set ears, strabismus	Array CGH and Fanconi's anaemia testing normal
					2	M	2mths	TOF, VSD		None	



CHD5

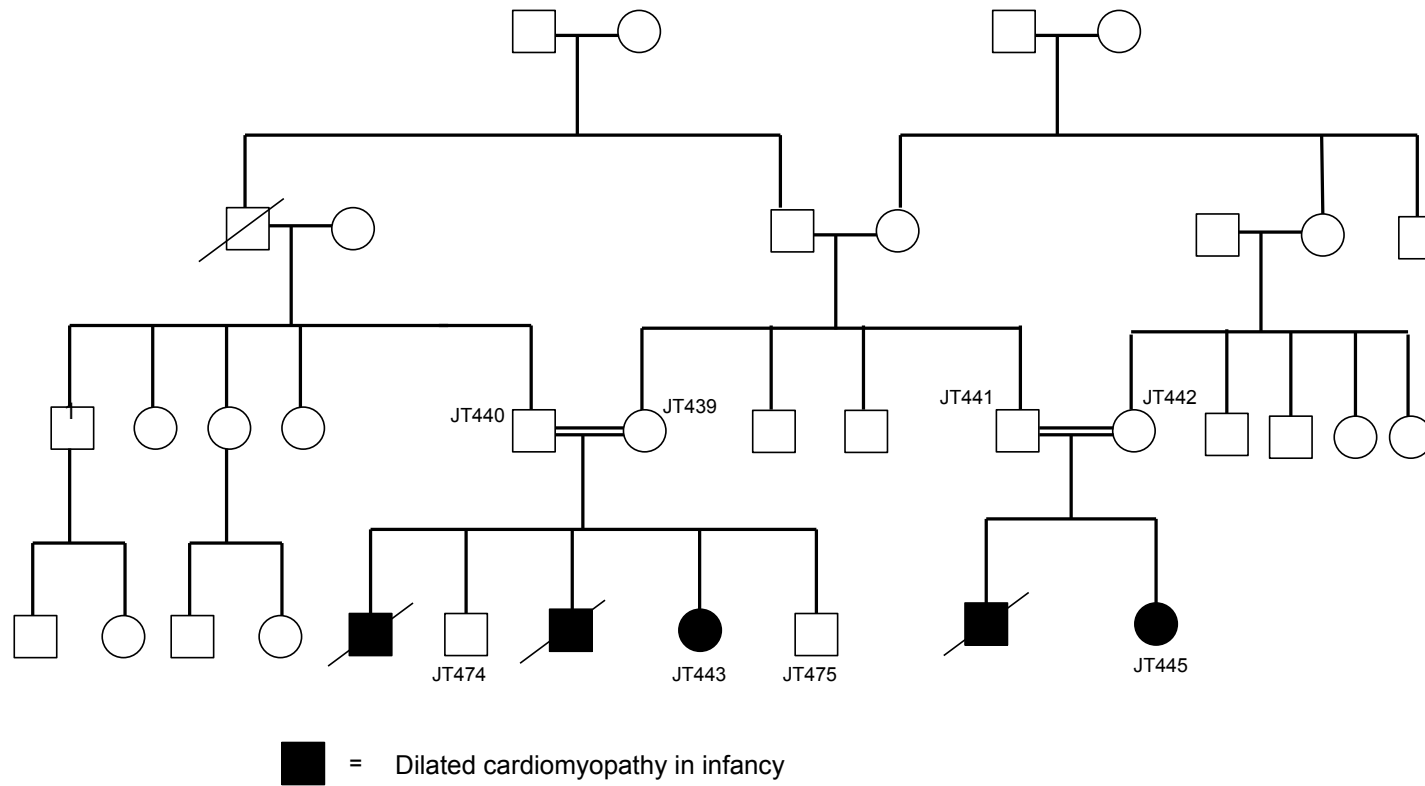


Figure 3-10: Pedigree of family CHD5

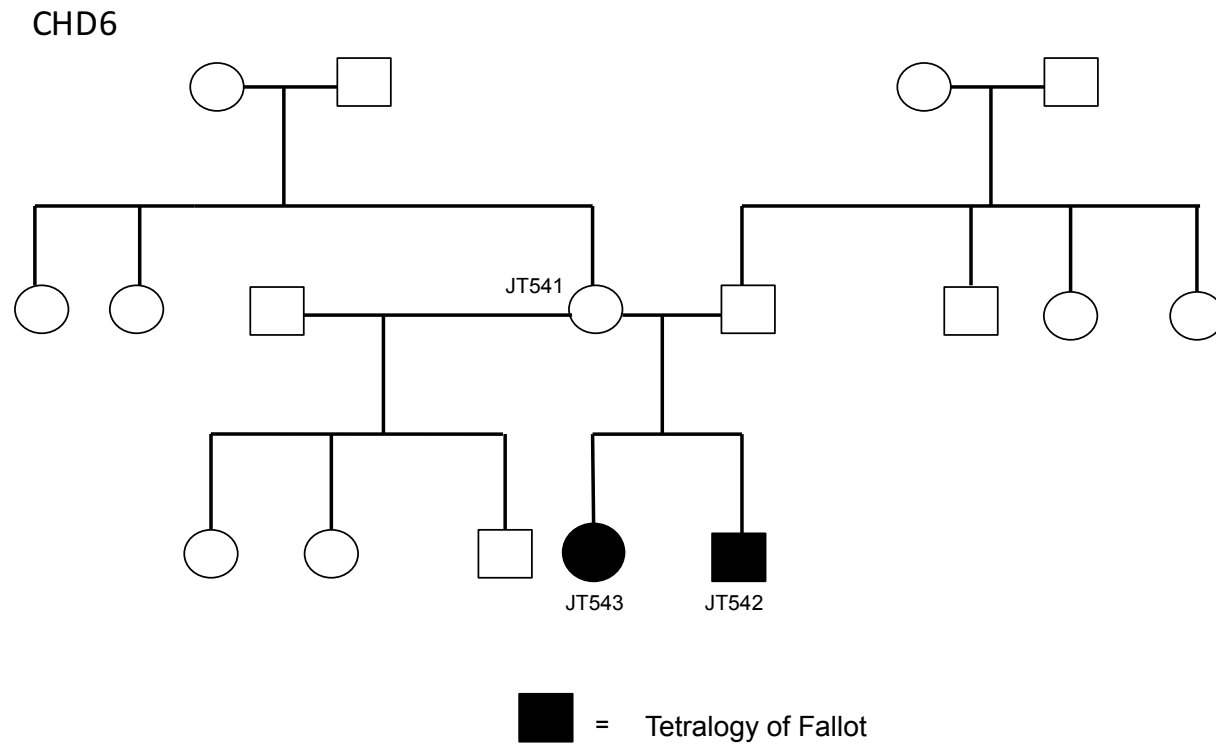


Figure 3-11: Pedigree of family CHD6.

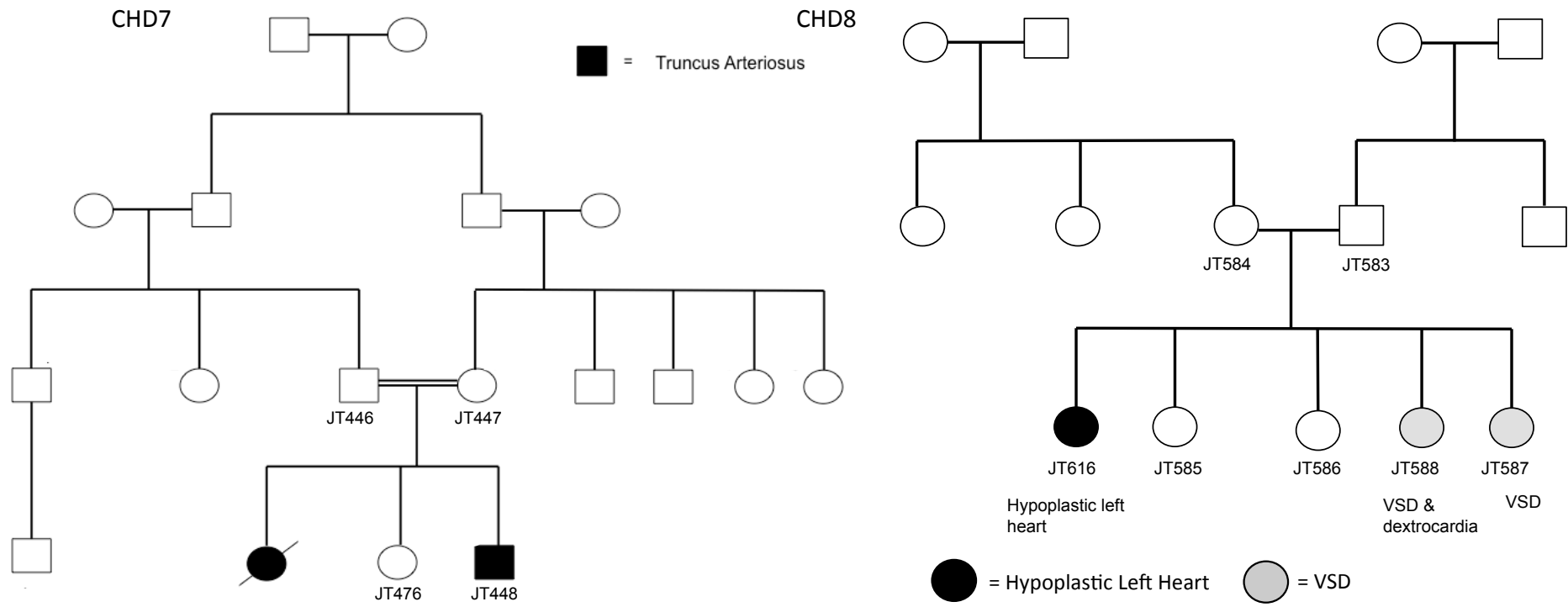


Figure 3-12: Pedigrees of families A: CHD7 and B: CHD8.

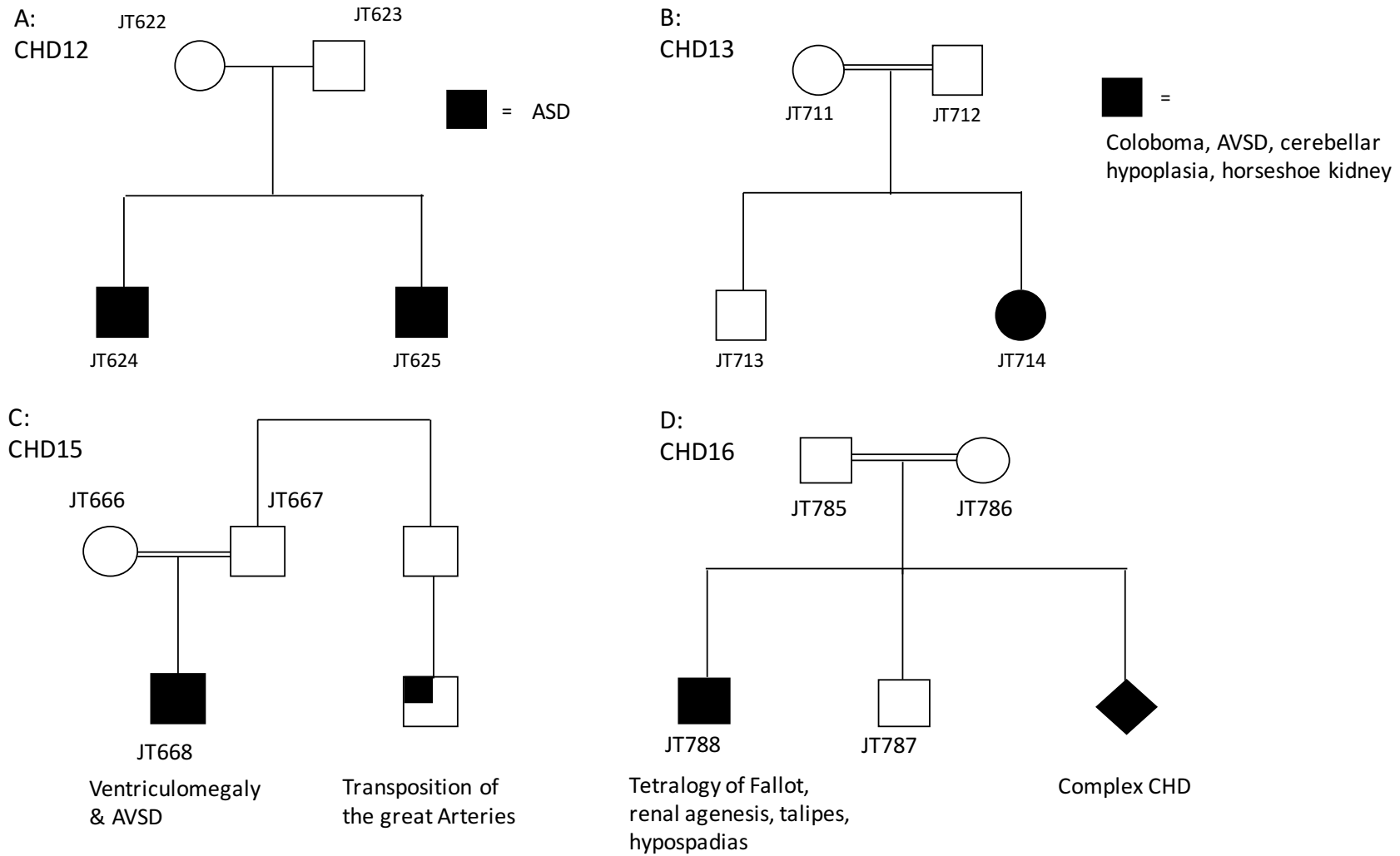


Figure 3-13: Pedigrees of families A: CHD12, B: CHD13, C: CHD15 and D: CHD16.

3.4.2 Whole Exome Sequencing

Variants were filtered as previously described (Chapter 2, section 2.2.14.3). Those that remained after filtering were rare (<1%) in control populations, predicted to have an effect on protein function, passed hard filters and their segregation was compatible with the assumed inheritance pattern in the family. Where more than one inheritance pattern was possible in the family, variants consistent with all possible inheritance patterns were assessed. Known disease-causing genes, appropriate to each phenotype, were assessed independently of this filtering. For a list of known genes assessed in each family see appendix G. Table 3-5 shows the variants remaining in each family following filtering. Only variants with a CADD score >15 and compatible with the assumed inheritance patterns are shown. For heterozygous variants remaining following filtering in family CHD3 see appendix H.

In family CHD5, there were no variants that remained after filtering, based on a recessive disease model if both JT443 and JT445 are considered as affected individuals. Variants were therefore assessed separately for either individual JT443 or JT445. Initially, this was performed assuming a recessive model of inheritance. These variants are shown in Table 3-5. *De novo* variants present in each individual were also assessed, but no candidates remained after filtering.

In family CHD13, analysis was performed to look for *de novo* variants in the affected individual, but none remained after filtering, variants reported are those identified under a recessive disease model. In families CHD15 and CHD16 only homozygous variants, identified under a recessive disease model are shown in Table 3-5 due to the large number of biallelic candidate variants that remained after analysis.

Table 3-5: Candidate variants remaining following filtering of WES data in CHD families.

Family, inheritance	Gene	Variant	Protein change	Polyphen 2	SIFT	Condel	CADD score	MAF GnomAD	Protein description or function	Omim disease	Homozygous region
CHD3, recessive	<i>TRBV5-4</i>	7:142168590 G>GAT;	44-45 fs; ENST0000 0454561.2	-	-	-	-	7/238432	T cell receptor		
	<i>TRBV5-4</i>	7:142168593 A>G;	S44P; ENST0000 0454561	Benign (0)	Tolerated (1)	Neutral (0)	0.019	8/238410			
	<i>PKHD1L1</i>	8:110374855 C>T	L16F; NP_803875	Possibly damaging (0.86)	Deleterious (0.04)	Deleterious (0.681)	25.0	8/239408	Polycystic kidney and hepatic disease 1- like 1		
	<i>PKHD1L1</i>	8:110437396 T>G	I927S; NP_803875	Benign (0.034)	Tolerated (0.07)	Neutral (0.321)	21.2	21/248582			
CHD4, recessive	<i>BCR</i>	22: 23653975 T>TCCGG	S/SGX; NP_004318	-	-	-	35	-		Leukemia chronic myeloid (608232)	
	<i>BCR</i>	22:23654017 G>A	D1062N; NP_004318	Benign (0.335)	Tolerated (0.15)	Neutral (0.346)	28.4	6/220486			
	<i>GPM6B</i>	X:13803811 C>T	A100T; NP_001001 995.1	Benign (0.002)	Tolerated (1)	Neutral (0)	10.16	7/204594	Membrane glycoprotein.		
	<i>GPRASP 1</i>	X:101913016 AC>A	D1392X; NP_055525	-	-	-	-	-	G protein- coupled receptor associated sorting protein.		

Family, inheritance	Gene	Variant	Protein change	Polyphen 2	SIFT	Condel	CADD score	MAF GnomAD	Protein description or function	Omim disease	Homozygous region
CHD5, JT443 as index case, recessive	<i>MFSD2B</i>	2:24247060 T>C	L470P; NP_001073942	Probably damaging (0.967)	Deleterious (0)	Deleterious (0.853)	16.70	2/251944	Major facilitator superfamily domain-containing protein		no
	<i>KLHL29</i>	2:23914668 G>A	A402T; NP_443152	Benign (0.358)	Tolerated (0.57)	Neutral (0.120)	15.51	2/156354	Kelch-like family member		no
	<i>ZFPM1</i>	16:8859928 6 TC>T	378 fs; NP_722520	-	-	-	-	-	Zinc-finger protein, transcription factor		no
		16:88600318 C>T	A651V; NP_722520	Benign (0.069)	Tolerated (0.25)	Neutral (0.041)	10.37	5/17348			no
CHD5, JT445 as index case, recessive	<i>DYM</i>	18:4681295 1C>T	G267S; NP_060123	Benign (0.02)	Tolerated (0.36)	Neutral (0.021)	18.18	1/251606	Dymecilin, secretory pathway, endochondral bone formation	Dyggve-Melchior-Clausen disease, (223800), Smith-McCort dysplasia, (607326)	no
	<i>MLXIPL</i>	7:73011601 G>A	A505V; NP_116569.1	Benign (0.011)	Tolerated (0.28)	Neutral (0.032)	12.32	98/233376	Helix-loop-helix leucine zipper transcription factor		no
	<i>EIF4H</i>	7:73609136 C>A	P179T; NP_071496	Benign (0.039)	Tolerated (0.55)	Neutral (0.009)	11.31	8/249532	Translation initiation factor		no

Family, inheritance	Gene	Variant	Protein change	Polyphen2	SIFT	Condel	CADD score	MAF GnomAD	Protein description or function	Omim disease	Homozygous region
CHD6, recessive	<i>SOX30</i>	5:15706529 3 T>C	R609G; NP_848511	Probably damaging (0.94)	Deleterious (0)	Deleterious (0.830)	25.3	13/229318	SRY-related HMG-box transcription factor.		
	<i>SOX30</i>	5:157065635 C>T	V495M; NP_848511	Possibly damaging (0.729)	Deleterious (0)	Deleterious (0.710)	27.6	537/282572 One homozygote			
	<i>INPP5D</i>	2:234112999 C>A	A1056D; NP_001017915.1	Benign (0.148)	Deleterious (0.02)	Neutral (0.396)	9.718	10/277136	N-terminal SH2 domain.		
	<i>INPP5D</i>	2:234113160 A>T	I1110F; NP_001017915.1	Possibly damaging (0.628)	Tolerated (0.15)	Neutral (0.460)	18.82	-			
CHD7, recessive	<i>TMEM260</i>	14:57088379T >C	C453R; NP_060269	Probably damaging (0.99)	Deleterious (0)	Deleterious (0.886)	25.7	2/250452	Unknown		yes
	<i>SULT4A1</i>	22:44229552C >T	V191M; NP_055166	Possibly damaging (0.563)	Deleterious (0)	Deleterious (0.637)	24.6	5/193618	Brain specific sulfo-transferase family		yes
	<i>C2Orf88</i>	2:191064718 CAA>C	K45X; NP_001035984.1	-	-	-	22.4	16/252072	Protein Kinase A Anchoring protein		yes
	<i>RREB1</i>	6:7232014 G>A	E1228K; NP_001003699.1	Benign (0.094)	Tolerated (0.09)	Neutral (0.303)	20.5	123/281976	Zinc finger transcription factor, binds to Ras.		yes

Family, inheritance	Gene	Variant	Protein change	Polyphen 2	SIFT	Condel	CADD score	MAF (GnomAD)	Protein description or function	Omim disease	Homozygous region
CHD8, recessive	<i>SYDE2</i>	1:85643954A >G	V849A; NP_115560	Probably damaging (0.999)	Deleterious (0.01)	Deleterious (0.896)	31	28/210048	GTPase activator activity		
		1:85656421C >T	A254T; NP_115560	Benign (0.002)	Deleterious (0)	Neutral (0.445)	23.5	1040/2635524 homozygotes			
CHD12, recessive	<i>FAM136A</i>	2:70524477G >C	L121V; NP_116211	Probably damaging (0.935)	Deleterious (0.02)	Deleterious (0.706)	26.0	3412/25644	Unknown	? Meniere disease	
		2:70524452A >G	M236T; NP_116211	Possibly damaging (0.9)	Deleterious low confidence (0)	Deleterious (0.709)	25.6	6656/276472			
CHD13, recessive	<i>ZNF302</i>	19:35175751 C>T	S270L; NP_001012320	Possibly damaging (0.804)	Deleterious (0)	Neutral (0.455)	25.9	181/252078	Zinc Finger transcription factor.		
		19:35175676 C>G	T245R; NP_001012320	Probably damaging (0.939)	Deleterious (0.01)	Neutral (0.382)	22.9	20/252260			
	<i>PLEC</i>	8:144999751 C>T	R1586H; NP_958782	Unknown (0)	Unknown (0)	Neutral (0.405)	23.8	43/178284 one homozygote	Plectin, intermediate filament binding protein	EB and muscular dystrophy (226670)	
		8:144994877 G>C	L3175V; NP_958782	Benign (0.279)	Unknown (0)	Deleterious (0.543)	5.76	24/239344			
	<i>RASGRF1</i>	15:79254524 C>T	E1262K; NP_002882	Benign (0.006)	Tolerated (0.14)	Neutral (0.403)	23.6	90/250194	Guanine nucleotide exchange factor (GEF)		yes

Family, inheritance	Gene	Variant	Protein change	Polyphen 2	SIFT	Condel	CADD score	MAF GnomAD	Protein description or function	Omim disease	Homozygous region
CHD15, homozygous	<i>TBCA</i>	5:77072074T>G	D3A; NP_004598	Possibly damaging (0.682)	Deleterious (0)	Deleterious (0.719)	29.6	-	Tubulin-specific chaperone A		yes
	<i>CDH19</i>	18:64172411C>T	A653T; NP_066976	Probably damaging (0.999)	Deleterious (0.01)	Deleterious (0.628)	29.6	12/251684	Type II cadherin gene		yes
	<i>PIGR</i>	1:207108001G>A	S490L; NP_002635	Probably damaging (0.966)	Deleterious (0)	Deleterious (0.616)	28.6	123/282596	Polymeric immunoglobulin receptor		no
	<i>ZMYM6</i>	1:35477514A>T	S347T; NP_009098	Probably damaging (0.997)	Tolerated (0.12)	Neutral (0.409)	27.3	-	Zinc Finger protein		yes
	<i>SERPINB11</i>	18:61390376G>T	D308Y; NP_536723	Probably damaging (0.966)	Deleterious (0)	Deleterious (0.713)	24.8	30/246458	Serine and cysteine protease inhibitor		yes
	<i>ZBTB46</i>	20:62378338G>A	P572L; NP_079500	Probably damaging (0.913)	Deleterious low confidence (0.01)	Neutral (0.372)	23.8	16/268004	Zinc finger protein expressed in effector lymphocytes		yes
	<i>STAB2</i>	12:104126886G>A	D1796N; NP_060034	Possibly damaging (0.828)	Tolerated (0.11)	Deleterious (0.543)	23.4	5/ 252160	Binds to Dil-labeled acetylated low density lipoprotein		no
	<i>MDM4</i>	1:204501360C>T	A110V; NP_002384	Benign (0.002)	Tolerated (0.16)	Neutral (0.427)	20.9	-	Critical regulator of p53		yes
CHD15, <i>de novo</i>	<i>CLOCK</i>	4:56329925T>TTCC	GE161Gfs; ENST0000513440		-	-	-	-	Transcription factor	?circadian rhythm	

Family, inheritance	Gene	Variant	Protein change	Polyphen 2	SIFT	Condel	CADD score	MAF GnomAD	Protein description or function	Omim disease	Homozygous region
CHD16, homozygous	<i>RBM48</i>	7:92166242 A>C	R365S; NP_115496	Probably damaging (0.999)	Deleterious (0)	Deleterious (0.658)	26.7	3/237872	RNA binding motif protein 48		yes
	<i>KLRD1</i>	12:10464155 C>T	R86W; NP_001107868	Probably damaging (0.992)	Deleterious (0)	Deleterious (0.643)	25.1	12/252390	Killer cell lectin-like receptor		no
	<i>CYP34A</i>	7:99359829G>A	T363M; NP_059488	Probably damaging (0.997)	Deleterious (0)	Deleterious (0.818)	23.5	49/282476	Cytochrome P450 superfamily of enzymes		yes
	<i>ZNF441</i>	19:11891859G>A	R407Q; NP_689568	Probably damaging (0.945)	Tolerated (0.7)	Neutral (0.346)	23.1	16/ 252100	Zinc finger protein		no
	<i>PRLHR</i>	10:120353948 C>A	R270L; NP_004239	Benign (0.1)	Deleterious (0)	Neutral (0.447)	23.3	6/238416	Prolactin-releasing hormone receptor		yes
	<i>ACADSB</i>	10:124802613 C>T	H245Y; NP_001600	Benign (0.368)	Deleterious (0.03)	Deleterious (0.559)	22.8	10/ 252392	Acyl-CoA dehydrogenase.	2-methylbutyryl-glycinuria (610006)	yes
	<i>NPAS4</i>	11:66192333G>A	G658R; NP_849195	Benign (0.163)	Tolerated low confidence (0.12)	Neutral (0.437)	17.01	100/ 249968	Neuronal PAS-domain protein, transcriptional regulator		yes

3.4.3 CNV analysis

CNV analysis was carried out using whole exome data in all samples where WES was performed in this study. WGS at low read depth provides a cost-effective method of CNV assessment (Dong et al., 2016). WGS gives coverage not only of the exonic regions, but also of the non-coding regions, meaning that CNVs are less likely to be missed, particularly if they are present in or overlap with non-coding regions. Resources were not available to perform WGS at low read depth in all samples and so two samples from two families were selected for this analysis. These samples were selected on the basis of a strongly suggestive recessive inheritance pattern, but no clear causative homozygous variant identified on WES data analysis.

3.4.3.1 Whole Genome Sequencing at Low Read Depth

WGS at low read depth was performed for samples JT381 (family CHD3) and JT443 (family CHD5). The results were analyzed for CNV calls, as described in Chapter 2, section 2.2.15, using the CNV-seq, CNVnator and ReadDepth programs. Tables 3-6 and 3-7 show a summary of CNV calls that were called by all three programs. Appendix I contains a summary of CNV calls from all programs. As can be seen in these tables, the majority of calls were present in the Database of Genomic Variants (DGV) (<http://dgv.tcag.ca/dgv/app/home>), making them unlikely to be pathogenic. Variants were also assessed for CHD phenotypes present in patients with overlapping deletions or duplications in the Decipher database (<https://decipher.sanger.ac.uk>).

Table 3-6: A summary of CNV calls from patient JT381 which were called by all three CNV analysis programs; CNV-seq, CNVnator and ReadDepth.

Location	Del/Dup	Genes in region	Gene description/function	Present in DGV	CHD present on Decipher
1:149041231-149279969	Del	<i>NBPF23</i>	Neuroblastoma breakpoint family	yes	none
1:149315068-149462199	Del	<i>FCGR1C</i> <i>FAM72C</i>	Fc of IgG	yes	Single ventricle
4:69374731-69499460	Del	<i>TMPRSS11E</i> <i>UGT2B17</i>	Transmembrane protease, serine 11E UDP Glucuronosyltransferase 2 family,	yes	none
5:12515668-12744002	Del	<i>CD49</i>	Homo sapiens cancer/testis antigen 49 (non-protein coding)	yes	AVSD, VSD
12:88325711-88477691	Del	<i>C12ORF50</i> <i>C12ORF29</i> <i>CEP290</i>	Associated with Joubert syndrome	no	PS, VSD.
12:88494691-88566434	Dup	<i>TMTC3</i> <i>CEP290</i>	Transmembrane and tetratricopeptide repeat containing 3 Joubert syndrome	no	
15:24439711-24548858	Del	none		yes	none
22:24308696-24399168	Del	<i>GSTT2B</i> <i>DTL</i> <i>DDT</i> <i>GSTT2</i> <i>GSTTP1</i> <i>GSTT1</i> <i>CABIN1</i>	Glutathione S-transferase theta 2B D-dopachrome tautomerase-like D-dopachrome tautomerase calcineurin binding protein 1, cell receptor signalling	yes	none

Table 3-7: A summary of CNV calls from patient JT443 which were called by all three CNV analysis programs; CNV-seq, CNVnator and ReadDepth.

Location	Del/Dup	Genes in region	Gene description/function	Present in DGV	CHD present on Decipher
1:196713091-196827120	Del	<i>CFHR1</i> <i>CFHR2</i> <i>CFHR3</i> <i>CFHR4</i>	Complement factor H-related 1, associated with Hemolytic-uremic syndrome (HUS)	yes	-
6:162678378-162856400	Del	<i>PARK2</i>	Proteasomal degradation. Mutations in this gene are known to cause autosomal recessive juvenile Parkinson disease	Yes	PFO, ASD, VSD
7:8815943-8872405	Del	near <i>NXPH1</i>	Neurexophilin 1	yes	PDA, VSD
8:7232106-7469594	Del	<i>SPAG11B</i> <i>DEFB4B</i>	Sperm associated antigen 11B Defensin, beta 4B	yes	ASD, VSD, PDA
8:7560068-7831482	Del	<i>SPAG11A</i> <i>DEF84A</i>		yes	ASD, VSD, Pulmonary stenosis,
8:11955251-12106284	Del	<i>DEFB130</i> <i>USP17L7</i>	Ubiquitin specific peptidase 17-like family member 7	yes	ASD, Pulmonary stenosis, coarctation of the aorta, TOF.
8:39225268-39396877	Del	<i>ADAM3A</i>	ADAM metallopeptidase domain 3A	yes	VSD
13:58313491-58406893	Del	none		yes	No patients with cardiac defects

3.4.3.2 CNV Analysis Using Exome Depth

CNV analysis using the Exome Depth program was performed on one affected sample from each family. A full list of identified copy number variants can be found in appendix J. No putative pathogenic CNV containing a candidate gene could be linked to the phenotypes of any of the families in the study. Where heterozygous CNVs occurred, exome data was analyzed to look for any candidate SNPs in the hemizygous state (ie appearing in the homozygous state because a CNV was on one allele and a disease-causing mutation was on the other). In this manner, all candidate CNVs were ruled out as a cause for disease.

3.4.4 Comparison of Variants to Ciliopathy Databases

All identified candidate variants were compared to databases of cilia-related genes. None of the identified variants showed an overlap with these databases.

3.4.5 Fetal RNA Data Analysis

One publically available dataset of RNA expression data from a fetal heart sample (sample number SRR643779, gestation not given; downloaded from <http://www.ncbi.nlm.nih.gov/sra>) was analyzed and compared to results from WES analysis. The expression levels of candidate genes for families CHD1 to 16 are listed in Table 3-8, as a strategy to increase the evidence that any genes were good functional candidates. Where genes were expressed, an FPKM value is given (Fragments Per Kilobase of transcript per Million mapped reads). Candidate genes with the highest expression in fetal heart were *EIF4H* and *TBCA*.

Table 3-8: Candidate genes from CHD patients and their expression in fetal heart sample SRR643779.

Gene	CHD pedigree	Transcript	FPKM in fetal heart sample SRR643779
<i>DNAAF1</i>	1		Not detected
<i>CCDC158</i>	3		Not detected
<i>HIP1</i>	3	NM_005338	6.868
<i>PKHD1L1</i>	3	NM_177531	0.388
<i>PLCB3</i>	3	NM_000932	1.184
<i>PSMB7</i>	3		Not detected
<i>RCE1</i>	3		Not detected
<i>SLIT2</i>	3	NM_004787	1.326
<i>TRBV5-4</i>	3		Not detected
<i>BCR</i>	4	NM_004327	2.422
<i>GPM6B</i>	4		Not detected
<i>DYM</i>	5		Not detected
<i>EIF4H</i>	5	NM_03199	113.646
<i>KLHL29</i>	5	NM_052920	0.441
<i>MFSD2B</i>	5	NM_001080473	0.137
<i>MLX1PL</i>	5		Not detected
<i>ZFPM1</i>	5		Not detected
<i>INPP5D</i>	6	NM_005541	0.975
<i>SOX30</i>	6	NM_007017	0
<i>C2ORF88</i>	7	NM_032321	2.163
<i>RREB1</i>	7		Not detected
<i>SULT4A1</i>	7	NM_014351	0.0792
<i>TMEM260</i>	7		Not detected
<i>SYDE2</i>	8		Not detected
<i>KMT2D</i>	9		Not detected

Gene	CHD pedigree	Transcript	FPKM in fetal heart sample SRR643779
<i>PTPRM</i>	9	NM_001105244	2.241
<i>RBM25</i>	9		Not detected
<i>SOGA1</i>	9		Not detected
<i>PTCH1</i>	10		Not detected
<i>FAM136A</i>	12	NM_032822	23.634
<i>PLEC</i>	13		Not detected
<i>RASGRF1</i>	13	NM_153815	0.0129
<i>ZNF302</i>	13		Not detected
<i>PIGV</i>	14		Not detected
<i>RIMBP2</i>	14	NM_015347	0.00326
<i>SORT1</i>	14		Not detected
<i>TIE1</i>	14	NM_005424	4.411
<i>WDR65</i>	14		Not detected
<i>CDH19</i>	15		Not detected
<i>CLOCK</i>	15		Not detected
<i>MDM4</i>	15		Not detected
<i>PIGR</i>	15	NM_002644	0.556
<i>SERPINB11</i>	15	NM_080475	0
<i>STAB2</i>	15	NM_017564	0.583
<i>TBCA</i>	15	NM_004607	101.979
<i>ZBTB46</i>	15		Not detected
<i>ZMYM6</i>	15	NM_001195156	4.0264
<i>ACADSB</i>	16	NM_001609	10.75
<i>CYP34A</i>	16		Not detected
<i>KLRD1</i>	16	NM_001114396	1.896
<i>NPAS4</i>	16	NM_178864	0.096
<i>PRLHR</i>	16	NM_004248	0
<i>RBM48</i>	16		Not detected
<i>ZNF441</i>	16	NM_152355	3.0404

3.4.6 Sanger sequencing

Sanger sequencing was performed in family CHD7 to confirm and segregate the variants identified in *C2ORF88* and *RREB1*, as both were candidates of interest. Both variants were confirmed to segregate with disease status in the family. Figure 3-13 shows the results of *C2ORF88* sequencing. The frameshift variant, c.133_134delAA, is shown to be heterozygous in the unaffected mother, father and sibling and homozygous in the affected child. Figure 3-14 shows the results of *RREB1*

sequencing. The c.3682G>A variant is heterozygous in the three unaffected family members and homozygous in JT448, the affected child.

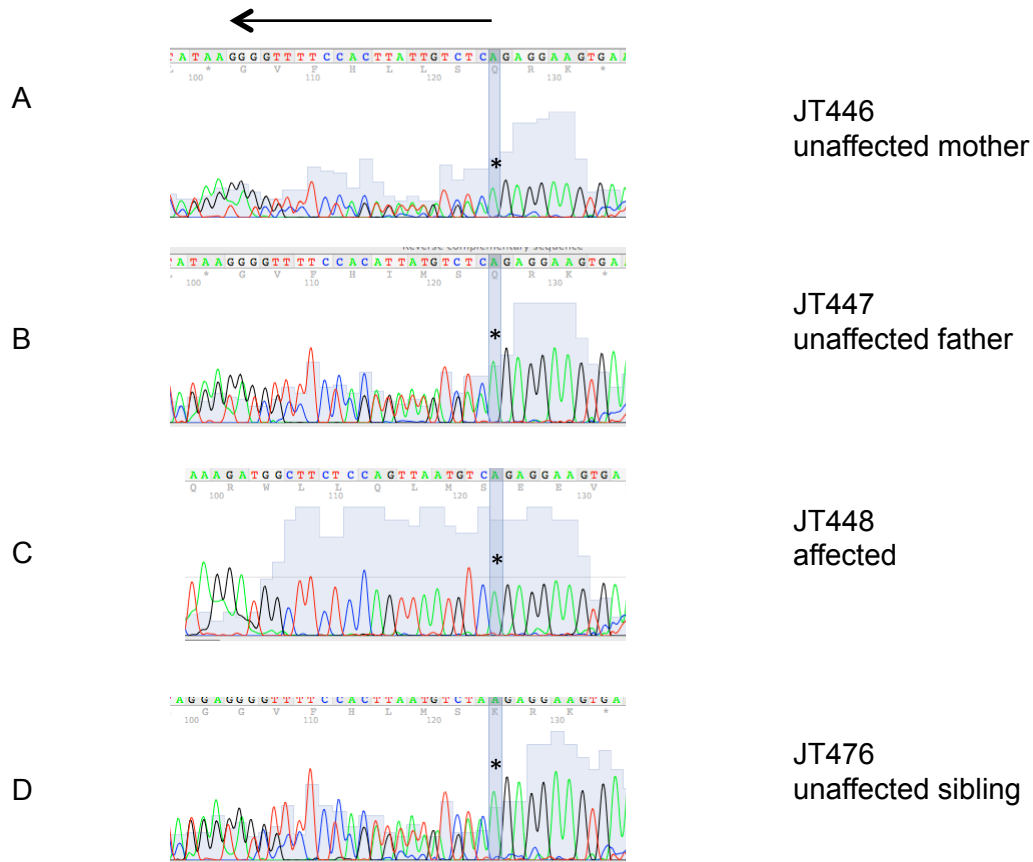


Figure 3-14: Sanger sequencing electropherograms to demonstrate a frameshift variant, c.133_134delIAA, in *C2ORF88* in family CHD7.

The reverse complementary sequence is shown. The frameshift mutation is shown to be heterozygous in the unaffected mother (A), unaffected father (B) and unaffected sibling (D) and homozygous in the affected child (C). The asterisk indicates the position of the mutation, grey shading indicates PHRED-score. The black arrow indicates the direction of sequencing, with discordance of the sequence following a heterozygous frameshift variant.

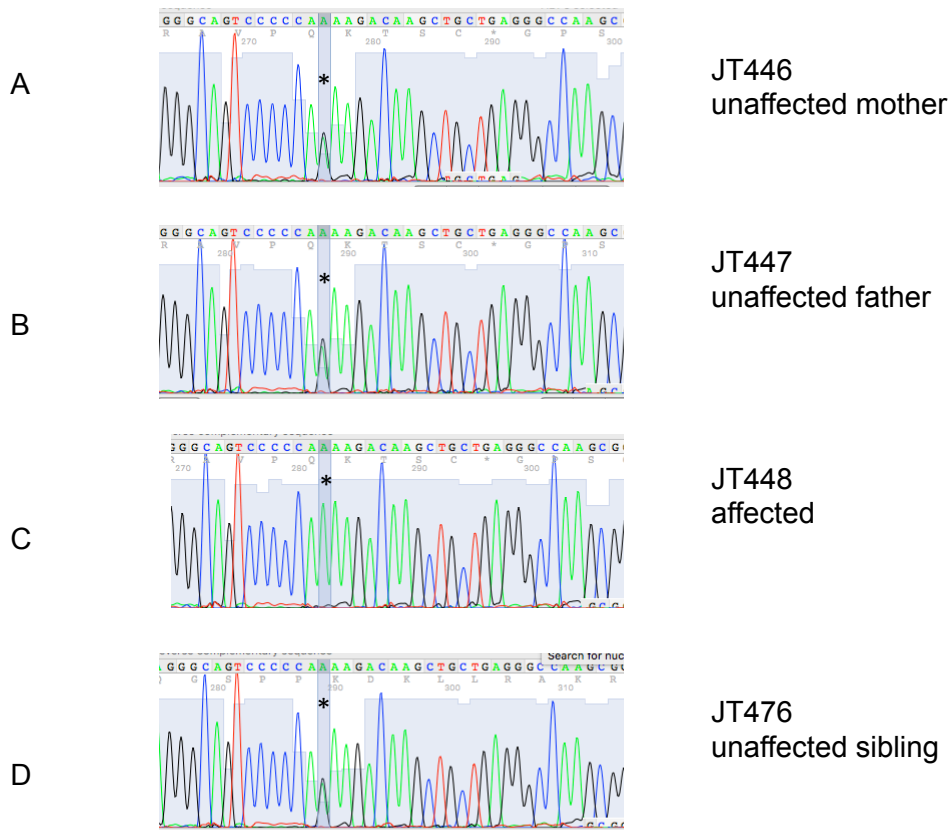


Figure 3-15: Sanger sequencing electropherograms to demonstrate the c.3682G>A variant in *RREB1* in family CHD7.

The variant is shown to be heterozygous in the unaffected mother (A), unaffected father (B) and unaffected sibling (D) and homozygous in the affected child (C). The asterisk indicates the position of the mutation.

Variants identified in *TMEM260* identified in this family were proven to segregate with disease and were confirmed on a diagnostic basis by Sanger sequencing by the Leeds Molecular Genetics Laboratory. Results were reported back to the referring clinician.

3.4.7 Discussion of Families with CHD

3.4.7.1 Discussion of Recruited Families

In this section we consider families that were recruited to the study but did not result in the identification of a mutation in a known CHD gene. The majority of families had an assumed autosomal recessive pattern of inheritance, comprising of unaffected parents with affected offspring or being consanguineous unions, for whom recessive disease is more likely to occur. Other inheritance patterns were possible in a number of the families, and so variant filtering was not strictly limited to a recessive disease model.

The phenotypes of the CHD families consisted of structural defects, such as ASD and Tetralogy of Fallot, myopathies such as Dilated Cardiomyopathy (DCM), and laterality

disorders. Both families with syndromic and non-syndromic disease were included. One family (family CHD3) was affected with aortic dilation and dissection. Although this is not strictly a CHD, this family was included due to the consanguineous nature of the family history, and early onset of disease without further risk factors, making a genetic cause more likely.

Overlap of CHD phenotype was seen in a number of families assessed, for example families CHD6, CHD7, CHD8, CHD12, CHD13 and CHD15 were affected with septal defects. Furthermore, Tetralogy of Fallot was seen in families CHD6 and CHD16.

Families with syndromic forms of CHD included families CHD13, CHD15 and CHD16. The phenotype in family CHD13 consisted of dysmorphic features, ASD, VSD, retinal colobomas, horseshoe kidney, cerebellar vermis hypoplasia, hypoplasia of the corpus callosum and slight elongation of the superior cerebellar peduncles. This would be consistent with 3C or Ritscher-Schinzel syndrome, a condition characterized by cerebellar, cardiac and craniofacial abnormalities. Cerebellar abnormalities are typically cerebellar vermis hypoplasia, Dandy-Walker malformation and enlargement of the cisterna magna. Craniofacial abnormalities include dysmorphic features, such as down-slanted palpebral fissures, low set ears and micrognathia, cleft palate and ocular colobomas. VSD and ASD are cardiac malformations seen in this syndrome (Leonardi et al., 2001). Horseshoe kidneys are not a typical feature of this condition, but have previously been reported (Seidahmed et al., 2011). Leonardi et al., 2001, have cited minimum diagnostic criteria for this condition, which the affected individual in family CHD13 would fulfill (Leonardi et al., 2001). This affected individual did have a depigmentation abnormality on one arm and leg, which is not known to be a characteristic of this condition, and can sometimes be an indication of chromosomal mosaicism, as well as a number of other syndromes.

The only gene previously identified to be associated with autosomal recessive 3C syndrome is *KIAA0196*, which was identified in a Canadian family who had homozygous splice site mutations in the gene. The c.3335+2T>A variant identified in these patients was shown to affect splicing (Elliott et al., 2013). Missense mutations in this gene were previously associated with a dominant spastic paraplegia (Valdmanis et al., 2007). This gene, and specifically this splice site mutation, was assessed in our affected patient and mutations were not present, indicating that there may be genetic heterogeneity in this condition. The described features in our family do have overlap with other genetic diagnoses. For example, optic colobomas and heart defects are

seen in CHARGE syndrome, Meckel-Gruber syndrome, CHIME syndrome (Coloboma-congenital heart disease-ichthyosiform dermatosis-intellectual disability-ear anomalies syndrome), and Kabuki syndromes. However, none of these other diagnoses fits with all of the features identified in this family.

Ventriculomegaly and AVSD were identified in the affected member of family CHD15. There are a large number of genetic conditions that could manifest with these features. A search on the London Dysmorphology Database (Guest et al., 1999) reveals 37 genetic diagnoses that could encompass this presentation, including a number of microdeletion syndromes, Smith-Lemli-Opitz syndrome and Kleeftstra syndrome. With only one affected pregnancy, this makes forming a differential diagnosis difficult. Genes associated with conditions that were identified as possible differential diagnoses were specifically assessed when analyzing the results from this family.

The phenotype of the affected individual in family CHD16 was of Tetralogy of Fallot, renal agenesis, deafness, talipes, hypospadias and developmental delay, with some mild dysmorphic features. Chromosomal diagnoses such as 22q11 deletion and Williams syndrome were ruled out by a normal result following array CGH analysis. Fanconi's anaemia was a possible differential diagnosis, but this is commonly associated with radial ray abnormalities, abnormal skin pigmentation, short stature and microcephaly, and a predisposition to bone marrow failure, which were not present in this child. However, an absence of these features does not rule out the diagnosis (Glanz and Fraser, 1982). McKusick-Kauffman syndrome was a further differential diagnosis for this presentation, but this is characterized by the presence of polydactyly, which was not present in this case (Schaefer et al., 2011). Genes known to cause these syndromic presentations were taken into consideration during the data analysis.

3.4.7.2 Discussion of Identified Candidate Variants in CHD Families

The candidate variants that remained after filtering of WES data for each family are described in this section, to highlight the strongest putative candidate gene in each family.

3.4.7.3 Family CHD3

In family CHD3, where a number of family members have been affected with aortic dilatation and dissection, only compound heterozygous variants in two genes remained following filtering. *TRBV5-4* encodes T-cell receptor beta variable 5-4 (Slightom et al., 1994), a T-cell receptor. This gene is unlikely to be related to aortic dilatation. The other gene was *PKHD1L1*. This gene is polycystic kidney and hepatic disease

autosomal recessive-like 1, a homolog of the autosomal recessive polycystic kidney disease gene (*PKHD1*). It is a large gene with 78 exons. It has been screened in a large number of patients with Autosomal Recessive Polycystic Kidney Disease (ARPKD), but no causative mutations have been identified (Hogan et al., 2003). It is predicted to encode a large receptor protein (fibrocystin-L) with a signal peptide, transmembrane domain and short cytoplasmic tail, similar to the PKHD1 protein. It is ubiquitously expressed at low levels. It was found to be up-regulated in T lymphocytes, so was suspected to play a role in cellular immunity (Hogan et al., 2003). ARPKD is an inherited childhood nephropathy. It is rare, affecting 1:20,000 live births, and is characterized by fusiform dilatation of the collecting ducts, alongside hepatic fibrosis. This is a severe condition: 30% of affected children die as neonates and progression to end stage renal disease is likely within 15 years (Harris and Rossetti, 2004). *PKHD1L1* initially seemed unlikely to be the cause of the phenotype in this family, as none of the family members are affected with renal disease. Furthermore, only one of the two variants was predicted to be pathogenic by *in silico* tools. However, the dominant form of polycystic kidney disease, caused by mutations in *PKD1* and *PKD2*, has been associated with vascular disease, including aortic aneurysms and dissections (Bichet et al., 2006), vascular endothelial dysfunction (Peterson et al., 2013) and a 10% risk of intracranial aneurysm (Chapman et al., 1992). Vascular disease is less common in ARPKD, but sub-arachnoid haemorrhage has been reported in three published cases (Chalhoub et al., 2013). ARPKD rat models also have impaired aortic vascular function, with a reduction in vasoreaction to vasodilators (Peterson et al., 2013). It is possible, therefore, that mutations in a gene with similar functions to *PKHD1* are related to the aortic disease seen in this family, and this remains the most likely candidate in this family. However, further functional work would be required to assess the effect of these mutations on the protein, and the possible association between this and renal disease. This is beyond the remit of the current research project.

Family CHD3 were also assessed for heterozygous variants, due to the number of family members who were affected with disease and the uncertain status of the father of the five affected siblings, who died suddenly of a myocardial infarction in Pakistan, and did not undergo post mortem examination. It is possible that he also suffered an aortic dilatation. Of the heterozygous variants that remained after filtering, the variant in *SLIT2* has the strongest link to vascular disease. *SLIT2* is a secreted glycoprotein that is a ligand for the ROBO family of immunoglobulin receptors, specifically Robo1. The main role of this protein is in neuronal migration and axon guidance. However, there is literature that suggests a role in cardiac development (Mommersteeg et al., 2013). For

example, murine embryos lacking Robo1 have absence of parts of the pericardium and *Robo1/2* double mutants have reduced sinus horn myocardium and abnormalities of the caval veins (Mommersteeg et al., 2013). Although no abnormalities of the aorta were identified in these mice, *Slit2* is expressed in mouse aortic smooth muscle cells (Ni et al., 2016). An additional role for SLIT2 is in angiogenesis and endothelial cell permeability, where it has been shown to inhibit VEGF-induced migration, tube formation and permeability through the activation of Robo4 (Jones et al., 2008). There is, however, no strong link between mutations in *SLIT2* and aortic dilatation and dissection, as seen in this family.

CNV analysis was performed in family CHD3 by a number of methods. CNVs on chromosome 12 were consistently identified by the programs used and were not in a region of known variation. This region was called as a concurrent deletion and simple duplication (copy number 1.5), both predicted to affect the *CEP290* gene. This region included the gene *CEP290*, mutations in which cause Joubert syndrome, a very different phenotype to that seen in this family and not known to be associated with aortic dilatation, although other CHDs have been reported in this condition (Karp et al., 2012).

3.4.7.4 Family CHD4

The only compound heterozygous variants that remained after filtering in family CHD4 were in *BCR*. In this family two children are affected with Situs Inversus Totalis (SIT). This gene is disrupted as part of a Philadelphia chromosome translocation. In this translocation, a fusion protein is created which contains some sequence from *BCR* and some from *ABL*. This condition is found in most cases of Chronic Myelogenous Leukaemia (CML) and some cases of Acute Lymphoblastic Leukaemia (ALL) and Acute Myelogenous Leukaemia (AML) (Kurzrock et al., 2003). However, the other roles of *BCR*, apart from this, are largely unknown. The protein it encodes has serine/threonine kinase activity and it is a GTPase-activating protein for RAC1. There is no evidence to link this gene to situs inversus.

Two variants were identified on the X chromosome that would segregate with disease in this family. One, in *GPM6B* was present in a hemizygous form in the control datasets, making it unlikely to be the cause in this family. The other was in *GPRASP1*, a G protein-coupled receptor associated sorting protein, which may be involved in lysosomal degradation (Moser et al., 2010).

Deletion of *GPRASP1* is identified in Xq22.1 del syndrome, along with a number of other genes. This syndrome causes epilepsy, mental retardation, cleft palate and dysmorphic features in females, with failure to thrive and respiratory failure in affected males (Cao and Aypar, 2016). *GPRASP1* is one of four critical genes for this phenotype. A mouse model of the deletion has growth delay, respiratory failure, epilepsy and cleft palate (Zhou et al., 2014). There is no known association with situs inversus and so this variant seems an unlikely candidate in this family.

3.4.7.5 Family CHD5

Family CHD5 consists of two sib-ships affected with dilated cardiomyopathy (DCM) in infancy. No homozygous variants were present in both affected individuals that passed the filtering criteria for family CHD5. There could be a number of reasons for this. It may be that the mutation is in an area poorly covered by the exome data, perhaps in a non-coding or promoter region. The assumed inheritance pattern in this family could be incorrect, perhaps with the two separate occurrences of DCM being coincidental, or due to environmental factors. For this reason, the two individuals were also assessed independently.

In patient JTT443, biallelic variants in *ZFPM1* were identified, one a frameshift variant and the other a missense variant. However, the missense variant was not predicted as deleterious by *in silico* tools. *ZFPM1*, also known as *FOG1*, encodes a zinc finger protein that binds GATA1, an essential transcription factor for the development of erythroid and megakaryocytic cells (Tsang et al., 1997). Mutations in *GATA1* are associated with X-linked macrothrombocytopenia with dyserythropoietic anemia (Freson et al., 2003).

The other identified variants were in *MFSD2B* and *KLHL29*. The variant in *KLHL29* did not have strong *in silico* support for pathogenicity. This is a member of the Kelch-like gene family and very little is known about its function (Dhanoa et al., 2013). The variant in *MFSD2B* had a higher pathogenicity prediction on *in silico* tools. It is part of the major facilitator superfamily of transport proteins. There is little research into its function, but a previous paper suggested a likely role of MFSD2A in adaptive thermogenesis and the two proteins share many sequence and structural characteristics (Angers et al., 2008).

In patient JT445, the three variants that remained after filtering were in *DYM*, *MLXIPL* and *EIF4H*. They were missense variants present in the homozygous form, but none

had a deleterious call from *in silico* tools. Mutations in *DYM* have previously been associated with Dyggve-Melchior-Clausen syndrome (MIM. 223800), a recessive skeletal dysplasia (El Ghouzzi et al., 2003). The pathogenic mutations described previously include nonsense, splice site, or frameshift mutations. This gene encodes dymeclin, which is expressed in the brain and co-localises with the Golgi apparatus, perhaps playing a role in trafficking to and from the cytoplasm (Dimitrov et al., 2009). This condition is not known to be associated with cardiomyopathy.

MLXIPL encodes a helix-loop-helix transcription factor of the Myc/Max/Mad family. It is also known as carbohydrate-responsive-element-binding protein (ChREBP). It is known to form a heterodimeric complex with the MAX-like protein and activate lipogenic and glycolytic enzymes (Merla et al., 2004). In adipose tissue, *MLXIPL* is regulated by GLUT4, and therefore plays a role in insulin sensitivity (Herman et al., 2012). *MLXIPL* is present in the region of the genome that is deleted in William-Beuren Syndrome (WBS) (7q11.23 deletion syndrome). WBS has previously been associated with cardiomyopathy (Mauser and Bonnemeier, 2012), but it is proposed that the reduction in levels of elastin, rather than other deleted genes is responsible (Mauser and Bonnemeier, 2012). *EIF4H* is located near to *MLXIPL* and is also a gene deleted in William-Beuren Syndrome. It is a eukaryotic translation initiation factor and contains an RNA binding domain sequence (Richter-Cook et al., 1998). There is no strong evidence to suspect either *MLXIPL* or *EIF4H* as a candidate gene for DCM.

3.4.7.6 Family CHD6

In family CHD6 two children were affected with Tetralogy of Fallot and Pulmonary Stenosis with VSD. Filtering on a recessive disease model gave only two remaining variants. One variant in *SOX30* was found to be homozygous in the GnomAD database, therefore ruling it out as a candidate. This leaves the compound heterozygous variants in *INPP5D*. These variants did not have strong *in silico* support for pathogenicity. This gene encodes an inositol polyphosphate-5-phosphatase. This hydrolyses inositol 1,4,5-triphosphate (Ins(1,4,5)P₃), which mobilizes intracellular calcium. The mouse orthologue of this gene is expressed in early development, specifically during hematopoiesis and spermatogenesis (Liu et al., 1998) and is suspected to play a role in differentiation and maintenance of hematopoietic cell lineages (Liu et al., 1998). Homozygous knockout mice have failure to thrive and reduced survival. They have a myeloproliferative-like syndrome with lung infiltration with macrophages (Helgason et al., 1998). Congenital heart disease was not a described feature in these mice, making this an unlikely candidate gene.

3.4.7.7 Family CHD7

Family CHD7, a consanguineous union, have two children affected with Truncus Arteriosus. The variant that remained after filtering with the highest CADD score fell in *TMEM260*, a transmembrane protein. Until very recently little was known about the function of this protein, also called *C14ORF101*, although a polymorphism in this gene had been associated with survival in non-Hodgkins lymphoma (Yang et al., 2014) and risk of gastric cancer development (Wang et al., 2016a).

In April 2017, a paper was published identifying *TMEM260* to be the cause of disease in two families with a 'neurodevelopmental, cardiac and renal syndrome' (Ta-Shma et al., 2017). The authors present two families from consanguineous pedigrees, one of Ashkenazi Jewish heritage and the other of Arabic heritage. There were three affected children in the first family and two in the second. The key described features included cardiac defects (VSD in four, two had truncus arteriosus, one had TOF, and one just stated as CHD), Central Nervous System (CNS) defects (agenesis of the corpus callosum in two, microcephaly in one and no defect in one), and renal defects. The renal defects were variable and consisted of elevated creatinine, which was accompanied with oliguria in one case and anuria in another, and one patient had renal cysts. One individual also had polydactyly, but this was not a feature in the other patients. Some mild dysmorphic features, such as low-set ears, skin tags and webbed neck were noted. Death occurred in four of the patients before the age of one year, but one was surviving at two years of age. These families were both found to have homozygous mutations in *TMEM260*; c.1393C>T (p.Gln465*) and c.1698_1701del9 (p.Try567Thrfs*27) on WES.

On functional investigation, the authors found that both mutations map to regions of *TMEM260* that are present solely in the long isoform of *TMEM260* (another short isoform is also predicted). They assessed quantitative PCR of the long and short isoforms in patient and control lymphoblastoid cell lines and found the long isoform to be reduced in the patient samples compared to controls, whereas the short isoform was unaffected. They went on to assess translation-blocking morpholino oligonucleotides (MO) of *tmem260* in zebrafish embryos and noted that the MO-injected zebrafish had a smaller optic size and fewer intertectal neurons than uninjected controls. Rescue of the CNS abnormalities was shown with *TMEM260* long isoform mRNA and partial rescue was seen with the short isoform. c.1698_1701del *TMEM260* mRNA did not rescue the phenotype, supporting its pathogenicity. They also developed a CRISPR-cas9 zebrafish model, targeting deletions to the long isoform,

which demonstrated a reduction in the number of intertectal neurons and optic tecta size. However, the mutant zebrafish did not appear to have consistency with the cardiac defects identified; pericardial oedema was noted, but was not rescued, and may therefore have been a non-specific finding. A noted kidney filtration defect was also not rescued. These findings could also be due to cross-species differences.

Our patients displayed features consistent with the previously described phenotypes; they had Truncus Arteriosus, and low-set ears were also described. The first child of our family also died in the early neonatal period, at day 2, however the second is alive at the age of 3 and a half years. Our patients did not have the same renal or CNS abnormalities that were present in these described families. A review of the medical records of the second affected child in our family showed a creatinine level within the normal range at 29 μ M/L (normal range 15-31 μ M/L), although there was mention in the notes of a deterioration in renal function at 2 weeks of age. A renal ultrasound scan did show a degree of nephrocalcinosis bilaterally. This patient was born preterm, at 31 weeks gestation, which may have contributed to early renal deterioration. Our patients did not suffer from any known CNS abnormalities and cranial ultrasound scan in the second sibling at one month of age was normal. The variant we describe in *TMEM260* is a missense variant, where the other described variants are predicted to be protein truncating. Our variant does fall within exon 11, which is predicted to fall only within the long isoform of the protein, consistent with the previously described cases. This variant is predicted to be pathogenic by *in silico* pathogenicity predictors and is very rare in control populations (2 of 250452 alleles in the GnomAD database). Overall, we assess that this variant is likely to be the cause of disease in this family, but we cannot be certain without further functional assessment. This variant has been confirmed by Sanger sequencing, and reported back to the referring clinician. We have given the variant a class III, uncertain significance (Best Practice Guidance on Reporting of Unclassified Variants, <http://www.acgs.uk.com>).

There were other variants present in this family that could be linked to the phenotype. *C2ORF88* contained a homozygous frameshift mutation that was remaining following filtering in this family. This gene is also known as *smAKAP*. This is a ubiquitously expressed AKAP (protein kinase A anchoring protein). It is named *small membrane AKAP* as it localises to the plasma membrane. This molecule provides specificity for cAMP-dependent protein kinase (protein kinase A). *smAKAP* mRNA was detected at relatively high levels in human and murine heart samples (Burgers et al., 2012). The cAMP-PKA pathway plays an important role in cardiac development. smAKAP interacts

with PRKAR1A, a regulatory subunit of PKA (Rual et al., 2005) and has specificity for the R1 regulatory subunit of PKA over the RII subunit (Burgers et al., 2012). A mouse knockout of the R1 α regulatory subunit has disrupted development of mesoderm-derived structures and homozygous mutant embryos fail to develop a functional heart tube (Amieux et al., 2002). Furthermore, in the hearts of *Prkar1a* knock-out mice, there is a downregulation of cardiac transcription factors, such as Gata4 and Nkx2-5 (Yin and Kirschner, 2009). Patients with Carney complex, caused by mutations in *PRKAR1A*, have been reported to suffer from CHD (Stratakis et al., 2001). There is therefore some evidence that *smAKAP* may be a conceivable candidate gene for Truncus Arteriosus in this family.

In this family, there was also a variant in the gene *RREB1* remaining following filtering. This was a missense variant and, although *in silico* pathogenicity predictors did not support pathogenicity, there was a moderately high CADD score of 20.5. This gene encodes Ras-responsive element binding protein 1, a zinc finger transcription factor involved in histone modification. It appears to bind the distal Ras-responsive element in the calcitonin gene (Thiagalingam et al., 1996) and acts as a transcription modulator of several tumour suppressor genes. The *Drosophila* orthologue, *HNT*, is known to play important developmental roles including embryonic germ band retraction and retinal cell regulation (Ming et al., 2013). A recent study identified variants in *RREB1* in a whole exome sequencing study in individuals with 22q11.2 deletion syndrome. This study compared those patients with CHD (the majority of whom had conotruncal defects, such as Tetralogy of Fallot) to those without cardiac defects, in an attempt to identify genetic modifiers. This identified rare, deleterious SNPs in *RREB1* in 8 affected cases, but no control subjects. These variants were spread throughout the gene, but one was at an adjacent amino acid position to that identified in our study (p.Asp1229Gly). Other histone modifiers were also identified as important in CHD pathogenesis by this study, including *JMJD1C* and *KDM7A*. The authors propose that dysregulation of transcription via modification of histones plays an important role in CHD pathogenesis in 22q11.2 deletion syndrome (Guo et al., 2015). It is therefore possible that the variants identified in *RREB1* in our family are playing a role in the pathogenesis of Truncus Arteriosus.

Sanger sequencing confirmed the variants in both *smAKAP* and *RREB1* to segregate with disease in this family. It is difficult to assess the effect of these variants on the phenotype observed in this family. Due to the publication of other patients with similar phenotypes affected with mutations in *TMEM260*, this is the most likely disease

candidate. However, it may be that the variants in *smAKAP* and *RREB1* are playing a modifier role in the phenotype observed.

3.4.7.8 Family CHD8

Family CHD8 have a complex phenotype, with a number of children affected with CHD, but with a mix of abnormalities including VSD, dextrocardia and hypoplastic left heart syndrome. Due to the number of siblings affected it was assumed that a recessive genetic model was appropriate, however all of the variants that remained following filtering were present in a homozygous state in control databases, effectively ruling them out as the cause. In a case such as this, with differing phenotypes in the affected children, it may be that more complex genetics or even environmental factors are playing a role. It is also possible that WES is not identifying the causative genes due to non-penetrance in one of the parents or the presence of a mutation in a non-coding region.

3.4.7.9 Family CHD12

Family CHD12 consists of two siblings affected with secundum AVSD. The only variants remaining following filtering in family CHD12 were compound heterozygous variants in *FAM136A*. This protein is expressed in the inner ear where it has been shown to co-localize with COX4, a mitochondrial marker (Requena et al., 2015). Variants in this gene have previously been associated with Meniere's disease, a chronic disease that causes vertigo, hearing loss and tinnitus (Sajjadi and Paparella, 2008). There is no known link between this gene and cardiac disease.

3.4.7.10 Family CHD13

In family CHD13 there is one affected individual with a complex phenotype encompassing ASD/VSD, kidney abnormalities, retinal colobomas, and cerebellar vermis hypoplasia. The variants remaining following filtering in this family fell within three genes: *ZNF302*, *PLEC* and *RASGRF1*. *ZNF302* is a zinc finger protein of largely unknown function, it is known to be deleted in the heterozygous form in 19q13.11 deletion syndrome. This condition manifests as failure to thrive, microcephaly, developmental delay, feeding difficulties, ectodermal dysplasia and dysmorphic features including hypertelorism, low set ears and high-arched palate. Deafness, hand and foot abnormalities and Genito-Urinary (GU) abnormalities (such as hypospadias) have also been reported. *ZNF302* appears to belong to the minimal region of overlap for this syndrome (Gana et al., 2012), along with four other genes (*ZNF181*, *ZNF599*, *ZNF30* and *LOC400685*). There is a single report of ASD occurring in this syndrome (Chowdhury et al., 2014). However, this condition does not have significant overlap

with the phenotype seen in our family. Our family has horseshoe kidney and retinal colobomas, which are not described in 19q13.11 deletion syndrome. Furthermore, our family does not have ectodermal dysplasia, which is one of the key features of 19q13.11 deletion syndrome. It is also important to note that the described deletions at this locus are in the heterozygous state and so may not be informative in this case.

PLEC encodes plectin-1, a large filament-binding protein. It acts a cross-linking element in the cytoskeleton (Pytela and Wiche, 1980). Mutations in this gene have previously been associated with various forms of inherited Epidermolysis Bullosa Simplex (EBS), including EBS with muscular dystrophy (Smith et al., 1996) and EBS with Pyloric Atresia (Pfundner and Uitto, 2005), due to truncating or nonsense mutations in the gene causing a loss of protein (Pulkkinen et al., 1996).

RASGRF1 encodes Ras protein-specific guanine nucleotide releasing factor. RAS proteins regulate many signaling pathways in cell growth and differentiation. RAS activation requires the release of GDP. The product of the *RASGRF1* gene is known to augment this GDP release (Schweighoffer et al., 1993). Its expression is mainly at neuronal synapses and decreased expression has been identified in the brains of patients with intractable temporal lobe epilepsy (Zhu et al., 2013), although expression at the pancreatic islet cells has also been shown (Font de Mora et al., 2003). *RASGRF1* knockout mice exhibit reduced postnatal growth, low insulin and IGF I levels, glucose intolerance, impaired synaptic plasticity, and impaired hippocampal-dependent learning (Font de Mora et al., 2003, Giese et al., 2001, Brambilla et al., 1997). There is no known cardiac disease in these mice. Polymorphisms in this gene have previously been associated with myopia (Chen et al., 2015).

3.4.7.11 Family CHD15

In family CHD15, where a single pregnancy was affected with AVSD and ventriculomegaly, the variant with the highest CADD score was in *TBCA*. This was a rare variant, present in a large region of homozygosity. This gene encodes Tubulin-Specific Chaperone A. This is a molecular chaperone of beta-tubulin (*TUBB*), a key component of microtubules (Nolasco et al., 2005). *TBCA* knockdown with siRNA in two cell lines decreased soluble tubulin, created microtubule disruption and caused cell cycle arrest (Nolasco et al., 2005). Mutations in *TUBB* and other tubulins including *TUBBA1A*, *TUBB2A*, *TUBB2B*, *TUBB3* and *TUBG1* are associated with structural brain malformations, for example cortical dysplasia (Breuss et al., 2012), usually as a *de novo* occurrence, termed 'tubulinopathies'. It is possible that a mutation in a tubulin

gene could manifest as ventriculomegaly on USS (Fallet-Bianco et al., 2014), however there is no previous report of cardiac defects in the tubulinopathies. It is notable that *TBCA* was highly expressed in foetal heart tissue, making this the most likely candidate variant in this family.

3.4.7.12 Family CHD16

Family CHD16 have a child affected with Tetralogy of Fallot, renal agenesis, deafness, talipes, hypospadias and developmental delay. They have a further pregnancy affected with Tetralogy of Fallot. A variant in *RBM48* was the variant with the highest CADD score remaining following filtering. This is an RNA binding motif protein, containing one RNA recognition motif. There is very little previous research on the function of this protein. A missense variant in this gene was previously associated with disease in a patient with nephronophthisis, a renal ciliopathy, and Jeune asphyxiating thoracic dystrophy, a skeletal dysplasia (Braun et al., 2016). These diagnoses are not consistent with the phenotype seen in our patient. None of the other variants that remained following filtering in this family were considered as likely candidates for the phenotype.

3.4.7.13 Overlapping Findings in CHD Patients

As was suspected from the range of phenotypes of families recruited, there was very little overlap of candidate genes in this cohort. However, variants in genes in the same functional or developmental pathways were seen repeatedly in the cohort. The most prominently represented group was that of zinc finger proteins, including *ZFPM1*, *RREB1*, *ZNF302*, *ZMYM6*, *ZBTB46* and *ZNF441*. These proteins contain motifs that involve zinc ions for stabilization. They have diverse functions which include transcriptional regulation, RNA binding, protein folding and assembly and regulation of apoptosis (Laity et al., 2001). Zinc finger proteins are one of the most abundant forms of protein in the eukaryotic cell, which may account for their high representation in this dataset (Laity et al., 2001). However, zinc finger transcription factors are known to play key roles in cardiac development, for example *GATA 4*, *5* and *6* (McCulley and Black, 2012), mutations in which are associated with CHD. All three are expressed at high levels in the developing heart and are known to regulate expression of cardiac-specific genes (Pikkarainen et al., 2004). It may be, therefore, that the high incidence of zinc finger proteins and other transcription factors in our dataset is due to the role of these proteins in cardiac development. The candidate variants identified in these CHD families have been uploaded to the Matchmaker Exchange website (<http://www.matchmakerexchange.org>), via Decipher (<https://decipher.sanger.ac.uk>) in

an attempt to identify other families with similar phenotypes and variants in the same genes.

3.5 Patients with Laterality Disorders from the DDD Project

3.5.1 Introduction

The Deciphering Developmental Disorders (DDD) project is a research study based at the Sanger Institute (Cambridge, UK) investigating the cause of developmental disorders in children and adults using array CGH analysis and WES in a trio-based approach. Access was granted through a DDD-Complimentary Analysis Project (DDD-CAP) to access data from patients recruited to the project. These patients were those with 'laterality disorders' as classified by the following Human Phenotype Ontology (HPO) terms: abnormal anatomic location of the heart, abnormal special orientation of the cardiac segments, abnormality of abdominal situs and abnormality of pulmonary situs, plus all derivative terms.

The data was received in the form of annotated .vcf files for the affected patient and both parents. .vcf files from each family were initially merged and then data analysis carried out using the standardized pipeline, commencing at the filtering of variants by comparison with databases of known population variants.

3.5.2 Results

3.5.2.1 Clinical Summary

In total, data from 11 patients coded as 'laterality disorder' was received from the DDD project, as well as their parents. Phenotypic information was available, as provided to the study, from the patient's clinician. This included information such as classification of anomalies present, birth history, growth parameters, dysmorphic features and family history. Table 3-9 shows a summary of clinical information for all 11 patients.

3.5.2.2 Identification of Known Disease-Causing Genes

In four patients from the project, mutations in known disease-causing genes, associated with laterality, were identified. Where appropriate these findings have been reported back to the referring clinician. Table 3-10 shows a summary of these mutations.

Table 3-9: Clinical features of patients from the DDD project coded as 'laterality disorder'.

Pf = palpebral fissures , PS = Pulmonary Stenosis, ASD = Atrial Septal Defect, DORV = Double Outlet Right Ventricle, IVC = Inferior Vena Cava

Patient number	Dextrocardia	Abdominal situs	Retinopathy	Brain abnormality	Upper limb abnormality	Renal anomaly	Liver disease	CHD	Dysmorphic features
DDD1		+	Pigmentary retinopathy	Molar Tooth sign. Polymicrogyria.	Post axial polydactyly.				
DDD2	+				Absent radius. Short thumb.				
DDD3	+	+			Adducted thumb	Renal dysplasia	Liver failure	+	
DDD4	+	+							Synophrys, high arched palate
DDD5	+	+			Radial hypoplasia.			ASD, PS, DORV.	
DDD6	+				Absent thumb				Hypertelorism. Upslanted pfs
DDD7	+							+	
DDD8	Mesocardia.								Bilateral ptosis.
DDD9	+	+							Preauricular skin tag.
DDD10	+							PS	Bilateral ptosis
DDD11	+	+						Right aortic arch. Interrupted aortic arch, VSD. Interrupted IVC.	

Table 3-10: Mutations in known disease-causing genes identified in patients from the DDD project.

Patient no.	Gene	Variant	Protein change	Polyphen2	SIFT	CADD	MAF GnomAD	Known disease
DDD3	<i>NEK8</i>	17:27064442G>A	R246Q; NP_835464	Probably damaging (1.0)	Deleterious (0)	35	5/281560	NEK8- associated ciliopathy
		17:27064957T>G.	V337G; NP_835464	Probably damaging (0.997)	Tolerated (0.26)	24	-	
DDD4	<i>DNAH5</i>	5:13809290A>G	W2539R; NP_001360	Possibly damaging (0.608)	Deleterious(0)	24.8	-	Primary Ciliary Dyskinesia
DDD10	<i>MMP21</i>	10:127456129TTC>T	p.K461VfsTer14; NP_671724	-	-	-	2/ 252118	Published as a new laterality gene in 2016 (Akawi et al., 2015)
		10:127460912A>G	p.I285T; NP_671724	Probably damaging (0.999)	Deleterious(0)		17/282466	
DDD11	<i>MMP21</i>	10:127460819C>T	W316X; NP_671724	-	-	37	2/252418	
		10:127460919G>A	p.H283Y; NP_671724	Probably damaging (0.994)	Deleterious(0)	28.4	-	

3.5.2.3 Candidate Laterality Variants

As well as filtering on each sample individually, data was compared across all samples to identify variants falling within the same gene, in an attempt to identify novel disease candidates. This was done using filtered biallelic and *de novo* candidate gene lists, comparing only variants which passed hard filters and had a CADD score >15. None of these variants reoccurred in the remaining seven samples.

3.5.3 Discussion of Findings from the DDD Project

In 4 of the 11 patients analyzed, putative mutations were identified in known disease causing genes. Patient DDD3 was affected with situs inversus with polysplenia, but also had renal dysplasia, liver failure, complex CHD and hypertrophic cardiomyopathy. Compound heterozygous variants in *NEK8* were identified, which were in *trans*. These variants had high CADD scores and were predicted as deleterious by *in silico* tools. They were also rare in control datasets. *NEK8* (never in mitosis A related kinase 8) causes a ciliopathy that is associated with liver disease (congenital hepatic fibrosis, bile duct hypoplasia or ductal plate malformation), enlarged cystic dysplastic kidneys, and congenital heart malformations (Rajagopalan et al., 2016). Hypertrophic cardiomyopathy, situs inversus (Frank et al., 2013) and renal dysplasia (Rajagopalan et al., 2016) have also been reported. Skeletal defects and brain malformations can also be associated with this condition (Grampa et al., 2016). This diagnosis would be consistent with the features identified in this patient. The *NEK8* gene is a protein kinase, found to be critical for ciliary function and important in cell cycle control (Moniz et al., 2011). It is located in the 'inversin compartment' of the cilium, a proximal part of the axoneme (Grampa et al., 2016). A mouse knock-out model has cystic kidneys and heart-looping defects and a similar phenotype is seen in zebrafish (Liu et al., 2002). The mutations identified here have not previously been reported, but V337G falls in the Regulation of Chromatin Condensation 1 (RCC1) domain of the protein, where previous disease-causing missense mutations have been identified. This domain is critical for *NEK8* localization to cilia and is thought to be involved in centrosome recruitment (Zalli et al., 2012). R246Q falls in the serine/threonine catalytic domain, which we would predict to affect the protein kinase function (Zalli et al., 2012). Missense mutations have previously been identified as disease-causing in this condition and some are reported to have a gain-of-function effect, typically causing renal dysplasia, rather than enlarged cystic kidneys (Grampa et al., 2016). This would be consistent with the phenotype observed in our patient.

Patient DDD4 was affected with a clinical diagnosis of Primary Ciliary Dyskinesia (PCD), but also had the unexplained findings of learning disability and dysmorphic features. A homozygous missense mutation in *DNAH5* was identified. This variant scored highly on pathogenicity prediction tools. It is not a previously published mutation, but falls in a region of the protein that is commonly mutated in PCD. This finding does not explain the additional features of learning difficulties and dysmorphic features identified in this patient and so its pathogenicity is not certain.

Patients DDD10 and DDD11 were identified to have mutations in *MMP21*. Following identification of the mutations in these patients, we recognised that these exact families have already been published in the literature (Akawi et al., 2015). *MMP21* encodes a matrix metalloproteinase that is reported to modulate cell migration through effects on the extra cellular matrix (Ahokas et al., 2002). The authors modeled these identified variants in the structure of *MMP21* and showed them to be close to a conserved zinc-binding site and they were predicted to have a strong impact on enzymatic function (Akawi et al., 2015). Two mouse models with missense mutations close to the zinc-binding domain have heterotaxy and complex CHD. These mice had normal ciliary motility at the embryonic node, suggesting *MMP21* acts downstream of the ciliary beat in the formation of laterality (Akawi et al., 2015). Mutations in *MMP21* are suspected to be a relatively common cause of heterotaxy in humans (Guimier et al., 2015).

In this study we did not identify further overlapping causes of heterotaxy in the remaining samples from the DDD project without a mutation in a known gene. It may be that the genetic causes in these families are different, as there is phenotypic variability between the families in the cohort. It may also be that some cases of heterotaxy are caused by mutations in the non-coding parts of the genome, or that environmental factors play a role, as is already known for congenital heart disease.

3.6 Conclusions

We have shown that in families with a recurrence of CHD, WES can identify putative pathogenic mutations. We have described one family with mutations in *PIGV* causing Hyperphosphatasia and Mental Retardation Syndrome (HPMRS), with the novel features of absence of the bladder and congenital diaphragmatic hernia. We have described 10 families with no mutation in a known disease-causing gene, but with putative candidate variants identified by this approach. In 11 patients from the DDD project with 'laterality disorders' we have recognized four with mutations in known disease-causing genes, but no further overlapping candidates.

4 *PPA2* Mutations Cause Early Onset Cardiomyopathy and Sudden Infant Death

4.1 Introduction

This chapter describes the investigation of a family, CHD2, affected with sudden unexpected death in childhood associated with cardiomyopathy on post mortem examination. Using WES we were able to identify a likely causative mutation in a novel gene, *PPA2*. Collaboration allowed for the identification of three other families affected with mutations in the same gene and allowed us to delineate the disease phenotype. Functional analysis was undertaken to support the pathogenicity of these mutations and to give insight into the function of this important protein.

Of the work described in this chapter, Dr Hartill was directly involved in the recruitment of family CHD2 to the study, including taking informed consent, clinical phenotyping and DNA extraction from patient samples, as well as preparation of whole exome libraries, analysis of WES data and bioinformatic analysis of variants identified in family CHD2, identifying *PPA2* as the likely cause of disease. Dr Hartill made initial contact and formed an ongoing collaboration with an international research team, including colleagues in Newcastle-Upon-Tyne (Prof. Robert Taylor) who identified further families affected by mutations in *PPA2*. The functional work described in this chapter including pyrophosphatase activity assays described in the results section and western blot and oxygen consumption assays described in the discussion section were performed in full by international collaborators in Austria (Ms Lavinija Mataković and Dr Johannes Mayr and colleagues, Department of Pediatrics, Paracelsus Medical University Salzburg, Salzburg, Austria).

4.2 Results

4.2.1 Clinical Phenotype

Family CHD2 is a multiply consanguineous family of Pakistani origin, living in Leeds, UK. The parents are first cousins (see Figure 4-1).

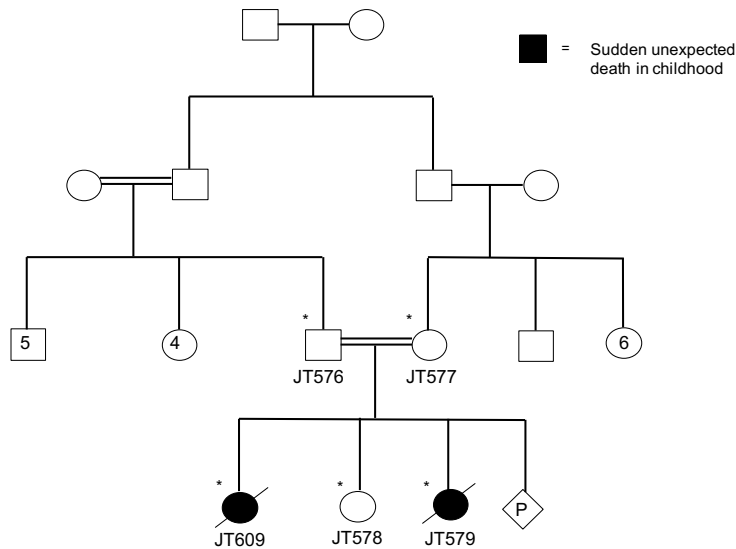


Figure 4-1: Pedigree of family CHD2.

The first child in this family (JT609) was well with normal growth and development until the age of 5 ½ months. She became unwell following a 24 hour history of a vomiting and diarrheal illness and suffered a seizure at home. She had further seizures on arrival at hospital and was intubated and ventilated. A Computerised Tomography (CT) head scan was normal. Echocardiogram showed poor contractility and a small amount of tricuspid regurgitation. She was transferred to a teaching hospital, although on transfer she unfortunately suffered multiple cardiac arrests, and she eventually died during the course of these.

On post mortem examination, the brain showed mild hypoxic injury. Skeletal survey, metabolic and toxicology screens were all normal. Fatty acid oxidation levels were performed on skin fibroblasts and were normal. Rotavirus was identified from stool samples taken on admission. Tests for myocarditis were normal. Cerebrospinal fluid (CSF) glucose was low, but this was undertaken on a post mortem sample. The heart appeared normal in size, shape and macroscopic structure. Cardiac histology showed areas of recent necrosis, and there was also evidence of long-standing myocyte loss with increased interstitial collagen and focal myocyte fibre disarray in the left ventricle and interventricular septum. This was described as a cardiomyopathy and the cause of death was described as sudden infant death due to a cardiomyopathy of unknown cause.

Their second child (JT578) is fit and well at 4 years of age. A recent echocardiogram was normal, and tests for lactate, acylcarnitine and Creatinine Kinase (CK) were all within the normal range.

The third child in this family (JT579) became acutely unwell at the age of 8 months. She had suffered symptoms of a viral illness and a week's history of increasing hypotonia and weakness. At this time, her CK was raised at 15,000 IU/L (normal range <250IU/L) and her plasma lactate was raised at 5mmol/L (normal range 0.5-2.2 mmol/L). Free carnitine was within the normal range at 43.6 μ mol/L (normal range 30-60 μ mol/L), but propionylcarnitine was raised at 2.38 μ mol/L (normal range <1.5 μ mol/L). Her urine organic acids and renal function were normal. Her weakness and hypotonia improved over the next two weeks and her CK level reduced. Echocardiogram at this time was normal.

She presented again at the age of 11 months with a further viral illness causing diarrhoea and vomiting. She became increasingly drowsy. An initial blood gas demonstrated a lactic acidosis. Following admission to her local hospital she suffered focal seizures, followed by a generalized seizure. CT head scan at this time was normal. She then suffered a cardiac arrest and was intubated and ventilated. Maximal cardiopulmonary resuscitation was continued for 20 minutes, but was not successful.

Post mortem examination showed extensive fibrosis of the heart muscle, in a similar distribution to that seen in her sister. There were no other macroscopic observations at post mortem. Norovirus infection was confirmed on stool samples from admission.

Mitochondrial disease was suspected and investigations were performed at Newcastle Mitochondrial Disease Service. These included *DGUOK* and *ACAD9* gene sequencing, which did not demonstrate a pathogenic mutation. Respiratory chain analysis on peripheral muscle tissue was normal, and histological and histochemical assessment of muscle biopsy did not reveal any major mitochondrial abnormalities. At the time of recruitment to the study, the mother was pregnant with the family's fourth child.

4.2.2 Whole Exome Sequencing

WES was undertaken in five family members from this family: JT576, JT577, JT578, JT579 and JT609 (see Fig. 4-1). The samples were prepared using the Illumina QXT methodology and Agilent v5 library kit. Samples were pooled with 18 other patient samples and run across 4 lanes on an Illumina HiSeq 2500 instrument.

Data analysis was performed as described previously (see Chapter 2, section 2.2.14.3). For this family, variants present at >1% minor allele frequency were removed

after comparison with the following databases: dbSNP 138 and previous, NHLBI Exome Sequencing Project (ESP) Exome Variant Server, the Exome Aggregation Consortium (ExAC), 300 local ethnically-matched exomes and over 3000 ethnically-matched control samples from the Born in Bradford project (mothers of babies recruited to the project) (<http://www.borninbradford.nhs.uk>). Variants were retained that were predicted to have a functional effect on the protein using Ensembl's Variant Effect Predictor (VEP), and these included missense, splice site variants and indels. Variants were retained which followed the predicted autosomal recessive inheritance pattern in the family i.e. were biallelic or homozygous in the affected siblings, but not in the unaffected sibling and heterozygous in the parents. Table 4-1 shows the number of variants remaining following each filtering step in the pipeline. Variants were manually inspected and retained if predicted 'pathogenic' by any one of Polyphen2, SIFT or Condel. Finally, variants were ordered by CADD score and those with CADD score >15 were retained. Data quality statistics were performed on WES data from this family and are shown in appendix F.

Table 4-1: Number of variants remaining at each stage of variant filtration in the analysis of family CHD2.

Stage in Analysis	Number of variants
Called with GATK	459205
Filter on dbsnp <1% MAF	123881
EVS <1% MAF	120698
BIB <1% MAF	108087
Exac <1% MAF	103935
Local exomes <1%	88749
Functional variants	1035
Biallelic shared by affected sibs	14
Pathogenic on <i>in silico</i> tools	6 (all homozygous)
CADD score >15	3

Following filtering, three variants remained (see Table 4-2). The variant in *PPA2* had the highest CADD score of 32. Figure 4-2 shows an IGV read out showing alignment of the sequencing reads in the 5 family members and the presence of the *PPA2* variant.

Table 4-2: Variants remaining following filtering WES data using the data analysis pipeline in family CHD2.

Gene	Variant	Protein change	Polyphen2	SIFT	Condel	MAF (GnomAD)	CADD	Known protein function
<i>PPA2</i>	4:106359135 G>A, c.500C>T; NM_176869.2	P167L; NP_789845	Probably damaging (1)	Deleterious (0)	Deleterious (0.945)	4/240938	32	PPases catalyze the hydrolysis of pyrophosphate to inorganic phosphate, localized to mitochondria.
<i>EPHA5</i>	4:66356221 C>T	V426I; NP_004430	Probably damaging (0.97)	Deleterious (0.04)	Deleterious (0.761)	206/245944	30	Ephrin receptor subfamily of the protein-tyrosine kinase family.
<i>RALGAPA1</i>	14:36190939 G>T	P741T; NP_001333175	Probably damaging (0.99)	Deleterious (0.04)	Deleterious (0.839)	Not present	26.4	Ral GTPase activator activity

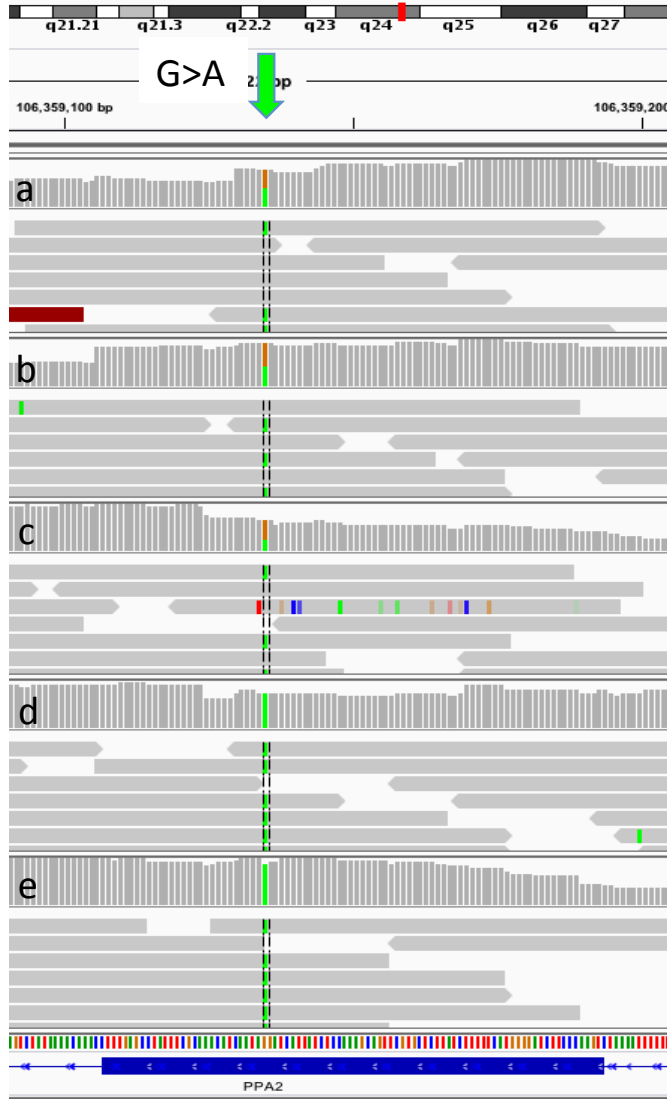


Figure 4-2: IGV read-out showing sequencing reads aligned to the reference genome (GRCh37) in family CHD2.

The variant in *PPA2*, c.500C>T, is highlighted by the green arrow, shown as G>A as present on the reverse strand. The variant is shown to be heterozygous in the parents and unaffected sibling (a: JT576, b: JT577 & c: JT578) and homozygous in affected patients (d: JT579, e: JT609).

4.2.3 Autozygosity Mapping

Autozygosity mapping was performed on WES data. Regions of more than 100000bp and containing at least 25 SNPs were identified which were shared by the affected family members, but not shared with the unaffected individuals. Table 4-3 shows the large homozygous regions that fit these criteria. The variant in *PPA2* falls within the largest homozygous region on chromosome 4.

Table 4-3: Output of SNP Viewer program showing large regions of homozygosity shared between affected individuals and not with unaffected individuals in family CHD2.

Chromosome position	Size of region (bp)
Chr4:89053656-114310106	25,256,450
Chr4:115586972-129770413	14,183,441
Chr4:66218728-77677570	11,458,842
Chr14:36334905-45494766	9,159,861
Chr4:55598202-64092721	8,494,519
Chr2:80846161-85133314	4,287,153
Chr14:31542391-35783948	4,241,557
Chr7:51327575-55238087	3,910,512
Chr2:98443531-101439243	2,995,712
Chr10:114887002-117856379	2,969,377
Chr10:86273249-88817578	2,544,329

4.2.4 CNV analysis

CNV analysis was performed using the Exome Depth software on WES data from individual JT579. A number of potential candidate CNVs were identified by the program, but a number were present in control datasets and none overlapped a likely candidate gene for cardiomyopathy. A full list of CNV variants identified is provided in appendix J.

4.2.5 Bioinformatic Analysis of the *PPA2* variant

The *PPA2* gene codes for an inorganic pyrophosphatase, an enzyme that catalyses the hydrolysis of pyrophosphate to inorganic phosphate. In humans, there are two forms of pyrophosphatase: PPA1 is cytosolic and present in the majority of cell types and PPA2 is specific to mitochondria. Following review of the literature, and due to the suspicion of mitochondrial disease in this family, this gene appeared a good candidate for the phenotype observed. Further bioinformatics analysis was performed to reinforce this hypothesis. The variant was assessed using *in silico* pathogenicity prediction tools. It was rated as 'probably damaging' by Polyphen2 (1), 'deleterious' by SIFT (0) and 'deleterious' by Condel (0.945) (see Table 4-2).

Investigation of this variant, c.500C>T, in publically available control databases showed it to be extremely rare. It was not present in 3000 ethnically-matched in-house control samples and was present as four heterozygous calls in 240938 alleles in the GnomAD database (all of South Asian origin), giving a MAF of 0.0000166. It was present in

dbSNP142 with a MAF of 0.0002. The variant falls in exon 6 of the gene. This affects the canonical transcript of the protein (ENST00000341695) at a highly conserved position conserved in mammalian species through to *Xenopus* (see Fig. 4-3).

P167L
↓

Homo sapiens	NP_789845.1	154	EKDKSTNCFGDNDPIDVCEIGSKILSCGEVIHVKILGILALIDEGETDWK	203
Pan troglodytes	XP_517378.2	154	EKDKSTNCFGDNDPIDVCEIGSKILSCGEVIHVKILGILALIDEGETDWK	203
Macaca mulatta	XP_001082969.2	154	EKDKSTNCFGDNDPIDVCEIGSKILSCGEVIHVKILGILALIDEGETDWK	203
Canis lupus familiaris	XP_535679.3	163	RKDKSTDCCGDNDPIDVCEIGSKVLSRGEVIVKILGILALIDQGETDWK	212
Bos taurus	NP_001069864.1	147	RKDKSTDCCGDNDPIDVCEIGSKVLSRGEVIVHVKILGVLALIDQGETDWK	196
Mus musculus	NP_666253.1	149	LRDKSTDCCGDNDPIDVCEIGSKVLSRGEVIVHVKILGTLALIDQSETDWK	198
Rattus norvegicus	NP_001129343.1	149	LKDKSTNCCGDNDPIDVCEIGSKVLSRGEVIVHVKILGTLALIDQSETDWK	198
Gallus gallus	XP_420502.3	298	HTDNITGCCGDNDPVDVCEIGSKVRSSEIVQVKVLGVLALLDEGETDWK	347
Danio rerio	NP_957027.1	107	HRDGDGTCCGDNDPIDICDIGSEVCSRGQVIQVKVLGTLALIDEGETDWK	156
Xenopus tropicalis	NP_001017130.1	123	HVDNDTKCCGDNDPIDVCEIGSKVLCARGDVIQVKVLGILALVDEGETDWK	172

Figure 4-3: Evolutionary conservation of p.P167L in vertebrates.

Output adapted from HomoloGene (<https://www.ncbi.nlm.nih.gov/homologene>).

There was no recorded knockout mouse model of *PPA2* at the time of investigation. Interrogation of the String database (<http://string-db.org>) showed interaction of *PPA2* with a number of ATPases and ATP synthases, for example *ATP5A1* (Fig. 4-4).

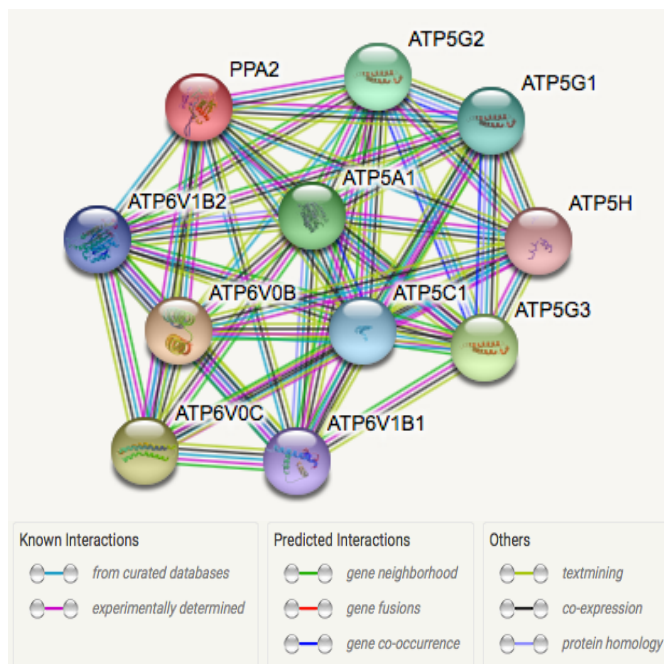


Figure 4-4: Output from the String database to summarize a query of interacting partners of *PPA2*.

Known interactions are demonstrated by light blue lines (curated databases) and pink lines (experimentally determined). Blue lines indicate predicted interactions due to gene co-occurrence, green due to similar gene neighbourhood and red due to gene fusions. Yellow associations are from text-mining, black associations from co-expression experiments and purple associations from protein homology.

4.2.6 Sanger Sequencing

Sanger sequencing was performed by the Yorkshire Molecular Genetics Diagnostic Laboratory and confirmed the presence of the mutation in *PPA2* in the family, being homozygous in the affected individuals and heterozygous in the parents and unaffected sibling. Results were reported back to the referring clinician.

4.2.7 Identification of Further Families with Mutations in *PPA2*

Having identified a likely candidate gene in family CHD2, we wished to identify further families with a similar phenotype who may have mutations in the same gene. For this, we formed a collaboration with Prof. Robert Taylor, head of the Rare Mitochondrial Disease Service for Adults and Children, Newcastle-upon-Tyne. Through this collaboration we were able to identify three further families with homozygous or compound heterozygous mutations in *PPA2* (here referenced as families 2b, 2c and 2d). WES was performed at a number of other international centers to identify these mutations in *PPA2*. Pedigrees of families 2b, 2c & 2d are shown in Figure 4-5.

We were also able to obtain, with consent, a skeletal muscle biopsy sample and skin fibroblast sample from patient JT579 (here also labeled P9), which had both been stored following post-mortem examination. The skeletal muscle sample was sent immediately to collaborators for functional assays as described below. The fibroblast sample was cultured in the Leeds laboratory, RNA was extracted and cells were frozen down prior to transfer to collaborators for functional analysis.

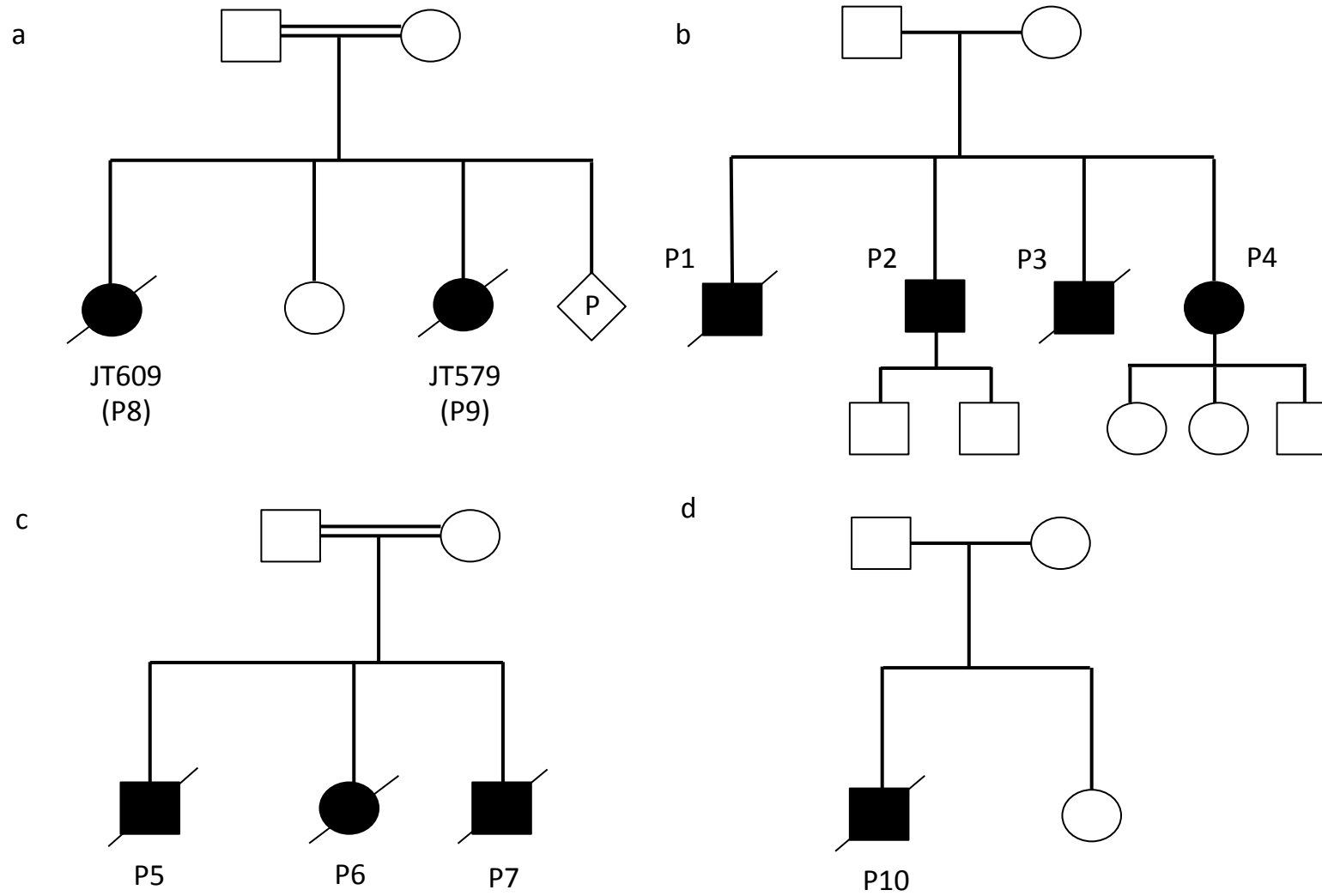


Figure 4-5: Pedigrees of families with *PPA2* mutations.

Family CHD2, from Leeds, is shown in panel a, notations b-d correspond to families 2b-d. Shaded circles/squares indicate affected individuals.

Family 2b is of Caucasian origin, living in New Zealand. The parents are unrelated. All four of their children exhibit extreme sensitivity to alcohol. This manifests as pallor and severe arm and chest pain following consumption of very small amounts of alcohol, for example having consumed less than 0.1g of alcohol in a cough medicine. Two of the siblings died in their teenage years following the consumption of small amounts (<1 unit) of alcohol. Sibling 1 died suddenly at the age of 15 years. Post mortem showed dilatation of both cardiac ventricles, with histology showing focal inflammation of neutrophils, lymphocytes and eosinophils. Skeletal muscle histology was normal. A diagnosis of myocarditis and sudden arrhythmic cardiac death was made.

Sibling 3 also died following consumption of <10g alcohol at the age of 20 years. Post mortem examination revealed a dilated left ventricle, with a distinctive circumferential scarring in the midmyocardium with focal sub-endocardial scarring. Microscopic examination of cardiac tissue revealed widespread, mature scarring of mid-myocardium (see appendix K).

Due to the family history, siblings 2 and 4 were assessed by cardiac MRI examination and both were found to have mid-myocardial fibrosis. On medical alcohol challenge, both suffered severe chest pain. Both siblings have now been fitted with an Implantable cardioverter defibrillator (ICD) device, although neither has had an event to date.

Family 2c is a family of Sri Lankan origin, living in Switzerland. The parents are first cousins. Family 2d are a non-consanguineous family of Northern Irish origin. Both families 2c and 2d are affected with a similar phenotype to that seen in family CHD2. They have had children affected with sudden, unexpected death between the ages of 11 days and 2 years. In the majority of cases a viral illness preceded deterioration and death. In both of these additional families, a metabolic cause was suspected with a history of lactic acidosis, seizures, hypotonia, and cardiac arrhythmia. Post mortem consistently demonstrated cardiac fibrosis and myocyte loss or disarray. In the majority of assessed cases, respiratory chain analysis was normal in fibroblasts and skeletal muscle. In P10, there was a slight reduction in complex I and IV in cardiac muscle.

4.2.8 Mutation Analysis

In total, four separate mutations in *PPA2* were identified across the four affected families (see Table 4-4).

Table 4-4: Mutations in *PPA2* identified in families 2b-d alongside *in silico* pathogenicity prediction scores and MAF observed in the GnomAD database.

Family	Mutation in <i>PPA2</i> ; c.DNA	Mutation in <i>PPA2</i> ; protein	Homozygous/heterozygous	Polyphen2	SIFT	GnomAD, heterozygotes (total alleles)
2b	c.514G>A	E172K	Heterozygous	Probably damaging (0.996)	Damaging (0)	140 (269750)
	c.683C>T	P228L	Heterozygous	Probably damaging (1.0)	Damaging (0)	57 (268654)
2c	c.500C>T	P167L	Homozygous	Probably damaging (1)	Deleterious (0)	4 (240938)
2d	c.380G>T	R127L	Heterozygous	Probably damaging (0.993)	Damaging (0)	48 (270540)
	c.514G>A	E172K	Heterozygous	Probably damaging (0.996)	Damaging (0)	140 (269750)

The identified mutations in *PPA2* were all scored as ‘pathogenic’ by *in silico* tools (see Table 4-4) and occurred at highly conserved positions of the protein. These variants were further assessed as to their frequency in control populations, for example the EXaC (<http://exac.broadinstitute.org>) and GnomAD databases (<http://gnomad.broadinstitute.org>). All were present at low levels, with no homozygotes present in control datasets. Table 4-4 shows the frequency of these variants in the GnomAD database.

Structural modeling of these variants was possible due to the similarities between *PPA2* and the yeast homologue IPP1. Three of the mutations were found to be located at the predicted enzyme active site (514G>A, 500C>T and 380G>T), so were predicted to disrupt the enzymatic function. The 683C>T variant was located on the outer surface of the molecule, and was predicted to disrupt dimerization due to its position at the dimerization boundary (see Figure 4-6).

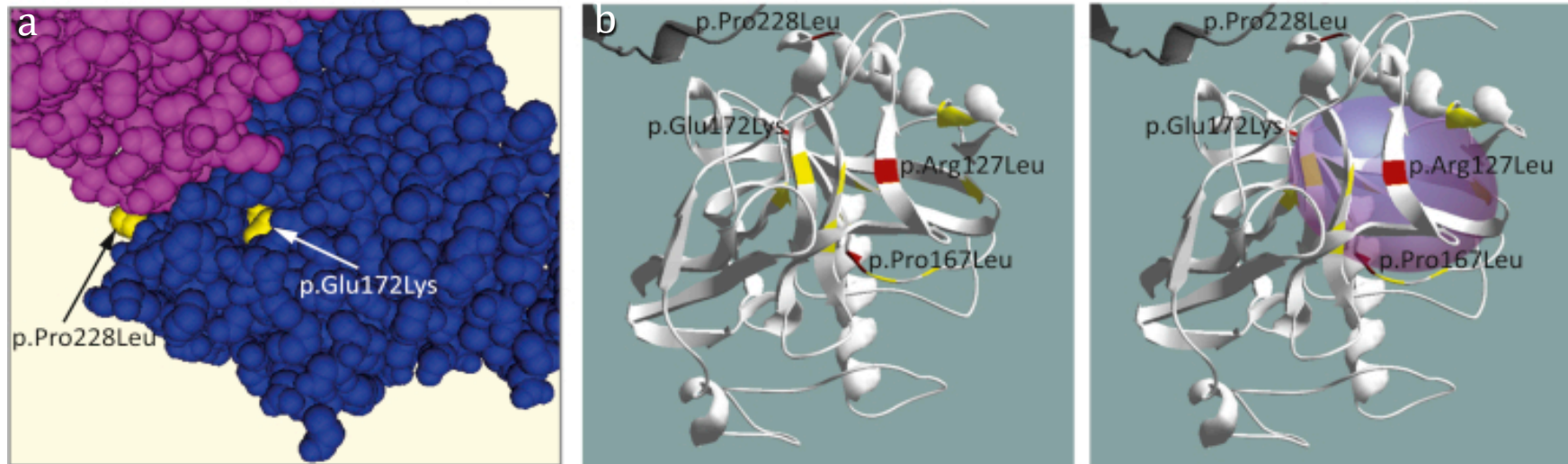


Figure 4-6: Structural modeling of PPA2 variants.

(a) p.Pro228Leu is shown to be located at the surface of the dimerization boundary. Image created using CN3D with reference PDB: 1M38. (b) p.Glu172Lys, p.Arg127Leu and p.Pro167Leu are found at the enzyme active site, modeled by the purple sphere. Image created using Swiss-PdbViewer with reference PDB: 1M38. Adapted from (Kennedy et al., 2016).

4.2.9 Pyrophosphatase Activity

Pyrophosphatase activity experiments were undertaken by the collaborating center in Austria, including wet laboratory work, quantification, statistical analysis, interpretation and confirmation. Pyrophosphatase activity was measured by determining formation of inorganic phosphate using a malachite green colorimetric assay (Van Veldhoven and Mannaerts, 1987). This was performed using fibroblast-extracted mitochondria from three patient samples and 15 control samples. Activity was significantly reduced in patients P5, P7 and P9 (see Fig. 4-5) when compared to controls, at various different $MgCl_2$ and PPI concentrations (see Figure 4-7). Inactivation with $CaCl_2$ was similar in affected and control individuals (Fig. 4-7).

Wildtype PPA2 c.DNA was cloned into expression vector pRSET B by the collaborating center. The cloned construct omitted the first 96 nucleotides of the protein, which corresponds to the N-terminal mitochondrial targeting sequence. The pathogenic variants identified in families CHD2 and 2b, c and d (c.500C>T (p.Pro167Leu), c.514G>A (p.Glu172Lys), and c.683C>T (p.Pro228Leu)) were introduced using site directed mutagenesis. These recombinant proteins were expressed in *E. coli*. Equal amounts of wildtype and mutant proteins, measured using Western Blot, were used in a Pyrophosphatase assay. Both the c.500C>T (p.Pro167Leu) and c.514G>A (p.Glu172Lys) variants had a 5-10% activity compared to the wildtype enzyme at PPI concentrations 18–500 $\mu\text{mol/L}$. The c.683C>T (p.Pro228Leu) variant showed a 24-28% activity in the same PPI concentration range (Fig. 4-7).

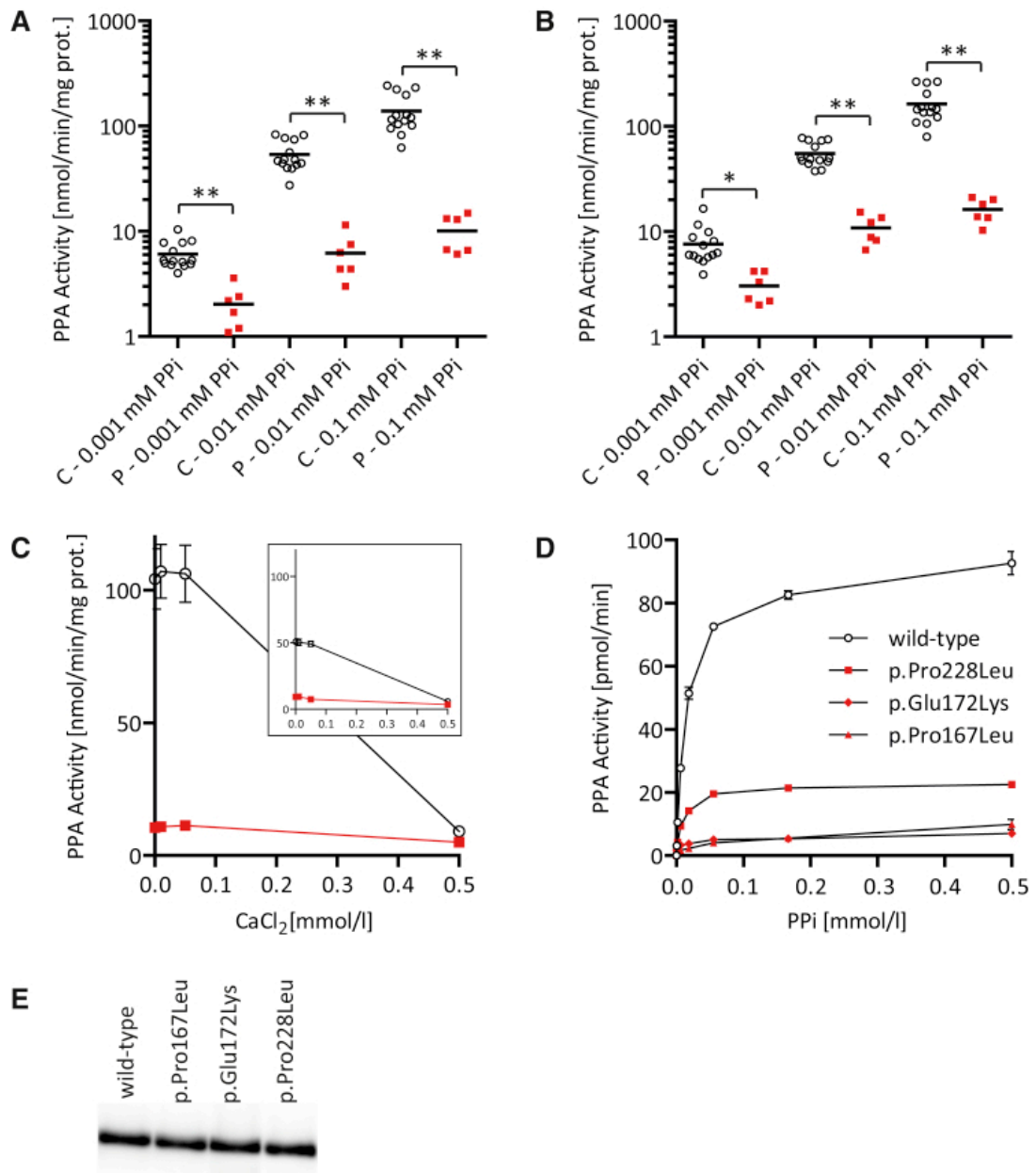


Figure 4-7: Pyrophosphatase activity in control and patient fibroblasts and recombinant proteins at different PPI concentrations.

A & B: Pyrophosphatase activity in patient (P) and control (C) fibroblast mitochondria isolations at 0.0001-0.1mM PPI concentrations at either (A) 0.5 mmol/L MgCl₂ or (B) 3.0 mmol/L MgCl₂. C: Pyrophosphatase activity in three control subjects (black line) compared to patient P5 (red line) at different CaCl₂ concentrations and 0.5 mmol/L MgCl₂. The larger graph shows a PPI concentration of 0.1mmol/l, with the small insert showing a PPI concentration of 0.01mmol/l. D: PPA activity of equal amounts of recombinant proteins at different PPI concentrations. E: Western blot to show addition amounts of recombinant proteins. Statistics shown are a Student's unpaired T-test. The error bars indicate the standard error of the mean * = P<0.01, **=P<0.0001. Adapted from Kennedy et al., 2016. This work was performed by the collaborating center in Austria.

4.3 Discussion

4.3.1 Candidate Variants in Sudden Unexpected Death with Cardiomyopathy

We have identified mutations in *PPA2* in four families affected with sudden unexpected death associated with cardiomyopathy. Cardiomyopathy is an inherited or acquired non-inflammatory disease of the myocardium (Chung and Kang, 2015). Clinically, cardiomyopathy is divided into hypertrophic, dilated and restrictive types, dependent upon echocardiographic assessment or the observation of the myocardium at post-mortem (Cecchi et al., 2012). In family CHD2, the disease was said to appear similar to hypertrophic cardiomyopathy, but did not fulfill the criteria for diagnosis, consisting mainly of fibrosis, cell disarray and scarring.

Families CHD2, 2c & 2d have a similar phenotype. In these families, children are affected with the condition in infancy, all children having suffered death before the age of 2 years. There was a consistent cardiac phenotype in these affected individuals with myocardial fibrosis and disarray identified on post mortem. In a number of these cases, there was a history of viral infection prior to the acute episode, and diagnosis of a viral illness was documented in P8 and P9. In family CHD2, one of the siblings had an episode of muscle weakness and hypotonia, prior to the onset of cardiomyopathy, but muscular disease was not a consistent phenotype in all of the four families. In family 2b, the phenotype was of later onset. The trigger for illness in this family appears to have been alcohol ingestion, to which the patients had an exquisite sensitivity. The mid-myocardial fibrosis seen in this family appears to be a pre-existing condition, having been identified as present in both living siblings.

The variant in *PPA2* in the Leeds family was identified from WES data. Following filtering, three variants remained. The *PPA2* variant was specifically prioritized as it was a good functional candidate and had the highest CADD score. However, the other two variants could not initially be ruled out as pathogenic. Both had high pathogenicity scores. The variant in *EPHA5* falls within a smaller homozygous region on chromosome 4. *EPHA5* codes for an ephrin receptor of the protein-tyrosine kinase family. It is involved in mediating developmental events, particularly in the nervous system (Das et al., 2016). This gene was a less convincing functional candidate than *PPA2*, as homozygous mutant mice have abnormal retinal mapping, but are otherwise phenotypically normal (Feldheim et al., 2004). The homozygous variant in *RALGAP1*

(*TULIP1*) falls outside of the largest homozygous regions. This gene encodes a major subunit of the RAL-GTPase activating protein and is a candidate for developmental delay, as zebrafish knockdown results in delayed brain development (Shimajima et al., 2009). There is no published literature to support a link between *TULIP1* and cardiomyopathy.

The variant in *PPA2* was therefore prioritized as the likely causative gene in this family based its segregation with disease in the family, high pathogenicity scores and previously published functional data.

4.3.2 Mitochondrial Function

A large proportion of the body's energy derived from fat and carbohydrate breakdown is created by oxidative phosphorylation (OXPHOS), which takes place in mitochondria. NADH and FADH₂ are created from the breakdown of glucose. During OXPHOS, electrons derived from NADH and FADH₂ combine with molecular oxygen and the energy released is used to drive the conversion of Adenosine Diphosphate (ADP) to Adenosine Triphosphate (ATP). Prior to energy release, these electrons are passed through a series of carriers, four complexes (I-IV) located on the inner mitochondrial membrane forming the Electron Transport Chain (ETC), and a final fifth complex, ATP synthase, that couples electron transport to ATP synthesis (Cooper and Hausman, 2016) (Fig. 4-8).

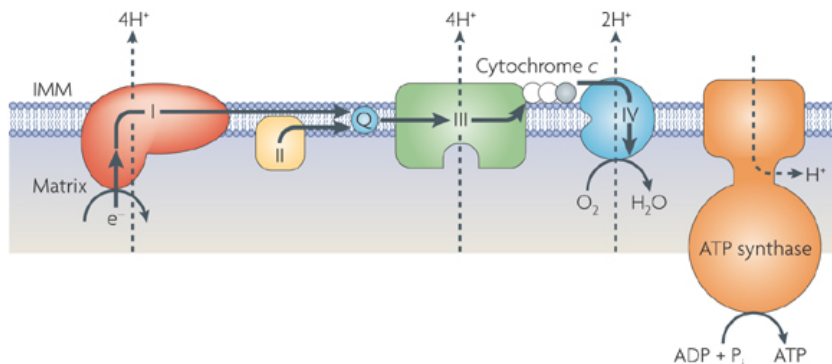


Figure 4-7: Diagram showing the complexes of the electron transport chain located at the inner mitochondrial membrane.

ATP synthase is shown converting ADP to ATP with the addition of Pi. Adapted from (Ow et al., 2008).

The phenotype identified in these families of cardiomyopathy, intermittent muscle weakness, hypotonia, raised CK and lactate was suggestive of a mitochondrial disease. Mitochondrial diseases often involve multiple organ systems and commonly present with neurological and myopathic features (Chinnery, 1993). Mitochondrial

diseases can also be specific to certain organ-systems, for example Leber Hereditary Optic Neuropathy (LHON), which affects only the eye. There is a known link between mitochondrial disease and cardiomyopathy. Cardiomyopathy is commonly seen in mitochondrial disorders, such as MELAS syndrome (Myopathy, Mitochondrial Encephalopathy, Lactic Acidosis, Stroke) (Jose et al., 2011) and the heart is known to be sensitive to the affects of mitochondrial disease as a high energy-expending organ. Cardiomyopathy has also been linked to mitochondrial disease in a number of animal models. For example, *OPA1* encodes a mitochondrial fusion protein and a mouse model carrying a nonsense mutation in *Opa1* develops late-onset cardiomyopathy (Chen et al., 2012). Furthermore, mitochondria are now being investigated as a potential drug target in the treatment of cardiomyopathy (Walters et al., 2012).

4.3.3 The Function of PPA2

Inorganic pyrophosphatases (PPAs) are enzymes that catalyze the conversion of inorganic pyrophosphate (PPI) into inorganic phosphate (Pi) (see Fig. 4-9). PPI is formed during many important cellular processes, for example DNA, RNA, protein and lipid synthesis. Pi formation is critical in the cell, as it is required for the formation of ATP (see Fig. 4-9). PPAs are therefore key cellular enzymes and are present in all forms of life. Eukaryotes have two forms of PPA: PPA1 is present in the cytoplasm and PPA2 is the mitochondrial form of the enzyme. PPA2 has a 60% homology to PPA1 and contains an N-terminal mitochondrial import sequence. PPA2 is ubiquitously expressed with the highest expression levels in liver, muscle, heart and kidney (Curbo et al., 2006). PPA2 has previously been shown to be located at the mitochondrial matrix (Curbo et al., 2006) and it has been suggested that the soluble catalytic part of the protein binds to an uncharacterized inner mitochondrial protein (Volk et al., 1983).

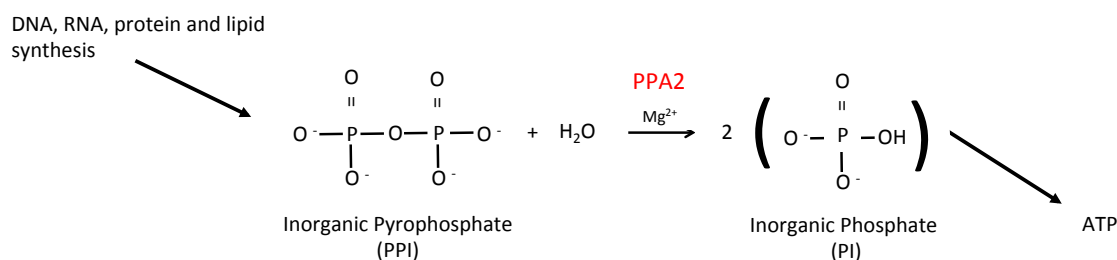


Figure 4-8: The formation of Inorganic Phosphate from Inorganic Pyrophosphate by PPA2.

Prior to this investigation, the key studies of PPA2 function involved the yeast *Saccharomyces cerevisiae*. Lundin et al., 1991, first cloned the *PPA2* gene in yeast. To investigate its function they disrupted expression using a URA3 marker. This created a

viable yeast strain, but it could no longer grow on respiratory carbon sources, for example acetate or ethanol. The yeast strain lacking *ppa2* was found to have a lack of mitochondrial DNA and the authors therefore concluded that *PPA2* was essential for the maintenance of the mitochondrial genome (Lundin et al., 1991). Based on this finding Curbo *et al.* sequenced *PPA2* in 13 patients with mitochondrial depletion syndromes, but were not able to identify a pathogenic mutation (Curbo et al., 2006).

From our findings, it seemed reasonable to suggest that mutations in *PPA2* were causing a mitochondrial dysfunction in our families, possibly due to a deficiency in mitochondrial Pi, causing a defect in mitochondrial ATP synthesis. This defect in mitochondrial function was presenting as a cardiomyopathy, a phenotype consistent with mitochondrial dysfunction.

4.3.4 Discussion of Functional Studies Performed in PPA2

A number of functional studies to investigate *PPA2* function in the patients identified in this study were undertaken by our collaborators in Austria. The results of these experiments are described in detail in the associated publication (Kennedy et al., 2016). In summary, Western blot analysis confirmed a normal level of *PPA2* protein in the majority of tissue samples tested from affected patients. Western blot was performed for all individuals where fibroblast, skeletal muscle or cardiac tissue samples were available. The only reduction in *PPA2* levels were in fibroblast mitochondria and skeletal muscle samples from patient P9, although these were normal in P6, who carried the same mutation, perhaps implying experimental or tissue-specific differences in expression, rather than a functional defect. In heart autopsy samples, patient P10 appeared to have a reduction in *PPA2* protein levels compatible with a complex I deficiency, showing a reduction of *NDUFS4*, a subunit of the complex I respiratory chain. P10 was the only patient identified to have a defect in respiratory chain analysis on cardiac tissue samples, affecting complexes I and IV, which would be in keeping with this finding. In cardiac tissue samples, patient P7 appeared to have a reduction in total mitochondrial number, demonstrated by a reduction in levels of *PPA2*, *NDUFS4*, *CS* and porin, but this was not mirrored in the fibroblast sample from P7 or the other patient samples available, and so may be due to changes in tissue composition post mortem. Overall, it did not appear that the identified mutations were consistently affecting *PPA2* protein levels in heart, muscle or fibroblast samples (Kennedy et al., 2016).

Oxygen Consumption Rate (OCR) was undertaken by our collaborators and was measured by micro scale respirometry by Seahorse Assay (XF96 Seahorse Biosciences, Agilent Technologies) in three patient and three control fibroblast samples. This analysis showed similar results for patient and control samples, except for following the addition of the mitochondrial uncoupler FCCP (Carbonyl-cyanide-4-(trifluoromethoxy)-phenylhydrazone), where patient samples showed almost double the rate of controls (Kennedy et al., 2016). An uncoupling protein dissipates the mitochondrial proton gradient inhibiting ATP synthesis without affecting the respiratory chain (Terada, 1990). Addition of FCCP is a measure of the reserve respiratory capacity (RRC), an estimate of the difference between basal and maximal mitochondrial activity, or how well mitochondria respond to stress (Sansbury et al., 2011). Although an effect on RRC is a relatively non-specific finding, it could be hypothesized that the observed difference is due to the limitation of ATP synthesis in patient cells due to a lack of Pi. To support this hypothesis, cells from patients with a specific ATP synthase deficiency, due to mutations in *TMEM70* or *ATP5E*, demonstrate an almost identical change in OCR (Cizkova et al., 2008, Mayr et al., 2010).

Due to the non-specific nature of the OCR results, further functional studies were used to measure pyrophosphatase activity in mitochondria extracted from patient fibroblasts and recombinant proteins were expressed to model the identified missense mutations. The results of both of these experiments supported the pathogenicity of the mutations identified in this gene. The c.500C>T (p.Pro167Leu) and c.514G>A (p.Glu172Lys) variants had a 5-10% activity compared to the wildtype enzyme, but the c.683C>T (p.Pro228Leu) variant showed a 24-28% activity when compared to wildtype (Kennedy et al., 2016). c.683C>T (p.Pro228Leu) is the variant identified in family 2b, who had a later onset of disease than the other families studied, which may be explained by the higher residual enzyme activity. None of the missense mutations appeared to have any effect on the stability of the protein.

4.3.5 Identified Mutations in *PPA2* are Hypomorphic

All of the *PPA2* mutations identified in this study are missense mutations. It is reasonable to hypothesise that biallelic loss-of-function mutations in this gene would be incompatible with life, and that the identified mutations are hypomorphic. To support this, in the large exome/genome project GnomAD, only 25 loss-of-function alleles are present in *PPA2* and none of these are present in a homozygous state. In the EXaC database, only 71 potentially pathogenic missense variants are listed in *PPA2* (those

with a SIFT score <0.05), which would result in an incidence for *PPA2*-related disease at 1 in 170,000. This suggests the *PPA2* mutations identified are hypomorphic, leaving some residual enzyme function that would be compatible with viability.

4.3.6 How is *PPA2* Dysfunction Linked to the Clinical Phenotype?

Some of the individuals identified in this study were shown to have a pre-existing cardiac fibrosis, proven by magnetic resonance imaging (MRI) investigation in the siblings of family 2b. Sudden demise or death in the described families was apparently triggered by an acute event. The pathophysiology of these trigger events is not clear and we are not able to readily explain the difference in phenotypic presentation observed between these families. In family 2b, the trigger appeared to be alcohol. In families CHD2, 2c and 2d, a viral illness was suspected to be the trigger. At the same time as the publication of our findings, a French group published a paper in the same journal describing three families with a very similar phenotype, where children had died of sudden cardiac arrest between the ages of 4 and 20 months and had biallelic mutations in *PPA2*. Similar cardiac findings were observed at post mortem, ranging from focal myocarditis to hypertrophic cardiomyopathy and cardiac fibrosis. One case was diagnosed with a myopathy prior to cardiac arrest. The same association with a 'trigger' was not recorded in these cases (Guimier et al., 2016).

Further work will be required to fully understand the pathophysiology of the clinical presentation associated with *PPA2* mutations. The observed pre-existing cardiomyopathy could be due to a chronic ATP deficiency, affecting the cardiac muscle. Cardiomyocytes are particularly sensitive to defects in ATP-generating processes and impairments in ATP synthesis can rapidly affect their contractile function (Doenst et al., 2013). Conversely, ATP depletion is seen in disease processes such as heart failure (Doenst et al., 2013). High levels of PPi could also be associated with the cardiac cell dysfunction seen, which beyond a certain threshold has resulted in sudden cardiac arrest. It is known that abnormally high levels of PPi can result in competitive inhibition of mitochondrial ADP/ATP translocase (Griffiths and Halestrap, 1993), which enables the exchange of ADP and ATP across the inner mitochondrial membrane, and this may be associated with a reduction in the ability of the cell to produce energy for contraction. Furthermore, the accumulation of PPi is known to inhibit a number of other important cellular processes, such as the biosynthesis of NAD^+ , the sequestration of Mg^{2+} ions and the polymerization of macromolecules (Heinonen, 2013). An excessive accumulation of PPi in respiring yeast cells leads to

cell cycle arrest, which is reversed with the removal of accumulated pyrophosphate. Fermenting yeast cells suffer not only cell cycle arrest, but also eventual cell death by autophagy (Serrano-Bueno et al., 2013). The phenotype observed in these families may therefore be a manifestation of chronic ATP deficiency, high PPI concentrations, or a combination of both of these factors.

Further to this, there is a known link between alcohol ingestion and PPA function. Ethanol is oxidized to acetaldehyde and further to acetic acid. Acetic acid is activated to acetyl co-enzyme A, which is accompanied by formation of stoichiometric amounts of PPI (Yamashita et al., 2001, Yamada et al., 1988). If PPA function is deficient, an excess accumulation of PPI in the body's tissues and, specifically cardiac tissues, may therefore be playing a role in the suspected fibrosis and sudden arrhythmia and death observed in family 2b. We can hypothesize that an excess of cellular PPI following alcohol consumption could lead to cardiac cell dysfunction and apoptosis, exacerbating the underlying cardiomyopathy, and acting as an acute trigger for disease.

4.3.7 The Importance of the Identification of Mutations in *PPA2*

Mutations in *PPA2* were identified in four families with a very severe phenotype, leading to cardiomyopathy and early death in affected family members. The phenotype investigated in these families is an extremely important one, as, although rare, it is a differential for Sudden Unexpected Death in Infants and Children (SUDIC). SUDIC is the most common cause of post-neonatal mortality in developing countries (Fleming et al., 2015). The cause of SUDIC is often unidentified, but potentially pathogenic variants in genes associated with cardiomyopathy, cardiac arrhythmia or cardiac ion channel dysfunction may be associated in a proportion of cases (Klaver et al., 2011) and metabolic diseases related to fatty acid oxidation and ketone body metabolism also require exclusion (Oshima et al., 2017). The families investigated in this report have not only lost one or more children to this condition, but have often been through a lengthy process in attempting to reach a diagnosis for their children, and in attempting to understand the likely risks to the surviving siblings and extended family members. The identification of a genetic cause for disease has provided information and knowledge for the families and clinicians involved in these cases and has directly affected clinical management.

The identification of mutations in this gene has provided key information for the families described. These families can now access information regarding future recurrence risk and prenatal testing could be offered if so desired. In family CHD2, the surviving sibling

has undergone genetic testing to ascertain their risk of disease and carrier testing is now available to further extended family members. This has been of great importance to the family who were very concerned regarding the risks to their surviving child and current pregnancy.

This case has provided an excellent example of a condition where reaching a genetic diagnosis has made a clear difference to patient management and has allowed life-saving measures to be employed. For example, in family 2b, the surviving siblings have been fitted with ICD devices, are having regular MRI surveillance, and have been educated on the avoidance of alcohol, with the aim of avoiding arrhythmia and sudden death.

Diagnostic testing in this condition can be performed by *PPA2* mutation analysis, but could also be aided by MRI investigation to assess for signs of cardiomyopathy and conceivably PPA activity levels could be used in diagnosis, although this type of testing is currently only available on a research basis. It may be that, as more families with this condition are identified, further genotype-phenotype correlation can be delineated and standardized strategies for the management of affected patients can be created.

It is not known whether heterozygous carriers of *PPA2*-related disease are at risk of cardiac anomaly, none of the known heterozygous carriers in the families described are affected with a clinical phenotype, but analysis of PPA activity could be used to further assess carriers for mild features of the disease. PPI is known to be an inhibitor of calcium phosphate deposition, where high serum Pi promotes ectopic calcification in a number of tissues, including blood vessels (Villa-Bellosta et al., 2017). Vascular calcification is associated with cardiovascular disease, for example coronary atherosclerosis (Otsuka et al., 2014). It is therefore reasonable to hypothesize that the *PPA2* gene could be a susceptibility gene for later-onset cardiovascular disease, with variants affecting PPA2 levels to a lesser degree than observed in our families with severe, young-onset disease. It would be extremely interesting to screen large cohorts of patients with later-onset cardiovascular disease for variants in this gene.

4.4 Conclusions

We have shown that mutations in *PPA2*, coding for an inorganic pyrophosphatase, are a rare cause of sudden unexpected death in infancy and childhood. This is an important clinical phenotype, as *PPA2*-related disease is a differential for sudden unexpected death in childhood. We also describe a family with an unusual phenotype

of extreme sensitivity to alcohol with mutations in this gene. *PPA2*-related disease encompasses these two very distinctive phenotypes, for which a molecular diagnosis should be sought if suspected. A molecular diagnosis of *PPA2*-related disease is of great importance as screening of family members, using cardiac MRI, mutation analysis or analysis of PPA activity could identify those individuals at risk. For family members at risk, interventions such as education on the avoidance of possible triggers and implantation of an ICD device may prevent the lethal consequence of disease.

The finding of mutations in *PPA2* in human disease highlights the critical role that *PPA2* plays in mitochondrial function and gives further insight into the pathophysiology of human heart disease.

5 Mutations in *DNAAF1* are Associated with Congenital Heart Disease in the Absence of Primary Ciliary Dyskinesia

5.1 Introduction

DNAAF1 is a dynein arm assembly factor and a known cause of PCD in humans (Loges et al., 2009). The link between PCD, laterality disorders and complex CHD has been long-established. Here we describe the identification of a missense mutation in *DNAAF1* (a known PCD disease gene and dynein arm assembly factor, MIM number 613193) in a family with heterotaxy and complex CHD, but without a clinical diagnosis of PCD. We have sought to identify typical clinical features of PCD in this family, but clinical investigations did not support this diagnosis.

We have collaborated to ascertain a small cohort of patients with mutations in *DNAAF1*. Those with frameshift mutations or deletions in the gene have a typical phenotype of PCD, often associated with cardiac defects. We hypothesize that the missense mutation identified in our family is hypomorphic, affecting the nodal cilia in embryonic development and leading to cardiovascular defects, but having no apparent effect on respiratory ciliary function. We have demonstrated the pathogenicity of this missense mutation by modeling in a *dnaaf1*^{-/-} null mutant zebrafish model.

We have also investigated the function of *DNAAF1* by assessment of protein-protein interactions and protein co-localisation in ciliated cells. The results of these experiments have identified novel interactions of *DNAAF1* and have provided new insights into the possible functions of *DNAAF1* during dynein arm assembly in motile cilia. Specifically, we have investigated the expression of *DNAAF1* and a novel interacting partner, RUVBL1/Pontin at the murine and zebrafish embryonic node and developing heart. Our data suggests that the *DNAAF1*-RUVBL1 interaction is required for symmetry breaking during the development of laterality in early embryogenesis.

The work described in this chapter was part of a multi-center collaboration, led by the researchers at the University of Leeds. Where the described work was performed by collaborating researchers, rather than by Dr Hartill, this is indicated in the text.

5.2 Results

5.2.1 Clinical Phenotype

Family CHD1 are a consanguineous family of Pakistani origin who have two children affected with complex CHD (Figure 5-1). Affected individual JT66 has right atrial isomerism (RAI) with complete AVSD, AV valve regurgitation, double outlet right ventricle, pulmonary atresia, and right pulmonary artery stenosis. Blalock-Taussig (BT) shunt and subsequent bidirectional cavopulmonary (Glenn) shunt have been performed. The affected sibling, JT67, has heterotaxy; dextrocardia, right atrial isomerism, left-sided liver, right-sided stomach and polysplenia. In terms of cardiac anatomy, he has double outlet right ventricle with the pulmonary artery to the right and the aorta to the left side. Both ventricles are described as small on MRI imaging and there is a large inlet VSD and severe sub-pulmonary stenosis. This child has had a previous Glenn shunt performed. Neither child complains of recurrent chest infections or chronic cough. They are non-dysmorphic and have normal learning, with average performance at school. Karyotype analysis has previously been performed and was normal.

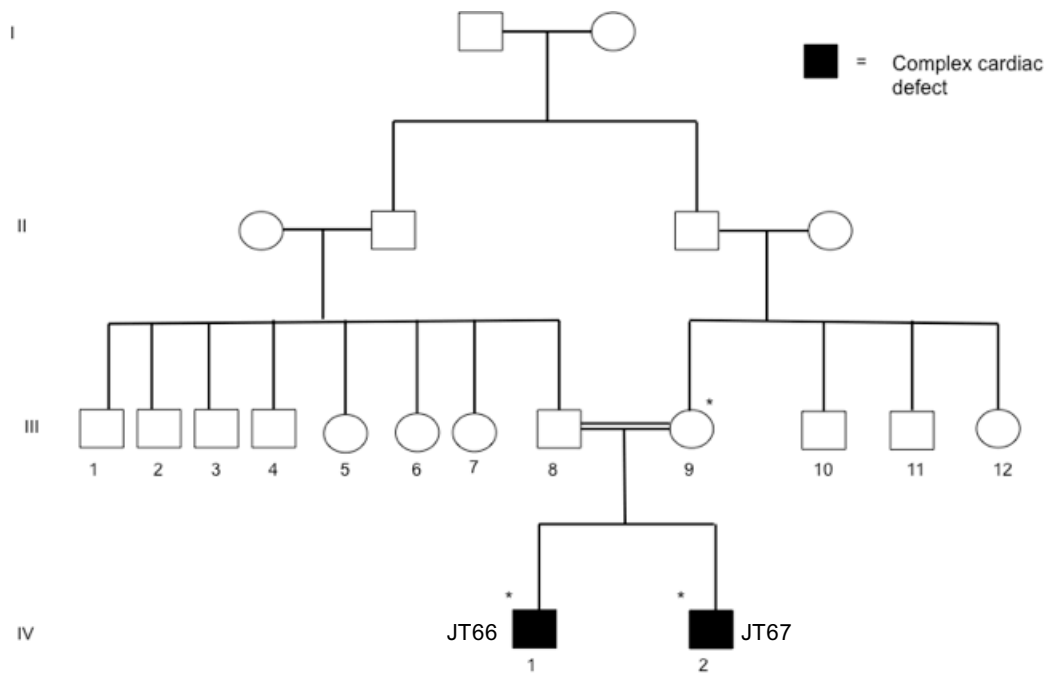


Figure 5-1: Pedigree of Family CHD1.

Shaded individuals are affected with complex CHD. DNA samples were available from individuals indicated by an asterisk.

5.2.2 SNP microarray

Affymetrix 6.0 SNP microarray analysis was performed on samples from both affected individuals in this family (JT66 and JT67). Large regions of homozygosity (>1,000,000 bp) shared between the affected individuals are shown in Table 5-1.

Table 5-1: Large regions of homozygosity, identified on SNP array, shared between the affected siblings in family CHD1.

Chromosome	Start position	End position	Size (bp)
chr13	30744080	43111120	12367040
chr1	49589806	51029387	1439581
chr2	228056368	234938781	6882413
chr2	237132461	243102476	5970015
chr5	45417643	50062543	4644900
chr12	130320647	133779461	3458814
chr16	1119946	3640274	2520328
chr11	2544340	4867962	2323622
chr14	21511789	23329446	1817657
chr1	147635483	149396403	1760920
chr5	129573666	130769785	1196119
chr20	25040884	26129447	1088563
chr5	41629963	42714692	1084729
chr14	66824734	67897700	1072966
chr16	83207071	84258711	1051640
chr1	155086148	156135241	1049093
chr8	50367785	51371466	1003681

5.2.3 Whole Exome Sequencing

WES was performed for both affected individuals. For patient JT66, this was performed on an Illumina MiSeq prior to the start of this research project. For JT67 this was performed using the Illumina XT kit, as previously described, on an Illumina HiSeq 2500 with 6 samples pooled per lane.

Data analysis followed the standardized pipeline. In this instance, variants were filtered out which were present at a Minor Allele Frequency (MAF) >1% in dbSNP 138, the NHLBI Exome Sequencing Project (ESP) Exome Variant Server, the Exome Aggregation Consortium (ExAC) database, and 200 locally-sequenced ethnically-matched control samples. Variants were also retained if they followed the predicted autosomal recessive inheritance pattern in the family (*i.e.* were homozygous in both affected siblings). Table 5-2 shows the number of variants remaining after each stage of data filtering and analysis.

Table 5-2: Number of variants remaining following each step in WES data analysis for family CHD1.

Stage in Analysis	Number of variants
Called with GATK	424054
Filter on dbSNP/EVS (<1%)	116753
<1% local exomes	92525
<1% ExAC	92062
Functional variants	732
Filter on family structure (homozygous shared between affecteds)	3
Predicted 'pathogenic' by <i>in silico</i> tools	2
CADD score >15	2

Two variants remained following filtering and analysis, which are summarized in Table 5-3.

Table 5-3: Variants remaining following filtering in family CHD1.

Gene	Variant	Protein change	Polphen2	SIFT	Condel	CADD	MAF (GnomAD)	Known Protein Function
<i>DNAAF1</i>	16:84188400C>T	p.L191F; NP_848547	Probably damaging (0.996)	Deleterious (0.01)	Deleterious (0.867)	25.6	-	Dynein assembly factor
<i>SPHKAP</i>	2:228883539C>G	R677S; NP_001136 116	Benign (0.044)	Deleterious (0.02)	Neutral (0.386)	22.6	257/246166, 4 homozygotes.	Protein kinase A binding

The variant in *DNAAF1* appeared to be the most likely candidate in this family. It fell within a large homozygous region, identified from the SNP data, on chromosome 16. An IGV readout showing this variant in homozygous form in both affected individuals is shown in Figure 5-2. WES quality scores and depth of coverage were assessed and showed high quality sequencing (see appendix F).

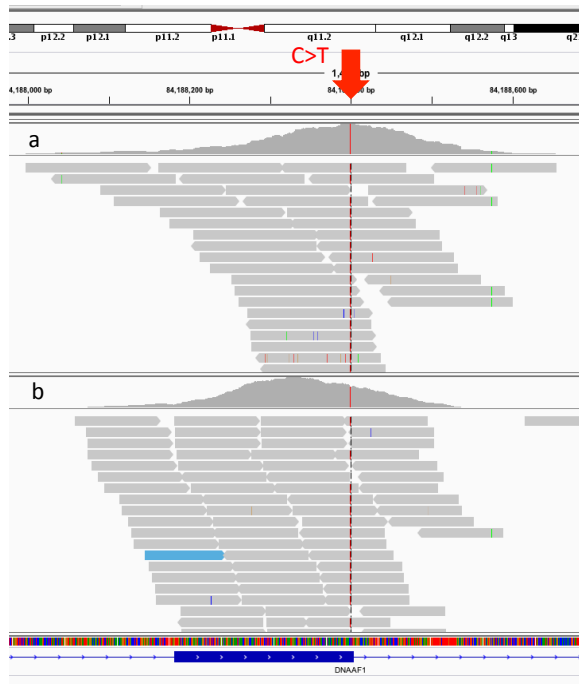


Figure 5-2: IGV readout showing the variant in *DNAAF1*, c.571C>T, which is homozygous in both affected individuals in family CHD1.

The variant is shown in affected individuals a) JT66 and b) JT67. The C>T variant is marked by the red arrow. Aligned reads are shown in grey.

5.2.4 CNV analysis

CNV analysis was performed using the Exome Depth software on WES data in individual JT67. A number of potential candidate CNVs were identified by the program, but none overlapped a likely candidate gene for complex CHD and heterotaxy. A full list of CNV variants identified is provided in appendix J.

5.2.5 Analysis of the Missense Variant in *DNAAF1*

The variant in *DNAAF1* was prioritized over the other filtered variants. Using *in silico* tools, it scored 'probably damaging' on PolyPhen2, 'deleterious' on SIFT and 'deleterious' on Condel. The CADD score was 25.6. This variant was very rare: it was absent from the ExAC and the GnomAD databases. In a local control dataset there was 1 heterozygote in 3000 ethnically-matched control samples, but no individual was

homozygous. The position of the variant is highly conserved in other vertebrate species, specifically mammals and amphibians, as shown in Figure 5-3.

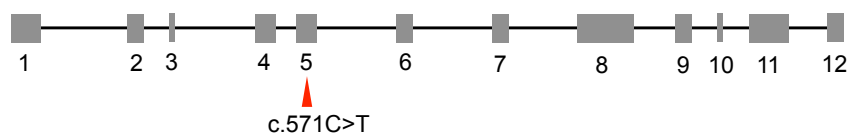
				L191F ↓	
Homo sapiens	NP_848547.4	149	AQTELRCLFLOMNLRLKIEPLEPLQKLDALNLSNYYIKTIENLSCLPVLN		198
Pan troglodytes	XP_511137.2	149	AQTELRCLFLOMNLRLHKKIENLEPLQKLDALNLSNYYIKTIENLSCLPVLN		198
Canis lupus familiaris	XP_536764.3	151	AQTELRCLFLOVNLRLHKKIENLESLQKLDALNLSNYYIKTIENLSCLPVLN		200
Bos taurus	NP_001030422.2	149	AQTELRCLFLOVNLRLHKKIENLEPLQKLDALNLSNYYIKTIENLSCLPVLN		198
Mus musculus	NP_080924.1	143	AQSELRCLFLOVNLRLHKKIENLEPLQKLDALNLSNYYIKTIENLSCLPVLN		192
Rattus norvegicus	NP_001014176.1	143	AQSELRCLFLOVNLRLHKKIENLEPLQKLDALNLSNYYIKTIENLSCLPVLN		192
Xenopus tropicalis	XP_004913593.1	125	AQTELRCLFLHQNLVHKKIENLDHLQKLDLNLNSNCKIKTIENLSCLKVLN		174

Figure 5-3: Conservation of variant p.L191F in DNAAF1 orthologues.

Conservation is shown within the region of LRR (leucine-rich repeats) in the proteins. The indicated protein sequences were aligned by HomoloGene. The arrow indicates the conservation of the Leu191 residue in all species down to *Xenopus tropicalis*.

This variant causes an amino acid substitution, leucine to phenylalanine, in exon 5 of *DNAAF1*. It lies within one of the LRR (leucine-rich repeat) domains of the protein. This domain contains a motif of repeats with a highly-conserved leucine present at specific points along the sequence. LRR domains are often involved in protein-protein interactions (Loges et al., 2009). The putative mutation would be predicted to substitute a phenylalanine for a conserved leucine (Figure 5-3 above) and is therefore likely to disrupt protein function by perturbing protein-protein interactions. The domains of the *DNAAF1* protein are shown in figure 5-4, with the position of the conserved Leu191 residue indicated.

NM_178452.5 (hg19)



NP_848547.4

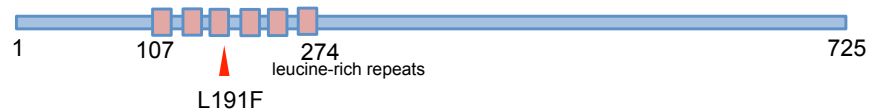


Figure 5-4: Position of c.571C>T in the *DNAAF1* gene and position of the missense substitution, L191F, in the *DNAAF1* protein.

The missense variant falls within the third leucine-rich repeat of the protein.

5.2.6 Sanger Sequencing Confirmation of Variant in *DNAAF1*

Confirmation of the variant in *DNAAF1* by Sanger sequencing was performed for both affected individuals and their unaffected mother. DNA from the father was not available for testing. Figure 5-5 shows the variant in *DNAAF1* present in the homozygous state in both of the affected individuals and in the heterozygous state for their mother. Primer sequences can be found in appendix E.

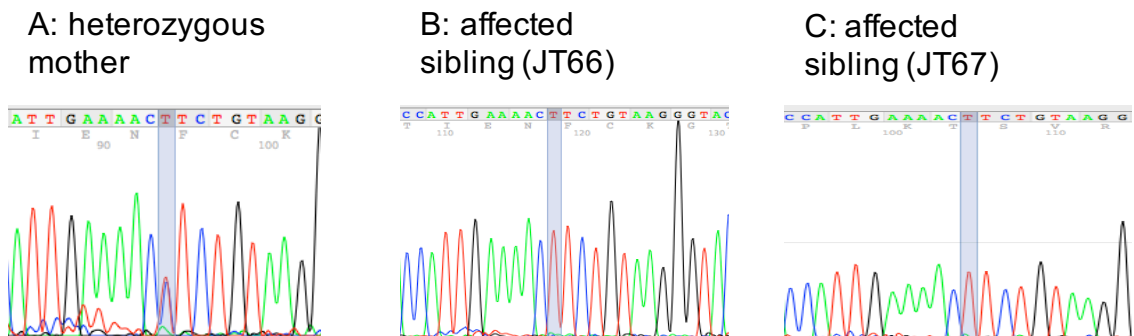


Figure 5-5: Sanger sequencing electropherograms to show the c.571C>T variant in *DNAAF1* in family CHD1.

Electropherograms showing (A) the mother of JT66 and JT67, showing the variant in heterozygous state, (B) individual JT66 (affected male) and (C) individual JT67 (affected male), showing the variant in the homozygous state. The variant in *DNAAF1* is highlighted with blue shading.

5.2.7 Investigation of PCD in Family CHD1

Due to the identification of a variant in *DNAAF1* in family CHD1, the family was further contacted to investigate a diagnosis of PCD in the affected siblings. Initially, a specific clinical history was taken: neither child had suffered from typical symptoms of PCD such as persistent neonatal respiratory distress, rhinorrhea, respiratory infections or chronic cough. JT67 had suffered from recurrent episodes of otitis media. Exercise tolerance was reduced in both siblings, but this was thought to be due to cardiac disease.

Clinical testing for PCD was performed by Dr. Amelia Shoemark, PCD senior clinical scientist, Royal Brompton and Harefield NHS Foundation Trust. Nasal cell brushings were taken concurrently from JT67 and a normal control individual. Video microscopy was initially performed on nasal epithelial cells to investigate ciliary beat and waveform. The sample from JT67 identified a sparsely ciliated epithelium. The cilia were effectively clearing surrounding debris and there was a mixed beat, with some slow, dyskinetic and static patches. The control sample taken concurrently showed a coordinated ciliary beat, but there were also some slow, dyskinetic and static patches,

similar to those seen in the patient, and likely resulting from sample storage and transport rather than an underlying pathology. The Ciliary Beat Frequency reported in affected individual JT67 was 8.7Hz (SD 2.2Hz; within the normal range), compared to 9.1Hz; SD 2.1Hz) in the normal control sample taken concurrently.

Transmission Electron Microscopy (TEM) was also undertaken and gave a normal result, demonstrating the presence of inner and outer dynein arms in the sample from patient JT67, although only 10 ciliary cross-sections could be assessed (data not shown). Nitric Oxide (NO) measurements in JT67 showed a level of 254 ppb (63 nL/min) in the right nostril (normal level >187 ppb) and 121 ppb (30 nL/min) in left nostril (low, but this nostril was blocked by a upper respiratory tract infection at the time of testing). The ambient level was 17 ppb. Overall, there was no evidence to support a diagnosis of PCD in affected individual JT67 on clinical testing, although a sub-clinical PCD could not be excluded.

5.2.8 Investigation of Mutations in *DNAAF1* in Other Heterotaxy Patients

Due to the finding of a putative pathogenic mutation in *DNAAF1* in a family with heterotaxy, we screened other patients with a similar phenotype for mutations in the same gene. Lymphoblastoid cell lines were available to our group from a cohort of additional patients with heterotaxy and other limited phenotypic information. Those with the most complete clinical information, compatible with a diagnosis of heterotaxy, were selected for DNA extraction. Sanger sequencing was performed to sequence all 15 exons of *DNAAF1* in 19 patients and the identified variants were analyzed using *in silico* tools. Table 5-4 shows a summary of the 19 sequenced heterotaxy patients, showing the known clinical information and identified variants. No likely pathogenic variants were identified in these 19 patients. For primer sequences see appendix E.

Table 5-4: Heterotaxy samples screened for mutations in *DNAAF1* by Sanger sequencing.

Sample name	Phenotypic information	Sequencing
IV5	Dextrocardia, situs inversus, VSD, PDA, isomerism	Full gene covered. No mutation.
IV9	AVSD, TGA, asplenia, trilobed lungs, RAI, symmetric liver, intestinal malrotation, pancreatic aplasia	Full gene covered. No mutation.
IV10	Right atrial isomerism, hepatic draining to IVC, aorta to right of spine, AVSD, DORV, asplenia	Full gene covered. No mutation.
IV15	Right atrial isomerism, AVSD, central liver.	Sequencing failed, likely poor quality DNA
IV19	Complete situs inversus	Full gene covered. No mutation.
CD5	Left isomerism sequence	Full gene covered. No mutation.
CD10	Right isomerism sequence	Full gene covered. rs9972733 homozygous, MAF 0.097 in ExAC.
CD20	Left isomerism sequence	Full gene covered. No mutation.
CD46	Left atrial isomerism, parents first cousins	Full gene covered. No mutation.
CD50	Right atrial isomerism, parents first cousins	Full gene covered. rs372390124 heterozygous, MAF 0.0001 in ExAC.
CD58	Tetralogy of Fallot.	Full gene covered. No mutation.
CD60	Laterality defect, complex CHD, consanguineous family.	Full gene covered. No mutation.
CD61	Laterality defect, complex CHD	Full gene covered. No mutation.
CD133	Left atrial isomerism, sister died of complex CHD	Full gene covered. No mutation.

5.2.9 Identification of Further Families with Mutations in *DNAAF1* Through Collaborating Centres

Through collaboration, we were able to identify additional families with mutations in *DNAAF1* who had been referred to the diagnostic PCD services based at University College London and the Yorkshire Molecular Genetics Laboratory. These families had been investigated by WES analysis and mutations in *DNAAF1*, but no other PCD gene, were identified. In contrast to the clinical features in family CHD1, all three families had typical features of PCD.

Family 1b have three children affected with PCD (see Figure 5-6). All three children suffered with recurrent chest infections, bronchiectasis and multiple hospital admissions. A diagnosis of PCD was confirmed in each sibling following nitric oxide

testing (for example Individual II:1 had an NO level of 40 ppb (10nL/min)) and completely immotile cilia were identified when assessed by high speed video imaging. TEM analysis was abnormal in the three siblings, showing the absence of ciliary outer dynein arms in 84%, 92% and 100% of cilia cross-sections, with occasionally absent inner dynein arms (affecting 21%, 22% and 30% of cross-sections, respectively). Two of these three patients were affected with *situs inversus*, associated with asplenia and RAI. One of these siblings had concurrent AVSD, pulmonary stenosis and anomalous pulmonary venous drainage, and the other had dextrocardia (azygous vein to left-sided superior vena cava) and anomalous pulmonary venous drainage. This family had a further child affected with RAI, but who died of cardiac complications before a diagnosis of PCD was explored. In Family 1b, heterozygous frameshift mutations in *DNAAF1*; c.281delA and c.1484delC, were identified by the team at UCL. These mutations segregated with disease in the family and were confirmed by Sanger sequencing.

Family 1c consists of a single affected male with confirmed PCD. He has typical features of the condition including longstanding otitis media, dextrocardia (but no other cardiac disease), and asthenospermia. This male has a homozygous mutation in *DNAAF1*, c.1484delC, resulting in a frameshift mutation, p.Pro495fs, identified by the Yorkshire Molecular Genetics Laboratory.

Family 1d consists of two siblings affected with PCD. The first (II:1) had suffered from respiratory distress as a neonate and then episodes of persistent rhinorrhea. He had *situs inversus* (including dextrocardia), transposition of the great vessels, VSD and tricuspid regurgitation. A diagnosis of PCD was made following ciliary brushings, which showed static cilia, and a lack of IDA and ODA on TEM. However, ODA stubs were noted to be present. The second child (II:2) had a history of chronic cough, but no evidence of cardiac disease. Clinical investigations showed static cilia and a lack of both IDA and ODA. Both affected individuals in Family 1d share a 7kb out-of-frame homozygous deletion in *DNAAF1* (GRCh37; Chr16:84,193,165-84,201,105) which affects exons 6 and 7 of the gene. This deletion was identified from whole exome sequencing data by Dr Chris Watson at the Yorkshire Molecular Genetics Laboratory. He was able to identify an absence of mapped reads at this position (Fig. 5-6). The breakpoints were confirmed by whole genome sequencing (WGS) and a breakpoint-spanning Sanger sequencing assay (see Figure 5-6).

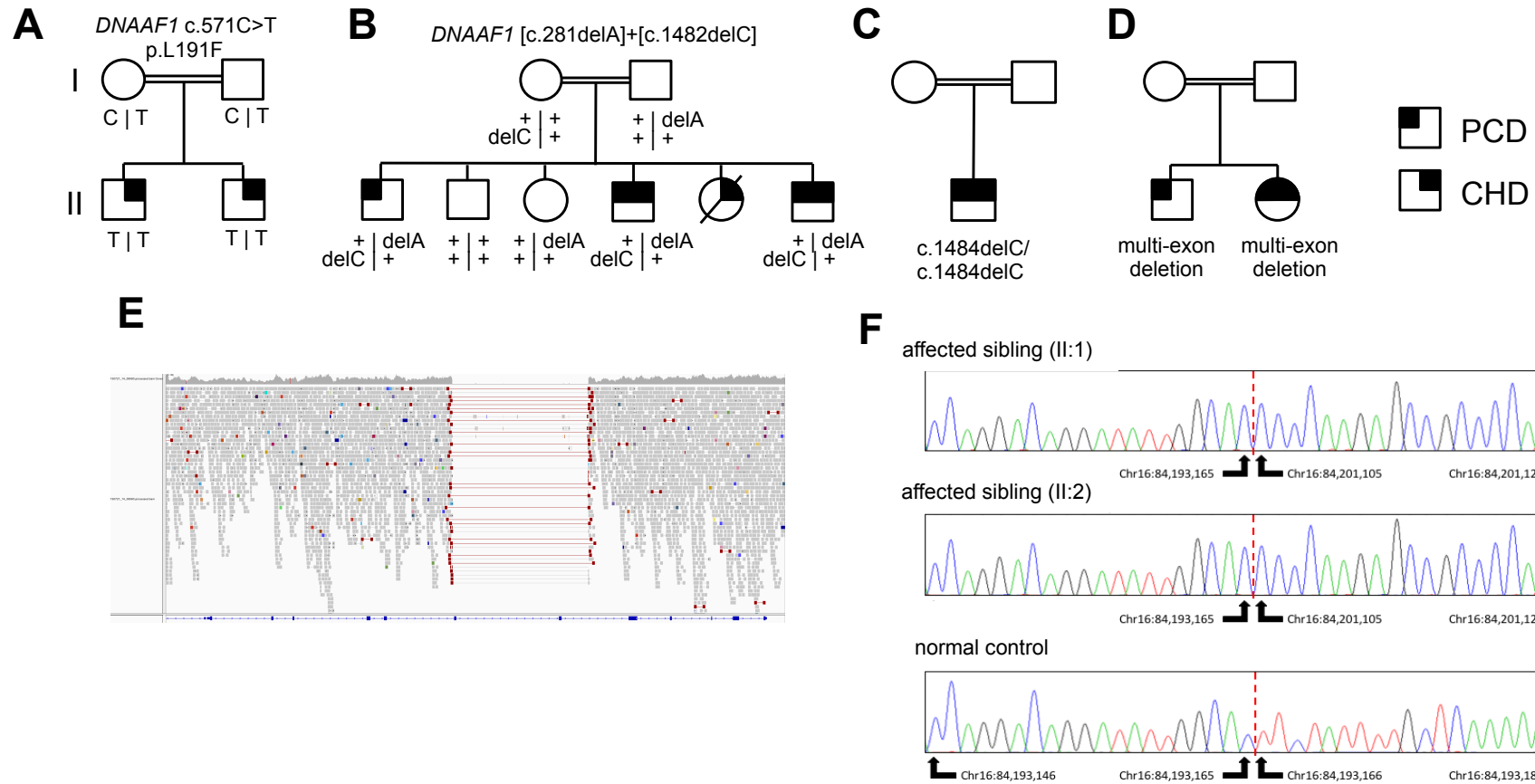


Figure 5-6: Pedigrees of families with *DNAAF1* mutations. Mapping of an identified deletion in *DNAAF1* using WGS and Sanger sequencing.

Pedigrees of A) family CHD1, B) family 1b, C) family 1c & D) family 1d. A shaded upper left quadrant indicates a clinical diagnosis of PCD, a shaded upper right quadrant indicates a diagnosis of Congenital Heart Disease (CHD). E: IGV readout of sequencing reads in family 1d, showing an absence of aligned reads at position 16: 84,193,165-84,201,105. F: Sanger sequencing assay to show breakpoints of homozygous deletion in family 1d.

5.2.10 Creation of DNAAF1 Constructs

Gateway cloning was used to create DNAAF1 N-terminal TAP (which includes a FLAG sequence) and pCS2+ constructs, for use in further experiments. The variant identified in family CHD1, c.571C>T, was introduced into the sequence by site-directed mutagenesis. Sanger sequencing was used to confirm the successful creation of the wildtype and mutant constructs. Figures 5-7 and 5-8 show maps of the N-terminal TAP construct and the pCS2+ constructs respectively.

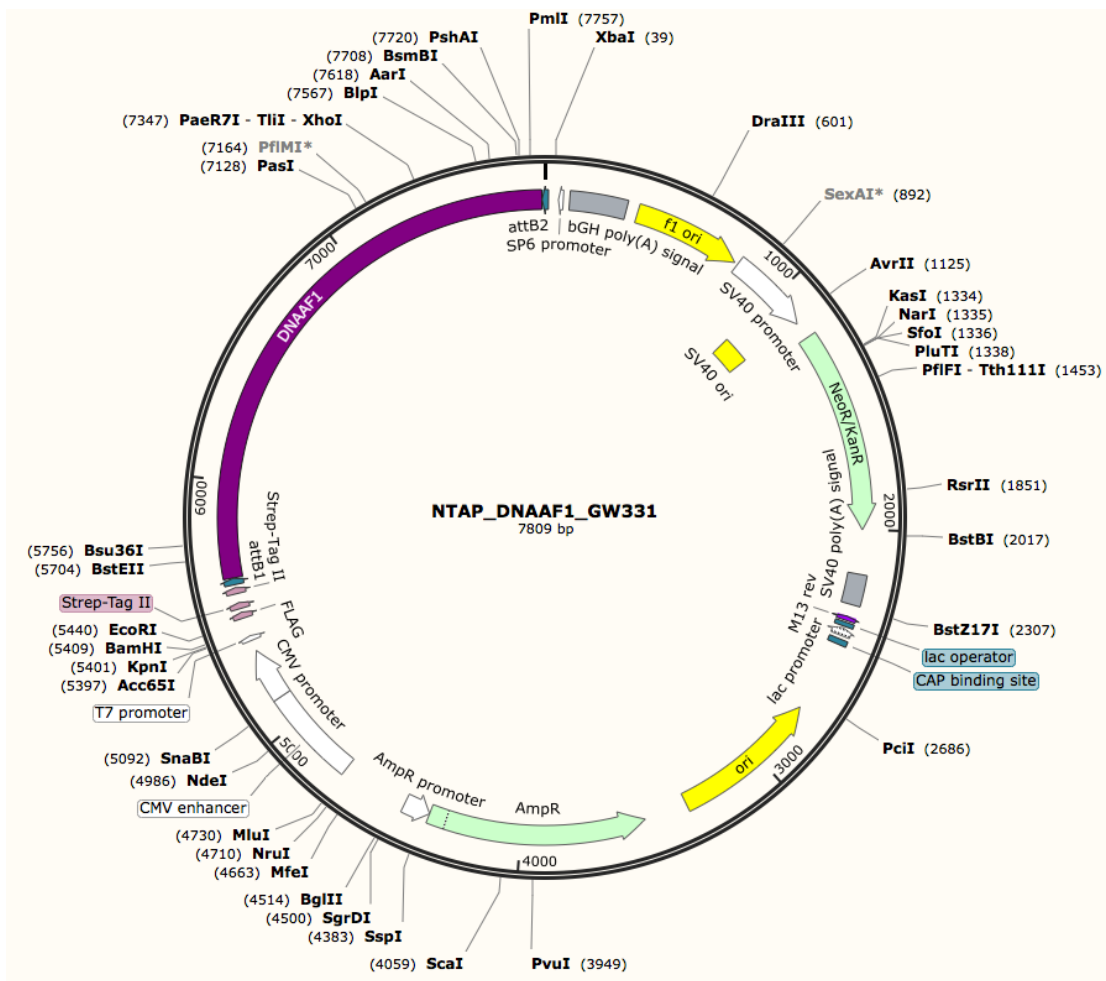


Figure 5-7: A representation of the N-terminal TAP DNAAF1 construct.

The DNAAF1 gene is shown in purple. Restriction enzyme sites are shown by black lines. Image adapted from SnapGene (<http://www.snapgene.com>).

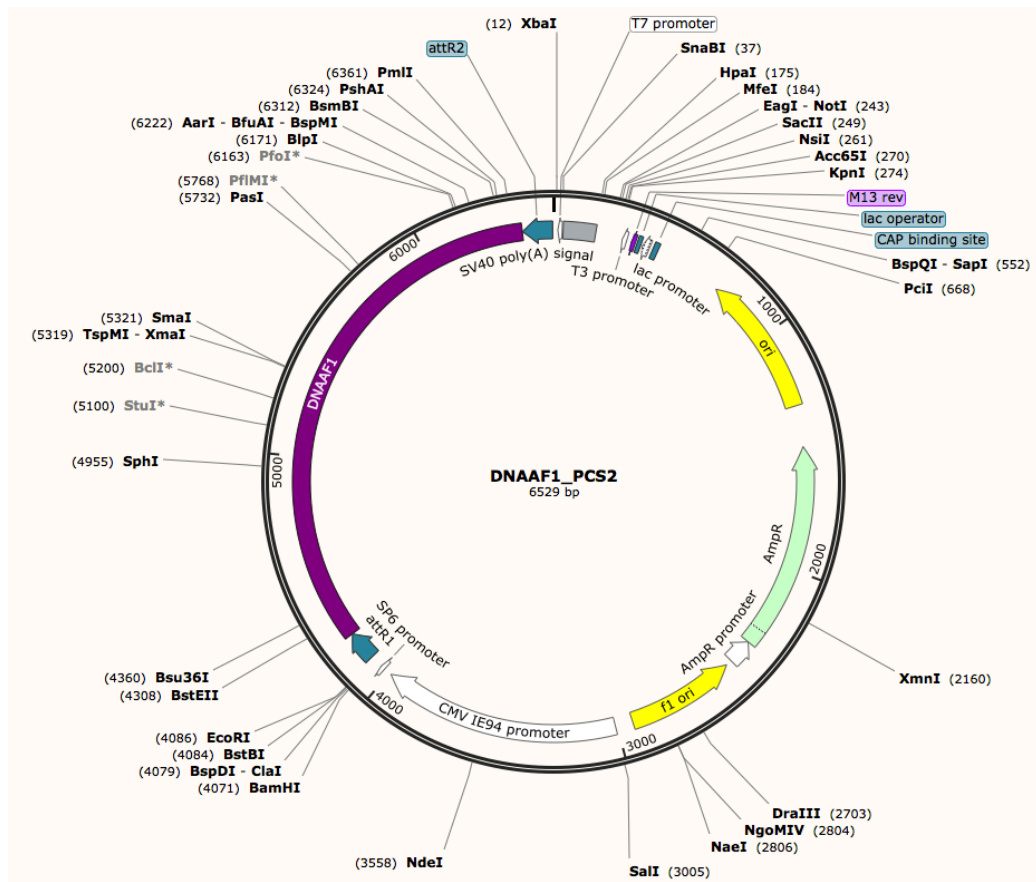


Figure 5-8: A map of the pCS2+ DNAAF1 construct.

The DNAAF1 gene is shown in purple. Restriction enzyme sites are shown by black lines. Image adapted from SnapGene (<http://www.snapgene.com>).

Prior to use in further experiments or transfer to collaborators, these constructs were Sanger sequenced to confirm orientation and absence of sequence errors from PCR. The pCS2+ constructs were linearized with the restriction enzyme *EcoRI* prior to their use as templates for *in vitro* transcription. Figure 5-9 shows agarose gel electrophoresis confirming a single band following linearization of the wildtype and mutant DNAAF1 pCS2+ plasmids.

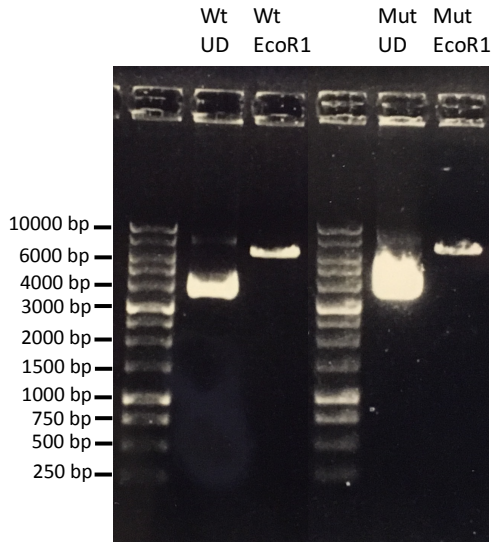


Figure 5-9: Agarose gel electrophoresis to show digestion of wildtype (wt) and mutant L191F (mut) PCS2+ constructs with EcoR1 enzyme.

UD= undigested. Bp= base pairs. Linearization of both plasmids is seen on digestion.

5.2.11 Rescue of Heart Looping Defect in *dnaaf1*^{-/-} Zebrafish

The pathogenicity of c.571C>T (Leu191Phe) missense variant identified in family CHD1 was investigated by modeling the mutation in the *Irrc50*^{-/-} (*dnaaf1*^{-/-}) null mutant zebrafish line. These experiments were performed by Dr Glenn van de Hoek, PhD student, University Medical Center, Utrecht, The Netherlands. Dr Hartill created and sequenced the pCS2+ DAAAF1 mutant and wildtype constructs required for mRNA production and Dr van de Hoek completed all of the other laboratory work, interpretation and statistical analysis.

The *dnaaf1*^{-/-} zebrafish have a T>A mutation creating a premature stop codon. They have a typical 'ciliopathy' phenotype with ventral body curvature, pronephric cysts and duct dilatation, and randomised heart looping. These fish have static cilia, previously demonstrated both at the Kupffer's vesicle and in the pronephros (Basten et al., 2013). mRNA of mutant (encoding p.Phe191) or wildtype (p.Leu191) *DNAAF1* was microinjected into embryos derived from *dnaaf1*^{+/-} heterozygote-heterozygote intercrosses at 72 hours post-fertilization and compared to mock-injected fish as a control. In mock-injected controls, around 16% of embryos showed reversed heart looping (Figure 5.10), less than the expected Mendelian ratio of 25% for a homozygous phenotype, because the laterality will be randomized (allowing 50% of affected fish to develop a left-sided heart and 50% a right-sided heart) in the mutant *dnaaf1*^{-/-} embryos (van Rooijen et al., 2008). In support of the pathogenicity of the mutation, rescue of heart looping defects was observed on injection with the wildtype *DNAAF1* mRNA, but

not the mutant (p.L191F) mRNA (see Fig. 5-10). Pair-wise comparisons were performed with the mock-injected zebrafish and the data shown is from three combined biological replicates. The background incidence of *situs* defects in wild-type zebrafish, kept under identical conditions in the same facility, was <5%.

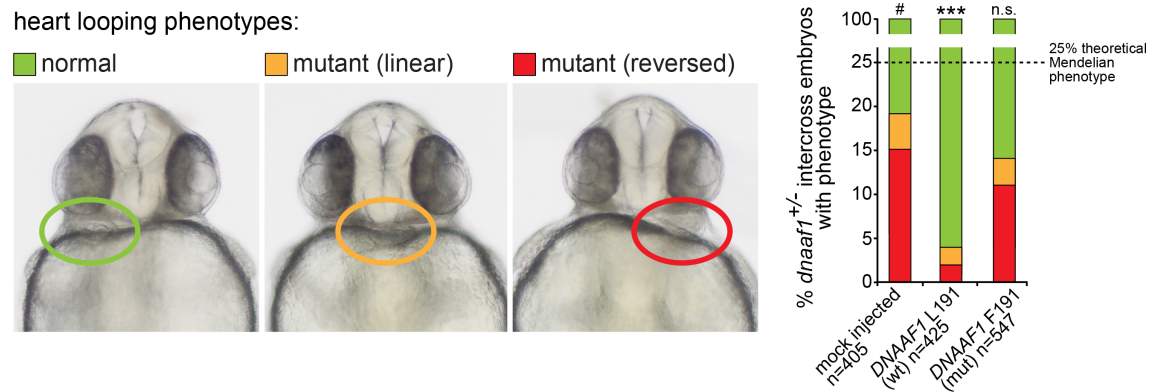


Figure 5-10: Phenotypic rescue of *dnaaf1*^{+/-} heterozygote-heterozygote intercrossed zebrafish embryos on injection of human wildtype, but not mutant L191F Dnaaf1 constructs.

Showing *dnaaf1*^{+/-} heterozygote-heterozygote intercrossed zebrafish embryos identified as control (mock-injected, n=405 embryos), or injected with mRNA expressed from human wildtype (wt, n=425 embryos) or p.Leu191Phe-mutant (mut, n=547 embryos) constructs. Left panels: Heart looping coded as normal (green), reversed (red) or 'linear' (orange), which is also an abnormal phenotype. Right panel: A graphical quantification of the three positions of heart looping following mock, Dnaaf1 wildtype (wt) or Dnaaf1 mutant (mut) injection. Phenotypic rescue is seen after injection of wildtype but not mutant mRNA. n.s.= not significant; *** p<0.001 (one-way ANOVA, n=3 biological replicates).

5.2.12 Tandem Affinity Purification to Identify Interacting Partners of Dnaaf1

The laboratory experiments described in this section were performed by Dr Kasia Szymanska, Post-Doctoral Research Fellow, University of Leeds, with data interrogation and statistical analysis performed by Dr Hartill.

Following the observation of a specific mutation in *Dnaaf1* associated with a phenotype of pure cardiac defects, but not typical PCD, we wished to gain further biological insights into the function of Dnaaf1. A Streptavidin-II/FLAG TAP experiment was performed to identify interacting partners of Dnaaf1, using wildtype and mutant (p.L191F) N-TAP constructs to pull-down peptides, which were then interrogated by mass-spectrometry. Table 5-5 shows the proteins associated with the highest number of pulled-down unique peptides in three biological replicates of this experiment, with results ordered from the highest to the lowest number of pulled-down peptides. Only those proteins with greater than 5 pulled-down peptides and present in at least two of

the biological replicates are shown. Proteins have been excluded that were observed in only one replicate, as well as those that were an effect of contamination and were present in BSA washes. For the full mass-spectrometry results see appendix L.

The peptide counts were highly reproducible between separate experiments, suggesting stable interactions. DNAAF1 pulled down itself, which validates the experimental technique, although the mutant protein was pulled down less efficiently, suggesting that the p.L191F mutation perturbed protein-protein interactions through the affected LRR repeat. As can be seen in Table 5-5, the mutant DNAAF1 appeared to have a number of novel interacting partners, not observed for the wildtype protein. Interacting partners of wildtype DNAAF1 were most commonly those involved in DNA damage/repair, mitochondrial membranes, ribosomes, vesicular trafficking, protein folding and viral infection. Of particular interest was the identification of the interaction of DNAAF1 with Heat Shock Proteins (HSPA1A, HSPA9, HSPA8 HSPA5, HSP90AA1 and HSPA6) and RUVBL1 and RUVBL2 (Pontin and Reptin), members of AAA+ (ATPases Associated with various cellular Activities) protein family. The RUVBL1 interaction, but not the interaction with RUVBL2, appeared to be perturbed by the L191F mutation, although this observation was not statistically significant (see Table 5-5).

Table 5-5: Proteins associated with peptides pulled-down in a TAP experiment, identified by mass spectrometry, using DNAAF1 wildtype or mutant (L191F) constructs as bait.

Only proteins with more than 5 pulled-down peptides in at least two of the biological replicates are shown. Results are ordered from highest to lowest number of pulled-down peptides. Log-fold change of wildtype to mutant protein is shown, where a positive value indicates an increased association with the mutant protein and a negative value indicates an increased association with the wildtype. A number of proteins show a novel interaction with the mutant construct (+) and several proteins of the chaperonin containing TCP1 (CCT) complex, also associated with the mutant protein only, have been removed from the table for brevity. The final column shows statistically significant differences in peptide numbers for wild-type compared to mutant conditions (Chi-squared test; n.s. not significant, * $p < 0.05$, **** $p < 0.0001$; for $n=3$ biological replicates. WT= wild-type)

Protein symbol	Protein function	UniProt accession no.	Protein size (kDa)	No. peptides associated with WT protein	Log-fold change (mutant vs wild-type)	Chi-squared test (mutant vs wild-type)
DNAAF1	dynein assembly factor	Q8NEP3	80	378	-0.058	*
HSPA9	heat-shock protein	P38646	74	121	-0.004	n.s.
HSPA8	heat-shock protein	P11142	71	89	+0.181	****
HSPA1A	heat-shock protein	P08107	70	86	+0.186	****
HSPA5	heat-shock protein	P11021	72	53	+0.075	n.s.
ACACA	acetyl co-A carboxylase	Q13085	266	50	-0.344	n.s.
RUVBL1	Pontin	Q9Y265	50	28	-0.049	n.s.
TUBB4B	tubulin	P68371	50	24	+0.602	****
TUBA1B	tubulin	P68363	50	20	+0.531	****
RUVBL2	Reptin	Q9Y230	51	19	+0.083	****
ACTB	beta actin	P60709	42	15	-0.030	n.s.
UBA52	ubiquitin-ribosomal protein	P62987	15	10	+0.146	n.s.
PC	pyruvate carboxylase	P11498	130	8	+0.026	n.s.
TUBB	tubulin	P07437	50	8	+0.495	****
SLC25A5	solute carrier	P05141	33	5	+0.643	****
EEF1A1	eukaryotic translation elongation factor	P68104	50	4	+0.176	n.s.
CCT6A	chaperonin containing T complex	P40227	58	3	+0.493	n.s.
ATP5A1	ATP synthase	P25705	60	3	+0.523	n.s.

Protein symbol	Protein function	UniProt accession no.	Protein size (kDa)	No. peptides associated with WT protein	Log-fold change (mutant vs wild-type)	Chi-squared test (mutant vs wild-type)
HSPA6	heat-shock protein	P17066	71	0	+	n/a
HSP90AA1	heat-shock protein	P07900	85	0	+	n/a
SUGT1	suppressor of G2 allele of SKP1	Q9Y2Z0	41	0	+	n/a

Prior to the start of this research project, our collaborators in the Netherlands had investigated additional wildtype DNAAF1 interactions using a GAL4 yeast two-hybrid assay, using an oligo dT-primed human retinal cDNA library. This screen identified only one positive clone, and it encoded intraflagellar transport protein IFT88. This result was confirmed using a “one-to-one” yeast two-hybrid experiment and DNAAF1 was found to interact with full length and N-terminally truncated IFT88, but not C-terminally truncated IFT88 (Dr S. Basten, University Medical Center, Utrecht, The Netherlands; unpublished data).

5.2.13 Co-Immunoprecipitation to Confirm Interactants of DNAAF1

Dr Mitali Patel, Post-Doctoral Research Fellow, University College London, performed the majority of experiments shown in this section, with Dr Hartill performing coIPs using a rabbit polyclonal antibody against the endogenous DNAAF1 protein. In a number of these experiments an IFT88 with C-terminal tag eYFP construct was used. This construct was obtained from Dr Pamela Tran, Harvard Medical School, Boston, Massachusetts, USA, and was received in a pLenti CMV vector (Tran et al., 2008). It was subcloned into a p.TriEx-1 vector and was sequenced verified prior to the start of this project.

Initially, co-immunoprecipitation (coIP) in HEK293 cells was used to confirm the interaction of DNAAF1 with RUVBL1 and RUVBL2, interactions identified in the TAP experiments. FLAG-DNAAF1 constructs with either the wildtype or Phe191 mutant sequence were overexpressed in this cell-line, then anti-FLAG antibodies were used for the immunoprecipitation pull-down and western blot of the membrane was performed with anti-RUVBL1 or anti-RUVBL2 antibodies, to assess interaction. Figure 5-11 shows the confirmation of the interaction of DNAAF1 with RUVBL1 and RUVBL2, both of which were pulled-down in this experiment (Figure 5-11:A & B). Neither

interaction appeared to be affected by the presence of the p.Phe191 mutation. The reciprocal co-immunoprecipitation was also attempted, with overexpression of FLAG-DNAAF1 then pull-down with anti-RUVBL1 or RUVBL2 antibody and western blot of the membrane with anti-FLAG antibody. Unfortunately, this experiment was not successful due to the anti-FLAG antibody giving a number of non-specific bands, which were not interpretable.

A similar co-immunoprecipitation was also performed with the overexpression of FLAG-DNAAF1, pull-down with endogenous DNAAF1 antibody and blot with anti-FLAG antibody and then, following stripping, with anti-RUVBL1 antibody (see Figure 5-11:C). Unfortunately no bands were seen in the IP lane, even on a number of repeats when the input protein concentration was increased. This was thought to be due to a problem with the specificity of the anti-DNAAF1 antibody. After three attempts this experiment was discontinued.

Further colPs were performed to confirm the interaction between DNAAF1 and IFT88. CoIP was performed with overexpression of IFT88-eYFP and FLAG-DNAAF1 constructs (wildtype Leu191 and mutant Phe191), followed by immunoprecipitation pull-down with anti-FLAG antibody and western blot with the proprietary anti-native GFP "Living Colors" antibody. This confirmed the interaction of IFT88 with wildtype DNAAF1. Interestingly, the interaction of IFT88 with the mutant DNAAF1 appeared to be abrogated, with a loss of the largest band after immunoprecipitation (Figure 5-11). Full blots from all colP experiments can be found in appendix M.

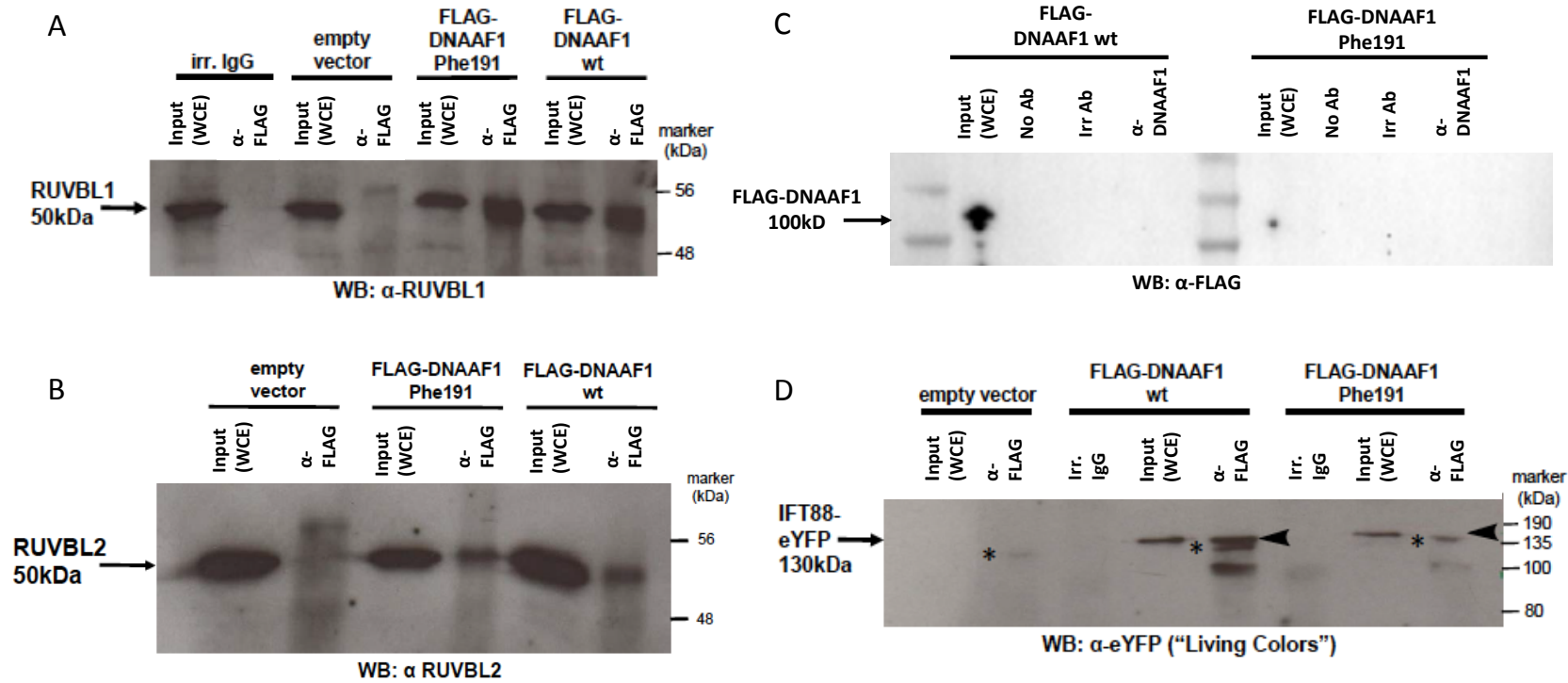


Figure 5-11: Co-immunoprecipitation assays to investigate interacting partners of DNAAF1.

A) Overexpression of wildtype or mutant (Phe191) FLAG-DNAAF1, pull-down with anti-FLAG antibody and blot with anti-RUVBL1 antibody. Controls are empty vector and irrelevant IgG controls. Input lane is the whole cell extract (WCE). The results confirm the interaction between DNAAF1 (wt and mut) and RUVBL1/Pontin. B) Overexpression of wildtype or mutant (Phe191) FLAG-DNAAF1, pull-down with anti-FLAG antibody and blot with anti-RUVBL2 antibody. An empty vector control is shown. The results confirm the interaction between DNAAF1 (wt and mut) and RUVBL2/Reptin C) Overexpression of wildtype or mutant (Phe191) FLAG-DNAAF1, pull-down with endogenous anti-DNAAF1 antibody and blot with anti-FLAG antibody. No bands are seen in the immunoprecipitation lanes. D) Overexpression of wildtype or mutant (Phe191) FLAG-DNAAF1 and IFT88-eYFP, pull-down with anti-FLAG antibody and blot with anti-GFP "living colours" antibody. Irrelevant IgG lanes and empty vector are shown as a control. The interaction of wildtype DNAAF1 with IFT88 is indicated by an arrowhead and is lost in the mutant DNAAF1 experiment. The presence of non-specific bands are indicated by asterisks.

5.2.14 Cellular Localization Studies to Assess Co-localisation of DNAAF1 with Identified Interacting Partners

We next assessed the co-localization of DNAAF1 and the identified interacting partners from previous experiments: RUVBL1, RUVBL2 and IFT88. Initially the localization of DNAAF1 was assessed in the primary-ciliated cell-lines, human hTERT-RPE1 and mouse IMCD3s. DNAAF1 was found to localize to the base of cilia in these cell lines using an endogenous DNAAF1 antibody and an antibody that recognized acetylated alpha tubulin as a ciliary marker. To further characterize this localization, immunofluorescence confocal microscopy using antibodies against RPGRIP1L (recognizing the ciliary transition zone), CEP290 (ciliary transition zone), Centrin-3 (distal lumen of the centrioles) and γ -tubulin (ciliary basal body) was performed in hTERT-RPE1 cells. Endogenous DNAAF1 was identified close to RPGRIP1L, CEP290 and Centrin-3, but overlapping with γ -tubulin, suggesting a specific basal body localization (Fig 5-12).

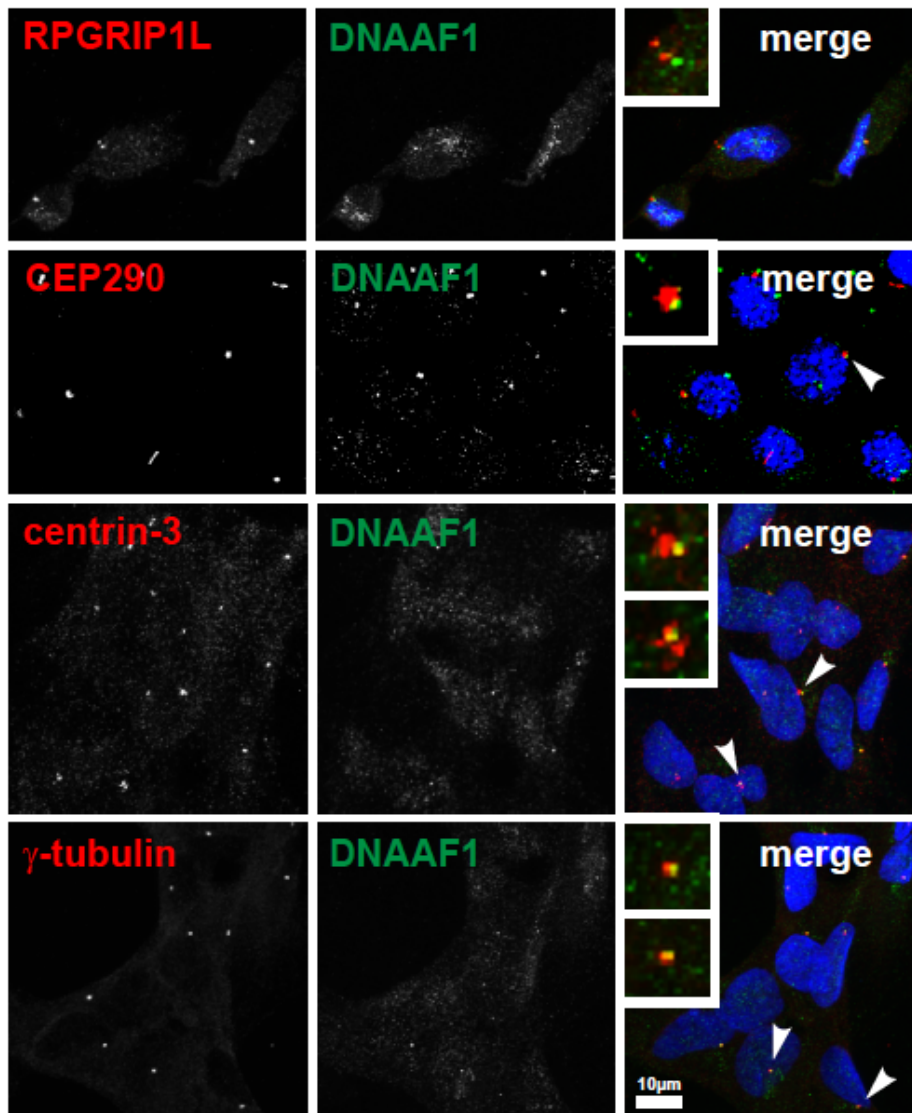


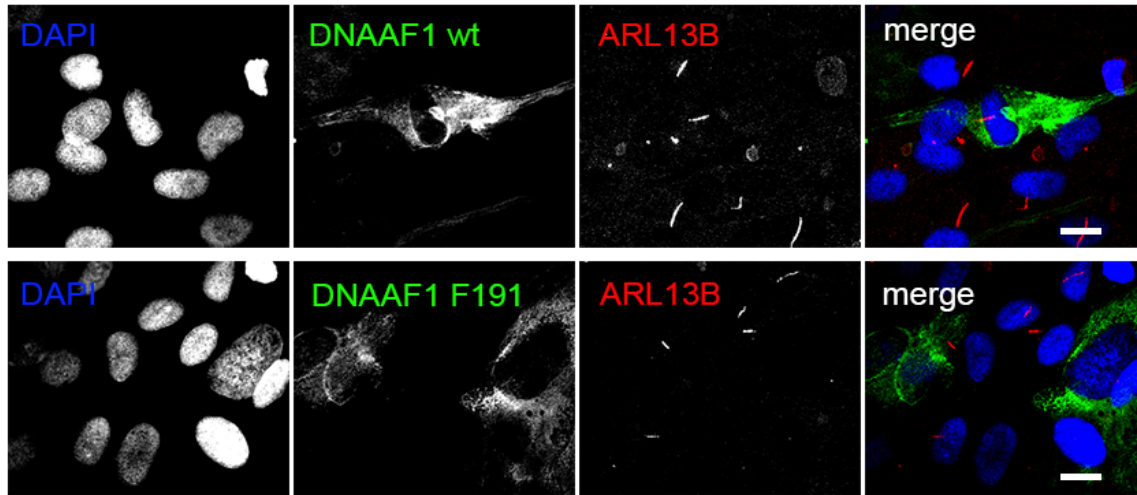
Figure 5-12: Co-localization of endogenous DNAAF1 (green) and ciliary markers, RPGRIP1L, CEP290, Centrin-3 and γ -tubulin (red).

DNAAF1 shows specific basal body localization. Scale bars = 10 μ m.

To gain further insights into the observed DNAAF1 missense mutation, experiments were performed with overexpression of wildtype and mutant (p.Leu191Phe) DNAAF1 N-terminal FLAG constructs (FLAG-DNAAF1) and immunofluorescence with anti-FLAG antibody in hTERT-RPE1 cells. The FLAG-DNAAF1 protein accumulated mainly in the cell cytoplasm, however there was specific localization of FLAG-tagged protein at the basal body, which was particularly obvious in cells where there was a lower level of transfection. Figure 5-13 shows RPE1 cells with overexpression of the constructs, co-stained with ARL13B as a ciliary membrane maker. The overexpressed wildtype and mutant constructs demonstrated similar levels of immunofluorescence intensity, as

demonstrated by assessing Corrected Total Cell Fluorescence (CTCF) analysis in ImageJ, suggesting no excessive degradation of the mutant protein (Fig. 5-13).

A



B

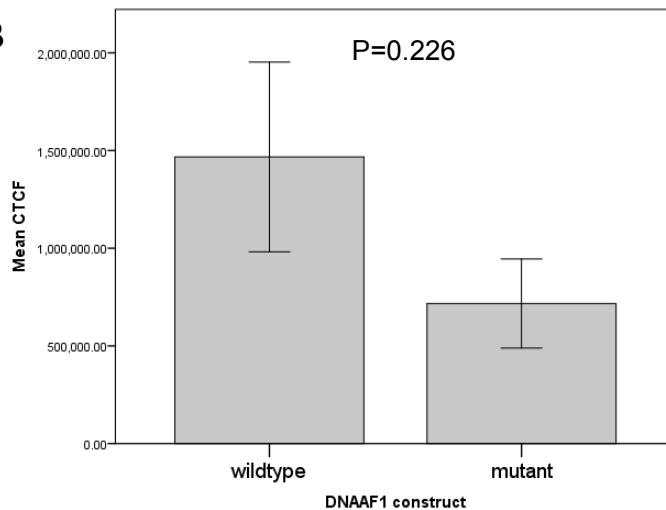


Figure 5-13: Immunofluorescence of overexpressed wildtype or mutant F191 FLAG-DNAAF1 in RPE1 cells.

A) IF of RPE1 cells with overexpression of FLAG-DNAAF1 wildtype or mutant F191, co-stained with ARL13B. Expression of DNAAF1 is grossly cytoplasmic, but co-localization to the basal body is visualized. Scale bars = 10 μ m. B) Comparison of Corrected Total Cell Fluorescence (CTCF) was assessed in RPE1 cells transfected with wildtype or mutant (F191) DAAAF1 constructs using analysis by ImageJ. A total of 40 cells were assessed. The P value is the result of Mann-Whitney-U test. Error bars represent standard error of the mean.

This experiment was repeated with overexpression of the FLAG-DNAAF1 constructs and immunofluorescence staining for endogenous IFT88, and a ciliary marker, γ -tubulin (Fig. 5-14). A clear co-localization was noted between DAAAF1, IFT88 and γ -tubulin. On quantitative assessment, the localization of IFT88 to DAAAF1 was not affected by the presence of the mutant construct, with a consistent finding of approximately 20% of IFT88 spots, per field of view, co-localized to the overexpressed DAAAF1 protein (Fig.

5-14). This was further assessed by overexpression of eYFP-IFT88 alongside FLAG-DNAAF1 and staining for the ciliary marker ARL13B. Again, co-localisation between DNAAF1 and IFT88 was observed. The co-localization of DNAAF1 to IFT88 was, similarly, not affected by the presence of expression of the mutant protein. Under these conditions, around 80% of cilia, per field of view, showed co-localisation of FLAG-DNAAF1 to IFT88 (Fig. 5-14). In the presence of overexpressed IFT88, the cilia marker ARL13B appeared to be disrupted (Fig. 5-14), with more diffuse staining than previously seen, perhaps due to the effect of excessive IFT88 on the correct targeting of ciliary proteins such as ARL13B or on ciliogenesis.

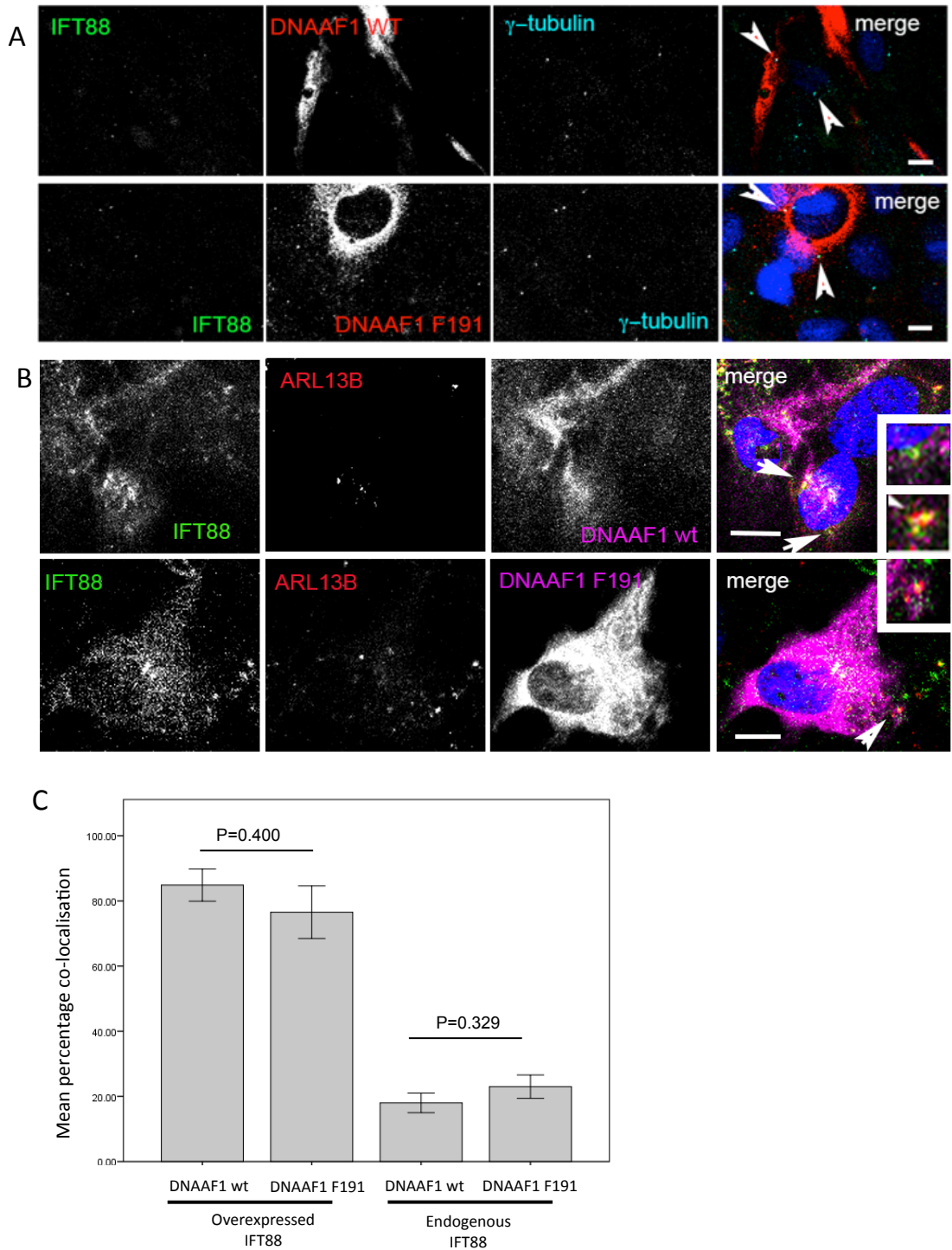


Figure 5-14: Immunofluorescence of overexpressed wildtype or mutant F191 FLAG-DNAAF1 in RPE1 cells and co-localisation with endogenous and overexpressed IFT88.

A) hTERT-RPE1 cells with overexpression of either wildtype or mutant FLAG-DNAAF1 (wt or mutant) and IF for anti-FLAG antibody (red), anti-IFT88 (green) and anti- γ -tubulin (cyan). B) hTERT-RPE1 cells with overexpression of either wildtype or mutant FLAG-DNAAF1 (wt or mut) and IFT88-eYFP (green) and IF for anti-FLAG antibody (magenta) and anti-ARL13B antibody (red). Scale bars = 10 μ m C) Mean percentage IFT88 co-localization with wildtype or mutant (Phe191) DNAAF1 when assessing overexpressed or endogenous IFT88. P values indicate the results of Student's T-test, comparing values for wildtype and mutant (F191) constructs under the conditions of staining for over-expressed IFT88 or endogenous IFT88. Error bars show the standard error of the mean.

Once RUVBL1 and RUVBL2 (Pontin and Reptin) were identified as interactors of DNAAF1, immunofluorescence was used to assess their localization. Both Pontin and Reptin had predominantly nuclear staining and there was overlap of Pontin/Reptin and DNAAF1 staining within the nucleus (Fig. 5-15).

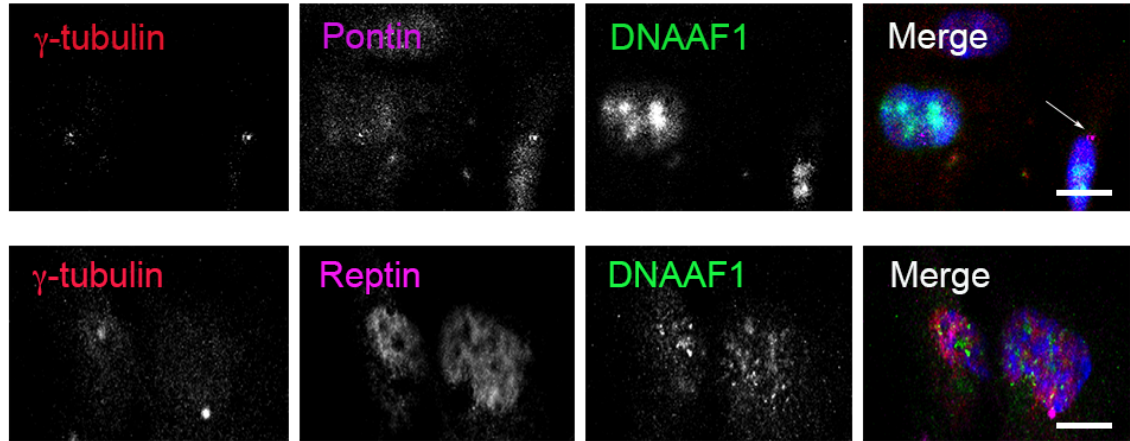


Figure 5-15: Immunofluorescence to show co-localisation of endogenous DNAAF1 and Pontin/Reptin in RPE1 cells.

IF of hTERT-RPE1 cells stained for anti-Pontin (RUVBL1) or Reptin (RUVBL2) (magenta), γ -tubulin (red) and DNAAF1 (green). Scale bars = 10 μ m. Co-localization of DNAAF1 and RUVBL1/2 was not consistently seen at the basal body, but there was nuclear overlap of staining.

5.2.15 siRNA Knockdown of Pontin (RUVBL1) and Reptin (RUVBL2)

Due to the identified interaction between DNAAF1 and Pontin and Reptin, we hypothesized that this interaction was part of a novel function of a Pontin/Reptin complex involved in dynein arm assembly. We theorized that knockdown of Pontin or Reptin might affect the localization of DNAAF1 to the basal body or to IFT88. Pontin or Reptin knockdown using siRNAs was performed in hTERT-RPE1 cells with concurrent overexpression of wildtype or mutant (Phe191) FLAG-DNAAF1 constructs. Previous qPCR experiments performed by Dr Sander Basten, University Medical Center, Utrecht, had shown successful knockdown of Pontin and Reptin in T47D cells, using these same siRNAs, under identical conditions (Fig. 5-16:A). In our experiments, western blot was used to confirm the knockdown of these proteins in RPE1 cells. The results showed a successful Pontin knockdown, although Reptin knockdown was only to 80% of wildtype, so only Pontin knockdown was continued in further experiments (Fig. 5-16:B).

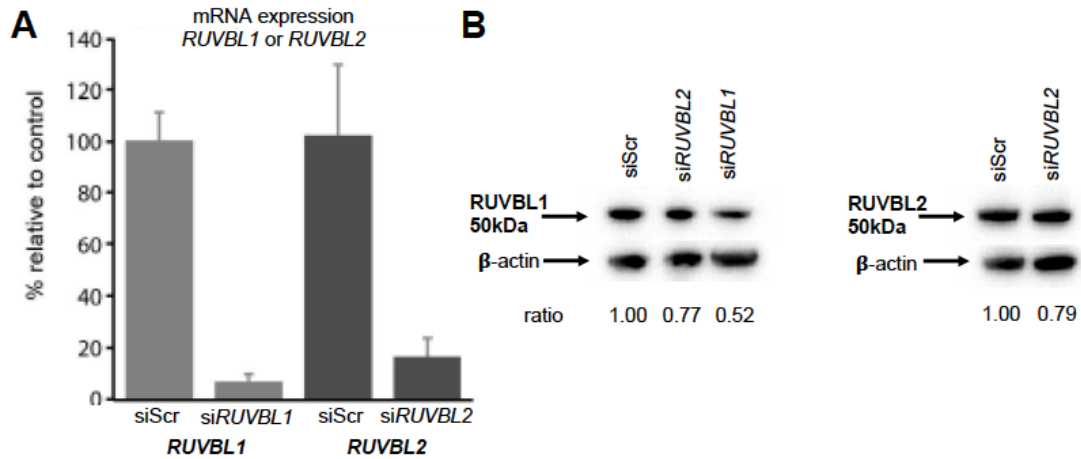


Figure 5-16: qPCR and western blot to demonstrate knockdown of RUVBL1/2 using siRNA.

A: qPCR to demonstrate siRNA knockdown of RUVBL1 (Pontin) and RUVBL2 (Reptin) in T47D cells. Scrambled siRNA (siScr) was used as a negative control. Experiment performed by Dr Sander Basten, University Medical Center, Utrecht. B: Western blot to demonstrate the effect of siRNA knockdown of Pontin or Reptin in RPE1 cells. β -actin was used as a loading control and relative reduction in expression (ratio) was compared to siScrambled knockdown.

Immunofluorescence microscopy was used to assess the localization of overexpressed FLAG-DNAAF1, endogenous IFT88 and γ -tubulin in these Pontin knockdown cells. As previously noted, there was clear localization between IFT88, γ -tubulin and DNAAF1 (both wildtype and mutant) in the scrambled siRNA transfected cells. This co-localization was also present in the Pontin siRNA treated cells (Fig. 5-17:A). On quantification, there was marginal disruption of co-localization between DNAAF1 and IFT88 in the Pontin knockdown cells ($p= 0.059$, Student's t-test, $n=2$ biological replicates), however the co-localisation of IFT88 to γ -tubulin remained unaffected (Fig. 5-17:B).

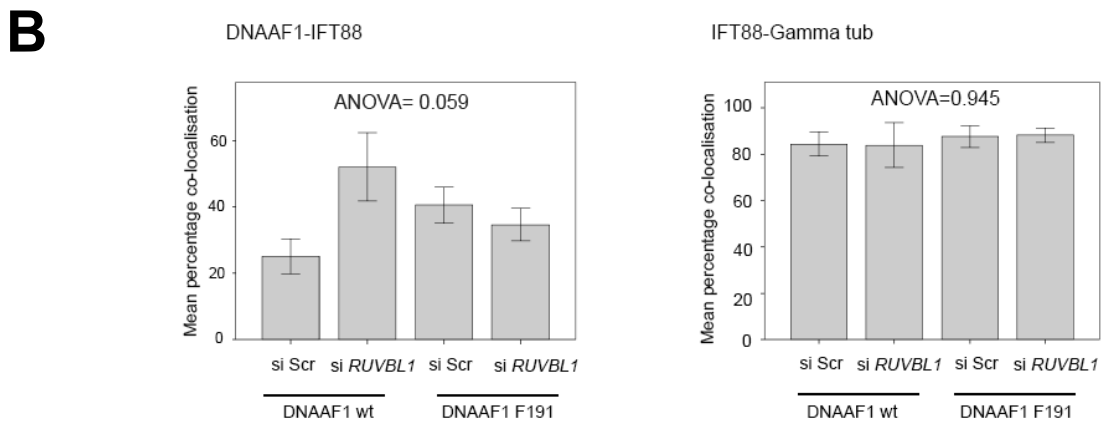
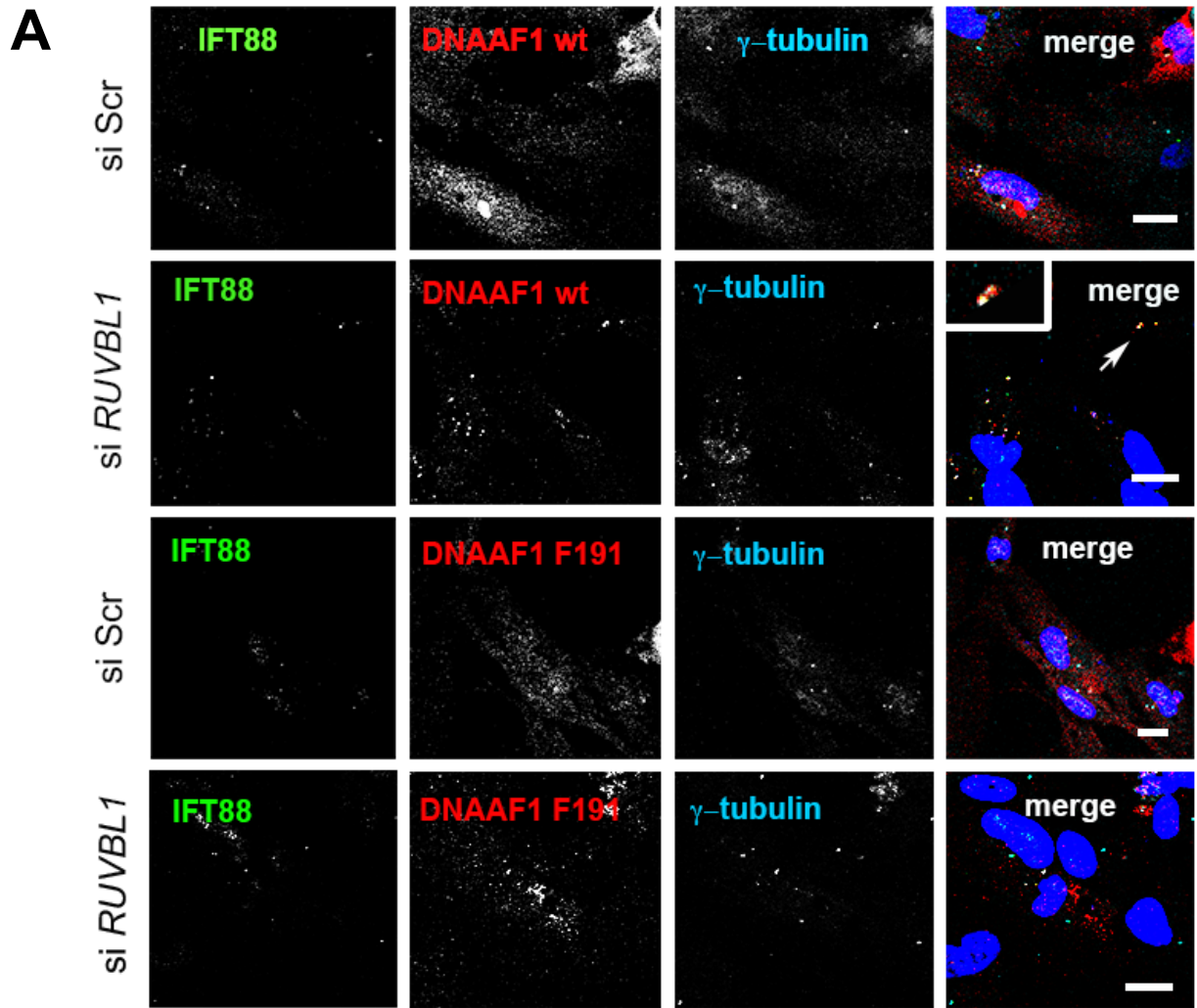


Figure 5-17: Results of siRNA knockdown of Pontin (RUVBL1) in RPE1 cells and affect on co-localisation of DNAAF1 to IFT88 and IFT88 to γ -tubulin.

A: Immunofluorescence microscopy to demonstrate co-localization of FLAG-DNAAF1 wildtype or mutant (F191) to IFT88 and γ -tubulin in the presence of scrambled siRNA negative control (siScr) or siRUVBL1 (Pontin) knockdown. Scale bars = 10 μ m. B: Quantification of the effect of siRUVBL1 knockdown on co-localization of DNAAF1 to IFT88 and IFT88 to γ -tubulin, in the presence of wildtype or F191 DNAAF1 overexpression. Statistical analysis is a one-way ANOVA. Error bars represent standard error of the mean. n= two biological replicates.

5.2.16 Mouse Embryo Whole Mount Immunofluorescence

The work described in this section was performed in close collaboration with Ms Dina Abdelmottaleb, Visiting Researcher, Leeds Institute of Molecular Medicine, Leeds, UK.

Due to the finding of pure heterotaxy in a family with a mutation in *DNAAF1*, we sought to further investigate the expression of DNAAF1 during murine cardiac development and at the mouse embryonic node. The localisation of DNAAF1 and RUVBL1 (Pontin) was assessed in wildtype BL6 mouse embryos using whole mount immunofluorescence on a Zeiss L1 light-sheet microscope. Both proteins were present in the E8.5 mouse heart. Pontin appeared to be present in a widespread distribution throughout the primitive ventricle and atrial chambers. DNAAF1 localisation was observed at the primary cilia in the developing cardiac cells, with one spot of DNAAF1 visualised per cell, presumably associated with the basal body (see Figures 5-18 and 5-19).

E7.5-E7.75 mouse embryos were assessed to visualise expression of these proteins at the murine embryonic node. DNAAF1 was expressed at the node and in the surrounding tissues, likely associated with primary and nodal cilia. Pontin expression was seen at the node itself, but not in the surrounding tissues. Interestingly the Pontin expression was always to the left side of the node and this striking left-sided expression was consistently observed in three replicates (Figures 5-18 and 5-19).

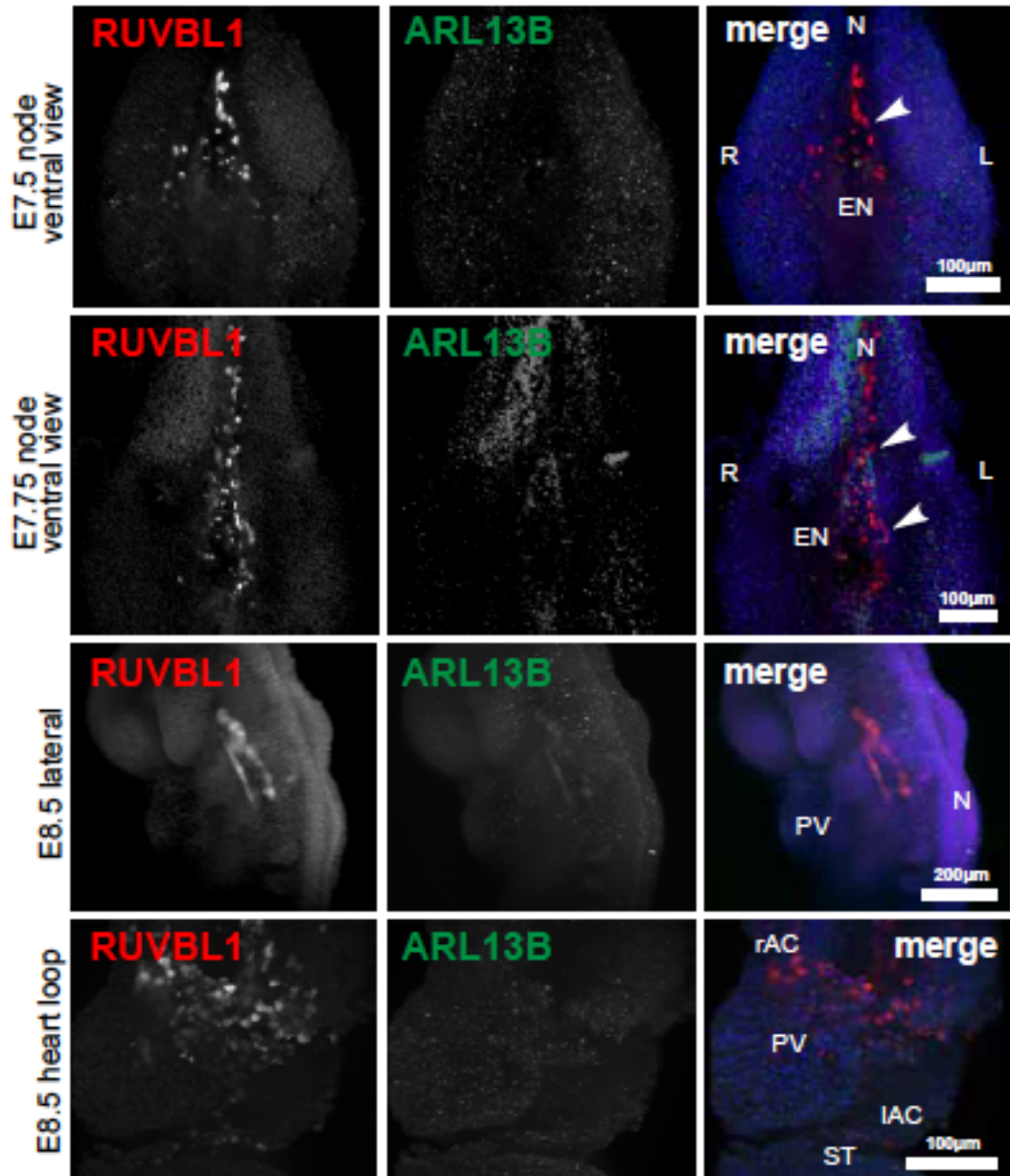


Figure 5-18: Light-sheet microscopy images showing immunofluorescence of RUVBL1 and ARL13B in staged mouse embryos.

A series of maximum intensity projections of light-sheet microscopy images to demonstrate RUVBL1/Pontin (red) and ARL13B (green) whole mount immunofluorescence in the E7.75 murine embryonic node and E8.5 heart loop. RUVBL1 is asymmetrically distributed at the embryonic node, with a predominantly left-sided expression (upper two panels, see arrow-heads). RUVBL1 is present in a widespread distribution throughout the atrial chambers and primitive ventricle (lower panels). EN= embryonic node, N= notochord, rAC and IAC = atrial chamber, PV= primitive ventricle and, ST= Septum Transversum. Scale bars = 100 or 200 μm , as indicated.

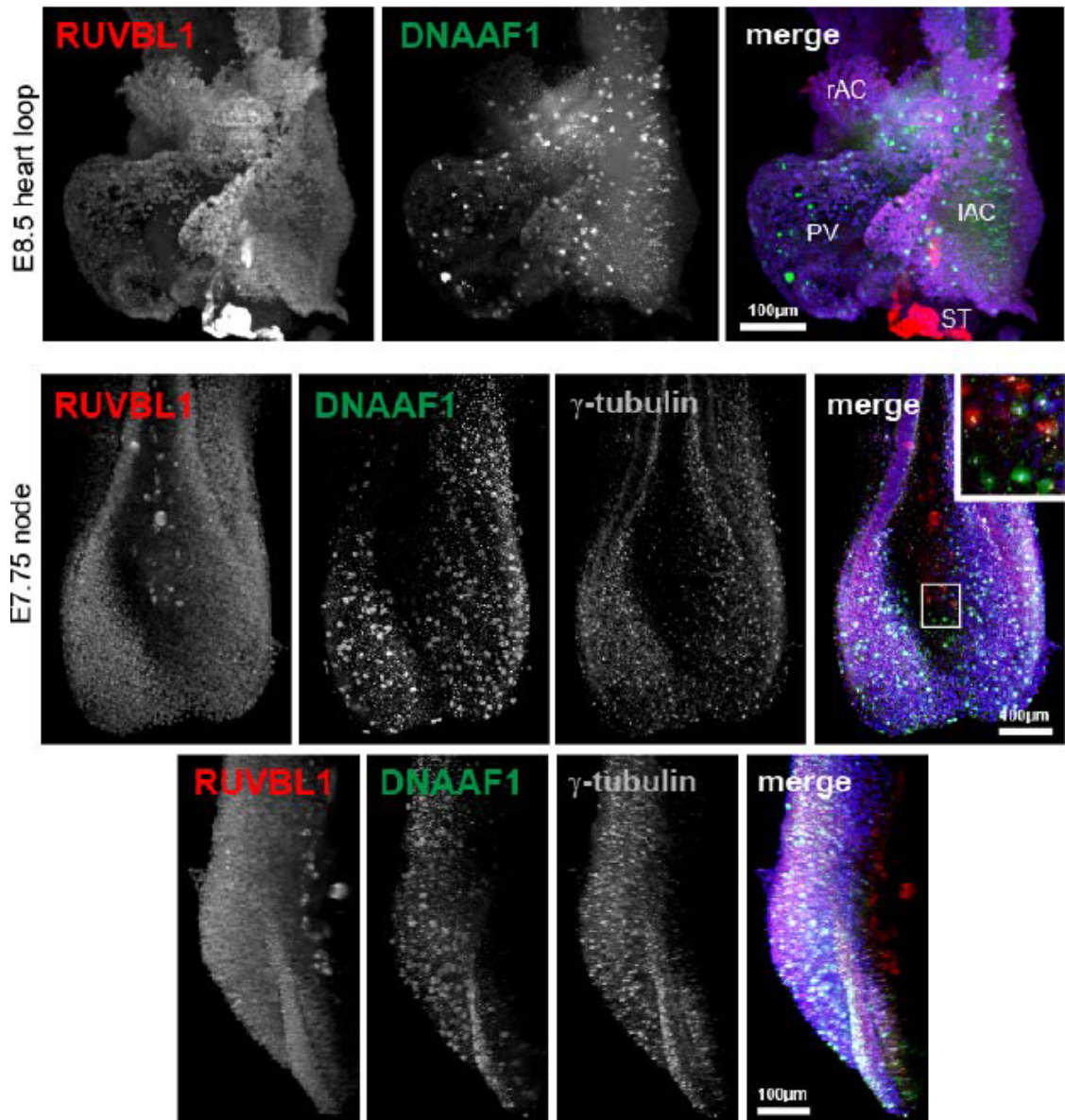


Figure 5-19: Light-sheet microscopy images showing immunofluorescence of RUVBL1 and DNAAF1 in staged mouse embryos.

A series of maximum intensity projections of light-sheet microscopy images to show RUVBL1/Pontin (red), DNAAF1 (green) and γ -tubulin (grey) whole mount immunofluorescence in the E8.5 heart loop (upper panels) and E7.75 murine embryonic node (lower two panels). rAC and IAC = atrial chamber, PV= primitive ventricle and, ST= Septum Transversum. Scale bars = 100 μ m.

5.2.17 Expression of DNAAF1 and Pontin at Kupffer's Vesicle in Zebrafish Embryos

The work in this section was performed by Mr Glenn van de Hoek, PhD student, University Medical Center, Utrecht, The Netherlands. Mr van de Hoek performed all of the laboratory work, confirmations, interpretation and statistical analysis.

Due to the observation of a left-sided distribution of RUVBL1/Pontin in the mouse embryonic node, the *dnaaf1*^{-/-} zebrafish were studied to assess Pontin distribution at the Kupffer's Vesicle (KV). Again, the distribution was found to be mainly left-sided in the wildtype zebrafish. However in the mutant *dnaaf1*^{-/-} zebrafish, this left-sided expression was abolished. This indicates that Pontin may play a role in the formation of asymmetry in the early embryo, and when nodal cilia are affected by *dnaaf1* mutation, this asymmetrical expression is lost (Figure 5-20).

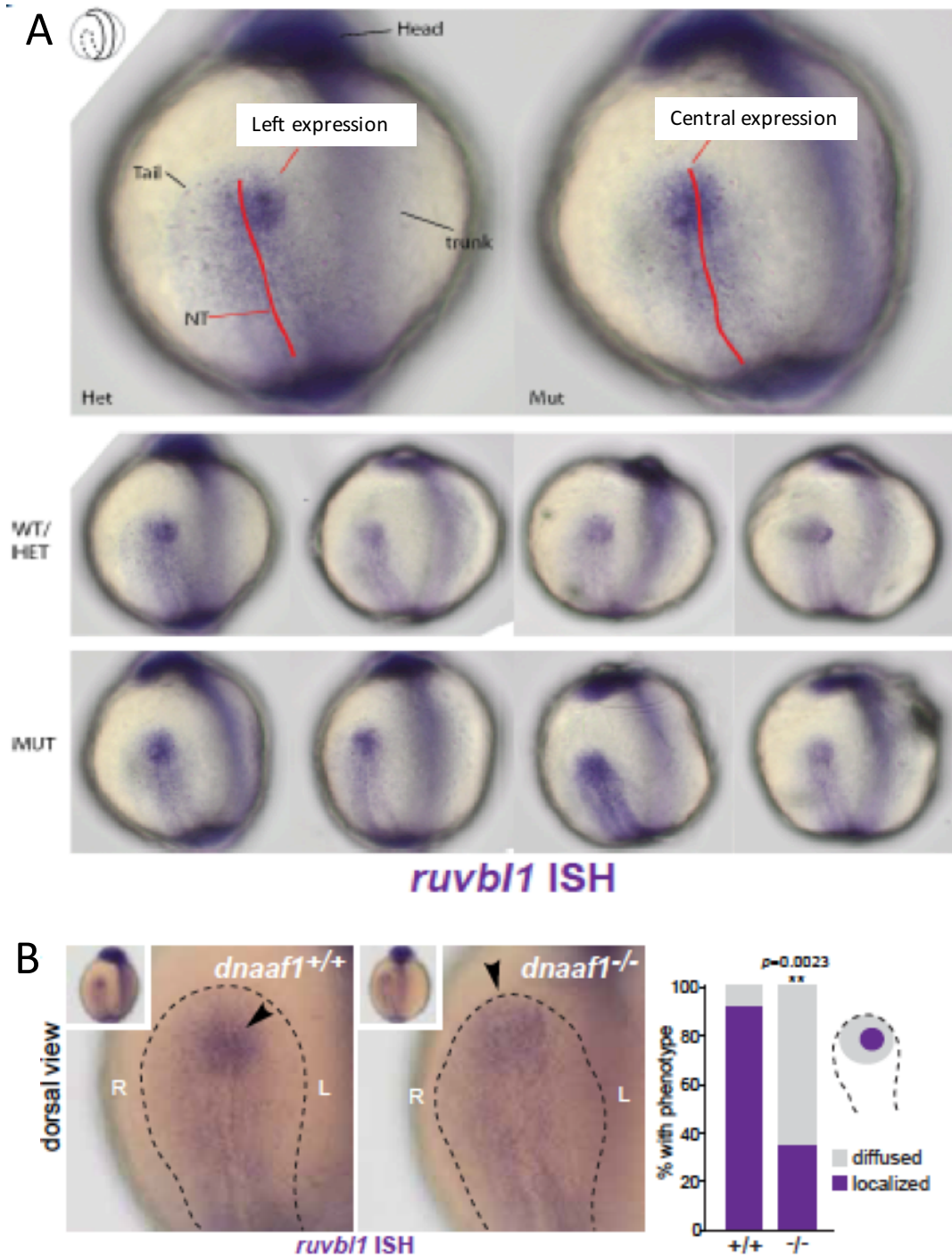


Figure 5-20: Pontin (*ruvb1*) whole mount in situ hybridization in zebrafish embryos.

A: Pontin (*ruvb1*) whole mount in situ hybridization in zebrafish embryos. Wildtype or heterozygous embryos show a predominantly left-sided expression of Pontin, and mutant *dnaaf1*^{-/-} embryos show a central expression. B: Quantification of *ruvb1* expression in wildtype and *dnaaf1*^{-/-} zebrafish embryos. A significantly higher proportion of wild-type zebrafish have left-sided localization of *ruvb1* compared to mutant embryos ($p=0.0023$, statistics are results of Student's t-test). Experiment performed by Mr Glenn van de Hoek.

5.3 Discussion

5.3.1 Discussion of Candidate Variants in Family CHD1

In family CHD1, the variant in *DNAAF1*, c.571C>T, p.L191F, was the only variant that passed all filters and was not present in a homozygous state in control databases. It appeared to be a good candidate for the phenotype, scoring highly on *in silico* bioinformatics tools and occurring at a highly conserved position of the protein. It also fell in a large (2,500,000 bp) homozygous region shared between the two affected siblings. *DNAAF1* is a known cause of Primary Ciliary Dyskinesia (PCD; MIM 244400) (Loges et al., 2009, Duquesnoy et al., 2009) and as there is a known link between PCD and CHD, it was considered a likely candidate gene in this family, although the clinical phenotype was not consistent with typical PCD.

5.3.2 Primary Ciliary Dyskinesia is a Disease of the Motile Cilia

Cilia were previously introduced as microtubule-based organelles that play a role in sensing chemical, morphogenic and mechanical stimuli (see section 1.6.1). Primary cilia are present on the majority of vertebrate cell types (Satir and Christensen, 2007). Motile cilia are a specialized type of cilia that exist on certain cell types, including those of the respiratory tract, Fallopian tubes and middle ear, where they beat in a rhythmical pattern to fulfill various functions, for example movement of mucus in the respiratory tract or movement of the ovum within the Fallopian tube. A further type of cilia exist at the embryonic node, “nodal cilia”, which are motile, and are involved in the creation of a fluid flow, which is a key step in breaking symmetry in the early embryo (see Chapter 1 section 1.5.2.3) (Norris, 2012). The nodal cilia also have differences from other motile cilia, since structurally they lack the inner microtubular doublets, giving them a 9+0 arrangement, but their molecular differences are not well understood (Hirokawa et al., 2009).

Primary Ciliary Dyskinesia (PCD) is a disease of the motile cilia, causing impaired mucociliary clearance in the respiratory tract, leading to neonatal respiratory distress, chronic respiratory tract infections, bronchiectasis, sinusitis and otitis media (Horani et al., 2014). Other associated features include infertility in both males and females, due to motile ciliary dysfunction in spermatozoa and the Fallopian tube. PCD is a predominantly an autosomal recessive disease with an incidence of 1:15000 (Katsuhara et al., 1972).

5.3.3 The Link between Primary Ciliary Dyskinesia and Congenital Heart Disease

As introduced in Chapter 1, the link between PCD and CHD has long been recognized. Around 50% of PCD patients have *situs inversus totalis*, a total reversal of body anatomy, due to defects in the nodal cilia and therefore dysfunction in development of laterality in the early embryo. As well as *situs inversus*, both heterotaxy and complex CHD are also more common in PCD than in the general population. In one study, the incidence of heterotaxy was found to be around 6% in a cohort of more than 300 PCD patients, including 8 patients with complex CHD lesions (Kennedy et al., 2007). Mouse models of PCD show even higher rates of CHD, perhaps representing a proportion of unrecognised early pregnancy loss in human studies (Francis et al., 2012). A recent study which described a large murine forward genetic screen for CHD aetiology, revealed a high mutational burden in cilia and PCD-related genes (Li et al., 2015).

As well as CHD and heterotaxy being prominent features of PCD, it appears that, conversely, ciliary dysfunction is common in heterotaxy patients. Nakhleh et al., assessed a cohort of over 40 patients with pure heterotaxy, and no previous diagnosis of PCD, and identified ciliary dysfunction in 42% of patients. An increased burden of variants in PCD genes was identified in this cohort. One had a single known *DNAI1* founder mutation, but none of the patients was found to have two pathogenic mutations in a PCD gene. This high burden of single PCD gene variants may indicate an increased mutational load to play a role in the aetiology of heterotaxy in these patients and the authors suggest that pre-screening heterotaxy patients for ciliary dysfunction may be a useful clinical tool prior to cardiac surgery (Nakhleh et al., 2012).

5.3.4 The Known Functions of DNAAF1 and Its Role in PCD

Motile cilia (except those present at the embryonic node) have a structure consisting of nine microtubule doublets surrounding a central pair (9+2). In motile cilia, the force required for motility is provided by molecular motors, which convert ATP into molecular energy. These motors are the inner and outer dynein arms (IDA and ODA), which connect to the microtubule doublets and drive their sliding past each other for motion (Nicastro et al., 2006) (see Fig. 5-21). It is considered that ODAs and IDAs provide slightly different functions in motility: ODAs are responsible for maintaining ciliary beat frequency and IDAs are required for the creation of a normal waveform. The dynein arms are incorporated into cilia as pre-assembled building blocks and this preassembly occurs in the cytoplasm, prior to transport to the ciliary axoneme. ODAs have a

complex structure and in human respiratory cilia ODAs consist of a globular head domain (heavy chain beta (DNAH11/DNAH9) and heavy chain gamma (DNAH5)), an intermediate domain (DNAI1, DNAI2) and a docking complex (CCDC114, CCDC151, ARMC4, TTC25), involved in anchoring the arms to the microtubule doublets (Nicastro et al., 2006). Multiple light chains are associated with this complex, required for motility and stability (Zariwala et al., 2007). The outer and central microtubules are connected by radial spokes (see Fig. 5-21), which provide support and were recently described to play an important role in the signal transduction pathway (Zariwala et al., 2007). PCD is known to be caused by mutations in 19 genes, all encoding proteins involved in parts of this complex machinery; radial spoke proteins, components of the inner and outer dynein arms and assembly factors (Horani et al., 2016).

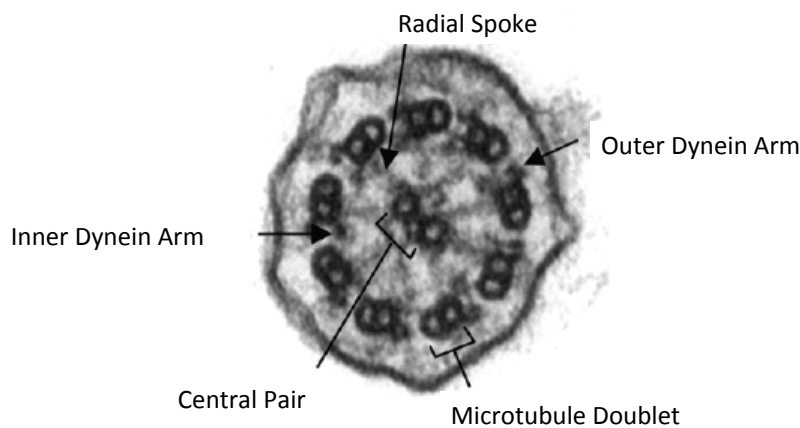


Figure 5-21: Transmission Electron Micrograph (TEM) of a human respiratory motile cilium.

Figure showing the inner and outer dynein arms, radial spokes, central microtubules and outer microtubule doublets. Adapted from reference (Shoemark et al., 2013).

DNAAF1 is orthologous to *ODA7* in *Chlamydomonas*. *ODA7* was initially identified as a dynein arm assembly factor with mutations in this gene causing a reduced ciliary beat frequency and block in ODA assembly in *Chlamydomonas* (Freshour et al., 2007). *Chlamydomonas* ODA globular head domains have a different structure to their human counterparts, consisting of heavy chain-alpha, heavy chain-beta and heavy chain-gamma subunits, plus an intermediate domain and docking complex. In *Chlamydomonas* with *ODA7* mutations, there is a build up of outer arm subunits in the cytoplasmic pool (Mitchison et al., 2012), with a specific increase in the intermediate domain protein DNAI2. Furthermore, in these mutants there is a reduction in heavy chain-beta and heavy chain-gamma and almost a complete loss of heavy chain-alpha in the axoneme (Mitchison et al., 2012). *ODA7* is therefore thought to specifically

regulate heavy chain- α stability in these organisms (Fowkes and Mitchell, 1998), perhaps being involved in the folding of the globular dynein head domain (Mitchison et al., 2012), but may also play a role in providing a structural link between inner and outer row dyneins (Freshour et al., 2007, Fowkes and Mitchell, 1998). In human respiratory cell lines, mutations in *DNAAF1* cause an absence of specific dynein heavy (DNAH5, DNAH9), intermediate (DNAI2) and light chains (DNALI1) from the ciliary axoneme (Loges et al., 2009). However, the mechanism by which *DNAAF1* mediates ODA pre-assembly and targeting of dyneins to the axoneme is unknown.

In human patients with mutations in *DNAAF1*, there are combined ODA and IDA defects, consistent with *in vitro* models (Loges et al., 2009). Furthermore, zebrafish *dnaaf1*^{-/-} mutants have abnormalities of the dynein arms, lacking either ODA, or both IDA and ODA, and have misalignment of the outer microtubules, indicating instability in the microtubular architecture (van Rooijen et al., 2008, Basten et al., 2013).

5.3.5 A Phenotypic Continuum Between CHD and PCD

In this study we have identified a missense mutation in *DNAAF1*, p.L191F, associated with a phenotype of pure CHD, without a clinical diagnosis of PCD in family CHD1. In this family, NO levels and ciliary beat frequency were within normal limits and did not support a clinical diagnosis of PCD. Furthermore, TEM revealed no inner or outer dynein arm defects. This is a novel observation. The three other families described in this study, with truncating mutations or deletions in *DNAAF1* (families 1b-d), had the typical clinical symptoms of PCD and diagnostic tests, such as exhaled Nitric Oxide levels. Defects in ciliary beat frequency confirmed this diagnosis. TEM gave variable results in these families: in family 1b ODA were mainly absent, and IDA were absent in 20-30% of cilia. In family 1d, IDA were consistently absent and only ODA stubs were present. Similarly, in the previously published seven cases of patients with *DNAAF1* mutations in the literature, IDA and ODA were consistently absent and the phenotype was consistent with typical PCD (Loges et al., 2009, Duquesnoy et al., 2009).

Our findings suggest that the truncating mutations in families 1b, c and d are associated with a full PCD phenotype, whereas the missense mutation in family CHD1, p.L191F, is hypomorphic and is associated with a predominant phenotype of cardiac disease. However, missense mutations in the same region of the gene, for example p.L175R, have been shown to cause the full PCD phenotype in other studies (Basten et al., 2013). There is *in silico* support for the pathogenicity of p.L191F, and this assumption is strongly supported by absence of genetic complementation in *dnaaf1*^{-/-}

mutant zebrafish embryos. To explain this unexpected finding, we hypothesize that there are subtle defects in motile ciliary function present in family CHD1, which were not identified on clinical testing. These subtle defects may be sub-clinical in terms of their effect on respiratory function, but caused disruption of nodal ciliary function, therefore affecting correct cardiac looping, either due to the sensitive nature of the ciliary beat at the node during embryonic development, or due to the inherent differences between nodal and respiratory motile cilia, which are, as of yet, not well described. In this instance, the missense mutation may affect unrecognised DNAAF1 functions, distinct from its known role in the assembly of dynein arms. Our findings suggest that CHD and PCD form a phenotypic continuum rather than comprising distinct clinical entities. Patients with complex CHD may therefore harbour mutations in PCD genes, even if respiratory ciliary function is apparently normal. Identification of these genetic aetiologies may help to refine the medical and surgical management of patients with these complex conditions, which carry a high burden of morbidity and mortality (Moller and Anderson, 2013).

5.3.6 Novel Interactants of DNAAF1 and Insights into its Function

To further understand the role of DNAAF1 in the assembly of motile ciliary dynein arms, we investigated if particular functions were being affected by the missense mutation identified in family CHD1. The results of our TAP and co-immunoprecipitation experiments have demonstrated DNAAF1 to interact with novel proteins, including RUVBL1 and RUVBL2 (Pontin and Reptin), intraflagellar transport protein IFT88 and members of the Heat Shock Protein (HSP) group.

IFT88 is an intraflagellar transport protein that is a member of the IFT-B complex. IFT88 is important in the transport of large cargo molecules along the axonemal microtubules of primary and motile cilia, required for their assembly and maintenance (Pazour et al., 2002). Homozygous mutations in *IFT88* cause a Meckel-Gruber syndrome-like ciliopathy (McIntyre et al., 2012), and mice with mutations in this gene have photoreceptor outer-segment abnormalities and retinal degeneration (Pazour et al., 2002). As well as providing data to support a biological interaction between DNAAF1 and IFT88, we observe co-localization of DNAAF1 and IFT88 at the basal body of primary cilia. The role of DNAAF1 in primary cilia is not known, but it has been identified to localize to the basal body in previous studies (Basten et al., 2013) and the immortalized primary ciliated cell lines used in this study are a useful model system which enables assessment of putative protein interactions. Our results suggest that

DNAAF1, as well as being involved in the assembly of dynein arm complexes, could mediate the delivery of these components directly to the intraflagellar transport machinery. This hypothesis would be supported by the experiments performed by Duquesnoy et al. in *Trypanosoma brucei*, which showed that on knockdown of the *DNAAF1* orthologue, dynein arm components such as DNAI1 were synthesised, but not transported from the cytoplasm to the axoneme (Duquesnoy et al., 2009).

Our TAP and coIP experiments also confirmed interactions of DNAAF1 with both RUVBL1 (Pontin) and RUVBL2 (Reptin). Immunofluorescence microscopy studies demonstrated co-localisation of DNAAF1 with both RUVBL1 and RUVBL2, although their location was mainly nuclear.

RUVBL1 and RUVBL2 are DNA-dependent AAA+ ATPases involved in many cellular processes, such as energy metabolism, chromatin remodelling, transcription factor function, the DNA damage response and snoRNP assembly and maturation (Rosenbaum et al., 2013). Pontin has recently been identified as playing an important role in cell cycle progression (Breig et al., 2014). Pontin and Reptin form hetero-hexamers and dodecameric-like structures. These structures form the core of a number of important assembly and chaperone complexes, for example the RT2P complex, which is involved in the assembly of certain complex molecules, such as snoRNPs, RNA polymerase II and DNA damage response kinases (ATM, ATR and DNA-PK) (Lakshminarasimhan et al., 2016). Alongside the Pontin/Reptin structure, this complex is associated with PIH1D1 (Protein interacting with Hsp90) and RPAP3 (a Tetratricopeptide Repeat (TPR)-containing protein), two interactors of HSP90 (Boulon et al., 2010, Zhao et al., 2008). The exact mechanism by which the RT2P complex functions is not known (von Morgen et al., 2015), but it has been suggested that HSP90 and RPAP3 stabilize PIH1D1, which directly binds these components (Paci et al., 2012). Interestingly, our TAP experiments showed that DNAAF1 interacts with a number of HSP-associated proteins, so named because of their upregulation in response to heat, including both HSP90 and SUGT1 (a TPR-repeat containing protein and a known co-chaperone of HSP90 (Prus and Filipek, 2011)). In support of this finding, SUGT1 and HSP90 have been shown to interact with other LRR domain-containing proteins in plant models (Taipale et al., 2010).

We therefore hypothesise that dynein arm assembly involves a complex similar to the RT2P complex, involving a RUVBL1/RUVBL2 core, a TPR-containing protein (for example SUGT1) and a PIH1-domain containing protein. Appropriate PIH1-domain

containing proteins would include DNAAF2/KTU, mutated as a known cause of human PCD and recently shown to be involved in the preassembly of the dynein arms (Omran et al., 2008), PIH1D3, a recently described X-linked human PCD gene (Paff et al., 2017) and MOT48, which has links to dynein arm preassembly and is reported to be required for the stability of heavy chains (Yamamoto et al., 2010). Yamamoto *et al.*, 2010, raised the suggestion that PIH-domain containing proteins are involved in dynein preassembly due to their increased presence in organisms with motile cilia and flagella (Yamamoto et al., 2010). Furthermore, PIH1D3 has been shown to interact with HSP90, DNAAF2/KTU, DNAAF4 and an intermediate chain of the ODA, although not DNAAF1 (Dong et al., 2014, Paff et al., 2017). Reptin has recently been shown to interact with LRRC6/Seahorse (Zhao et al., 2013), another dynein arm assembly factor and recognised PCD gene (Inaba et al., 2016), which would be consistent with our model.

This work suggests a novel role of the RUVBL1/2 (Pontin/Reptin) complex in motile ciliary assembly, which confirms previous data that suggests a link between these proteins and motile ciliary function. In *Chlamydomonas*, Pontin and Reptin are upregulated following de-flagellation (Stolc et al., 2005). Zebrafish embryos with loss-of-function *ruvbl1* (Sun et al., 2004) or *ruvbl2* (Zhao et al., 2013) mutations have ventral body curvature and kidney cysts, demonstrating a typical “ciliopathy” phenotype. These zebrafish *ruvbl2* mutants have an absence of dynein arms in motile cilia, causing static cilia, further supporting the suggestion that RUVBL2 is involved in mediating dynein arm assembly or maintenance (Zhao et al., 2013). The same research group have demonstrated that Reptin genetically interacts with *ift172*, a subunit of the IFT-B complex necessary for ciliary assembly (Zhao et al., 2013).

To link our findings, we hypothesise that DNAAF1 is further required to couple the dynein arm components, pre-assembled by the RT2P-like complex, to the IFT machinery (see Figure 5-22). This may help to explain the finding of low levels of DNAAF1 in ciliary axonemes (Duquesnoy et al., 2009). Furthermore, Pontin knockdown appeared to perturb the co-localisation of DNAAF1 to IFT88, as shown in our siRNA experiments. This may suggest that a failure of the formation of the Pontin/Reptin dynein arm assembly complex leads to a disruption of DNAAF1 coupling with the IFT machinery.

The missense mutation identified in family CHD1 does not appear to affect the interaction of DNAAF1 with Pontin/Reptin, demonstrated here by colP, or the

interaction of DAAAF1 with members of the RT2P-like complex, as shown by the TAP data, where a number of these proteins appeared to be more strongly associated with the mutant form of DAAAF1. The only exception was Pontin, which was slightly more highly associated with the wildtype form of DAAAF1, although the results were not statistically significant. The mutation did appear to affect the interaction of DAAAF1 with IFT88, demonstrated by coIP, suggesting that the proposed subtle defect in motile ciliary function in family CHD1 may be related to delivery of the dynein arms to the transport machinery.

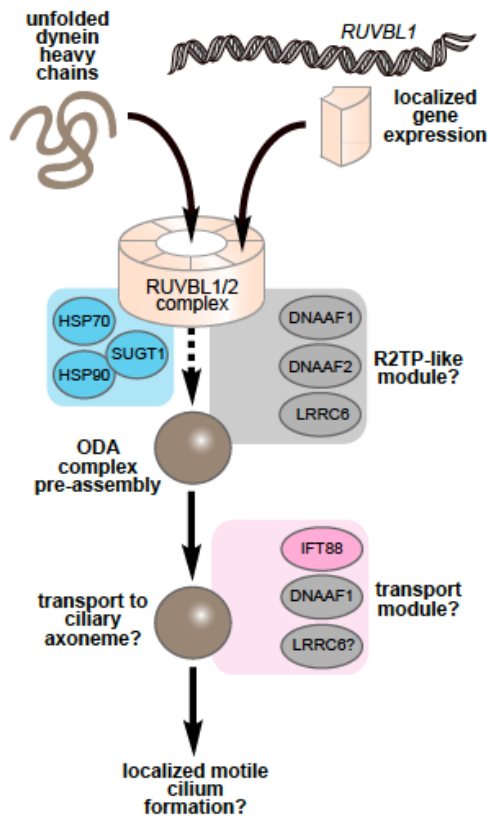


Figure 5-22: Schematic diagram to show the hypothesized model of an RT2P-like complex involved in dynein arm assembly.

The proposed complex involves a core RUVBL1/2 structure, HSP90, DAAAF2 and SUGT1. This complex is hypothesized to interact with dynein assembly components DAAAF1 and LRRC6. Transport to the ciliary axoneme is then coupled to the IFT machinery, for example IFT88.

5.3.7 Asymmetry of Pontin at the Embryonic Node

In murine whole mount immunofluorescence studies we observed the expression of DAAAF1 and RUVBL1 at the embryonic mouse heart and embryonic node in order to gain insights into the role of these proteins in early development. RUVBL1 (Pontin) was expressed across the primitive atria and ventricles during embryonic heart development, at gestational day E8.5. This finding is consistent with previous studies, which showed Pontin and Reptin to be associated with cardiac development, with heart

size in zebrafish antagonistically regulated by the action of Pontin and Reptin (Rottbauer et al., 2002). Furthermore, a number of Pontin/Reptin-containing complexes are known to be expressed in the developing heart (Ikura et al., 2000, Lough, 2002).

During the course of our experiments we unexpectedly observed asymmetric distribution of Pontin (RUVBL1/*ruvbl1*) at the embryonic node in mice and Kupffer's Vesicle (KV) in zebrafish. This expression was consistently left-sided in both animal models. This asymmetric distribution was dependent on the presence of DAAF1, as it was abolished in the *dnaaf1*^{-/-} zebrafish model. One explanation for this finding is that RUVBL1 is acting as a transcriptional regulator of asymmetrically expressed genes in early development. Gene transcription and chromatin modeling are previously known functions of both Pontin and Reptin (Jonsson et al., 2001, Gallant, 2007). They can act as ATP-dependent mediators of chromatin remodeling. They have been identified to play a role in a number of chromatin remodeling complexes, such as the Ino80 complex in yeast, involved in the transcription of around 5% of genes (Jha and Dutta, 2009, Jonsson et al., 2001). The asymmetric fluid flow from the ciliary beat at the embryonic node induces asymmetric gene expression on the left and right hand sides of the node. *Cerl2* (Cerberus-like protein 2) is one of these asymmetrically expressed genes and is expressed in a right>left pattern. This asymmetry of *Cerl2* mRNA is found to be controlled by Wnt signaling, with Wnt3 expression seen in a left>right pattern (Nakamura et al., 2012). In mouse models, knockout of the beta-catenin gene, a key downstream effector of the canonical Wnt pathway, results in randomization of *Cerl2* expression at the node. Furthermore, excessive levels of Wnt3A protein result in downregulation of *Cerl2* mRNA. This model of asymmetry necessitates an increased expression of Wnt signaling on the left hand side of the node (Nakamura et al., 2012). Pontin and Reptin are known antagonistic regulators of Wnt signaling by binding to beta-catenin (Rottbauer et al., 2002), with Reptin inhibiting the Wnt pathway and Pontin activating it (Bauer et al., 2000). It may be hypothesized that an asymmetrical distribution of Pontin at the node is related to the regulation of Wnt signaling in left-right patterning. In this model, the absence of *dnaaf1* in the knockout zebrafish leads to static cilia at the KV, abolishing asymmetrical gene expression, including asymmetrical Pontin expression (see Figure 5-23).

A further possibility is that dynein arm assembly in nodal cilia is initially specified solely on the left side of the node in early embryonic development, so an asymmetric distribution of Pontin is present to initiate early nodal cilia formation. This earlier development of nodal cilia on the left could contribute to the establishment of leftward

nodal fluid flow. Earlier dynein arm assembly on the left of the node could also allow immotile cilia on the perinodal crown cells at the edge of the node to respond to nodal flow before those on the right side of the node. This hypothesis might be supported by studies that suggest that left-right patterning is established very early in development of the node by a weak flow of fluid from a subset of nodal cilia. This weak flow only requires a very small number of cilia to initiate symmetry breaking (Shinohara et al., 2012). In the *dnaaf1*^{-/-} embryo, the lack of active *dnaaf1* could lead to a disruption of the dynein arm assembly complex, leading to a loss of asymmetrical Pontin. This intriguing observation of asymmetric distribution of RUVBL1/Pontin at the node requires further investigation to understand the unexpected role of Pontin (and likely also Reptin) in left-right axis determination.

5.4 Conclusions

We have shown that mutations in DNAAF1 can cause a phenotype of PCD, PCD with CHD or pure CHD, suggesting a phenotypic continuum between CHD and PCD. This has clinical implications, as CHD patients may have ciliary dysfunction, associated with poorer surgical outcome (Nakhleh et al., 2012), a situation that could be investigated by pre-screening prior to surgical intervention. We also suggest that CHD patients (even with apparently normal ciliary function) may harbor mutations in PCD-associated genes. This observation may allow for improved genetic diagnosis in CHD patients, which may allow for advances in medical and surgical management.

We have demonstrated novel interacting partners of DNAAF1, including IFT88, Pontin, Reptin and a number of HSP proteins, giving key insights into the function of DNAAF1 in dynein arm assembly. These investigations have allowed us to hypothesize regarding the function of DNAAF1 and to place it within a novel R2TP-like complex in dynein arm assembly, as well as suggesting a role for DNAAF1 in the transport of dynein arm components to the intraflagellar transport machinery. We have provided further evidence that Pontin and Reptin are important ciliary proteins, and have shown Pontin to be asymmetrically expressed at the developing embryonic node in mice and zebrafish. This suggests a role in the creation of laterality, possibly distinct from its suggested role in dynein arm assembly. This finding places DNAAF1 and Pontin as key players in the development of embryonic laterality, as well as generating many questions regarding the exact nature of this role. These questions deserve further investigation to ascertain the true nature of this exciting observation.

6 A Variant in *KMT2D* is Associated with a Syndrome of Athelia, Choanal atresia and Congenital Heart Disease

6.1 Introduction

In this chapter we present the results of investigations into a family with a specific phenotype of athelia (absence of the nipples), choanal atresia and congenital heart disease, associated with other developmental abnormalities. There was a single affected child in this family. Through exome sequencing we identified a *de novo* variant in *KMT2D*, a gene associated with Kabuki syndrome. Kabuki syndrome is characterized by developmental delay, distinctive facial features, skeletal abnormalities, short stature and haematological disorders (Dentici et al., 2015), a very different phenotype to that recognized here. Through collaboration we were able to identify a second family who had the same variant in *KMT2D* and were affected with a very similar phenotype. This variant was present in the father of the affected children and both affected individuals. We offer evidence to suggest that the father in this family is mosaic for the variant and is unaffected by features of the condition. We further investigate two other families affected by athelia and summarize the previous published cases of athelia and choanal atresia from the literature.

6.2 Results

6.2.1 Family CHD9

6.2.1.1 Clinical Phenotype

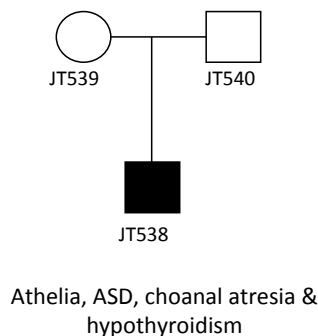


Figure 6-1: Pedigree of family CHD9.

This is non-consanguineous family with a single child affected with multiple congenital anomalies. On antenatal 12 week scan, an increased nuchal thickness of 3.5mm was noted, and Chorionic Villous Sampling (CVS) was normal. At the 20 week antenatal scan an echogenic bowel was noted and the parents were investigated for mutations in *CFTR*, which were normal. At 37 weeks gestation, cerebral redistribution was recognized and so induction of labour was performed. Following a normal vaginal delivery there was fetal distress and intubation was performed at 8 minutes of age. This patient remained on the neonatal unit for 3 months receiving treatment for interstitial lung disease and multiple abnormalities were noted. These included bilateral choanal atresia, absent thymus, branchial cysts, absent gall bladder, hypothyroidism, a caudothalamic cyst, preauricular pits, tongue tie, moderate to severe hearing loss and neonatal teeth. Cardiac anomalies included ASD and pulmonary hypertension. This patient required nasal oxygen until the age of 18 months due to interstitial lung disease and at the age of 3 ½ years still requires oxygen supplementation for flights. He has suffered recurrent breath-holding attacks. He had Intrauterine Growth Retardation (IUGR) and was 2.35kg (<0.2nd centile) at birth, and has continued to fail to thrive with height and weight remaining below the 0.2nd centile until the age of 3 ½ years. His head circumference is on the 50th centile. He is being considered for gastrostomy placement due to poor oral intake. He has bile duct dilatation. He is growth hormone deficient and receives regular growth hormone injections. His gross motor development was delayed, he first sat at 12 months, crawled at 15 months and walked at 3 years. He also has delayed expressive language, which is thought to be related to his hearing loss. He has normal cognitive function at 3 ½ years. He has been found to have a low CD8 T-lymphocyte count. This is being monitored the haematology team and he receives azathioprine coverage in the winter months. He had delayed tooth eruption. There are no ophthalmological concerns. Both parents are unaffected (Figure 6-1).

6.2.1.2 Whole Exome Sequencing

WES was performed using the Illumina XT method and sequencing was performed on the HiSeq2500 instrument with 6 samples pooled per lane. WES analysis was carried out for the affected individual and both parents. Data analysis was performed as previously described, in this case assuming a new dominant (*de novo*) mutation as the most likely cause, but also investigating for biallelic variants. Table 6-1 shows the number of variants remaining following each step in the filtering pathway for a *de novo* variant and table 6-3 shows the variants that remained after each filtering step. Table 6-2 shows the number of variants remaining following each step filtering on a recessive

disease model and Table 6-3 shows the variants remaining following filtering assuming a *de novo* (dominant) disease model.

Table 6-1: Number of variants remaining following each stage of filtering in family CHD9, assuming a new dominant disease model.

Stage in Analysis	Number of variants
Called with GATK	487857
Filter on dbSNP/EVS (<1% and b129)	67307
Filter on exac <1%	63894
<i>De novo</i>	187
Pass filters	48
Cadd >15	6
Not in HLA gene	3

Table 6-2: Number of variants remaining following each stage of filtering in family CHD9, assuming a recessive disease model

Stage in Analysis	Number of variants
Called with GATK	487857
Filter on dbSNP/EVS	69792
Functional variants	2090
Biallelic	49
Pass hard filters	42
CADD >15	13
No mucin/ HLA genes	3
<1% in exac	1

Table 6-3: Variants remaining following filtering in Family CHD9

Family, inheritance	Gene	Variant	Protein change	Polyphen 2	SIFT	Condel	CADD score	MAF (GnomAD)	Protein function	Omim disease
CHD9, <i>de novo</i>	<i>RBM25</i>	14:7354417 C>T	P85L; NP_067062	Probably damaging (1)	Deleterious (0)	Deleterious (0.945)	32	-	Splicing factor	
	<i>KMT2D</i>	12:49427932 C>A	G3553V; NP_003473	Probably damaging (0.997)	-	-	24.5	-		Kabuki syndrome
	<i>PTPRM</i>	18:8384644 A>G	D1322G; NP_002836	Benign (0.008)	Deleterious (0)	Neutral (0.445)	23.9	-	Protein tyrosine phosphatase	
CHD9, recessive	<i>SOGA1</i>	20:35421971 G>A	T1505M; NP_542194	Tolerated (0.12)	Probably damaging (0.99)	Deleterious (0.702)	17.09	104/185536	Suppressor of glucose, autophagy-associated protein 1	
		20:35443621 G>C	Q742E; NP_542194	Tolerated (1)	Benign (0.199)	Neutral (0.019)	0.008	162/282618		

Data from patient JT538, the affected individual, was filtered to identify variants present in the *CHD7* gene, which causes CHARGE syndrome and has possible phenotypic overlap with this syndrome, as well as *TTF2*, which causes Bamforth-Lazarus syndrome, a potentially overlapping phenotype. There was poor coverage of exon1 of *CHD7*, however, there were no mutations identified in these genes.

6.2.1.3 CNV analysis

CNV analysis of WES data for this family was done using the Exome Depth program in individual JT538. A full list of identified copy number variants is provided in appendix J. No likely causative CNV was identified that could be functionally linked to the phenotype in this patient.

6.2.1.4 Confirmation of Variant in *KMT2D* by Sanger Sequencing

The variant identified in *KMT2D* was thought to be a good candidate for the phenotype. It is a known cause of Kabuki syndrome, which has phenotypic overlap with the identified features. It was confirmed to be present in the heterozygous state in JT538 and was not present in his parents (Fig. 6-2).

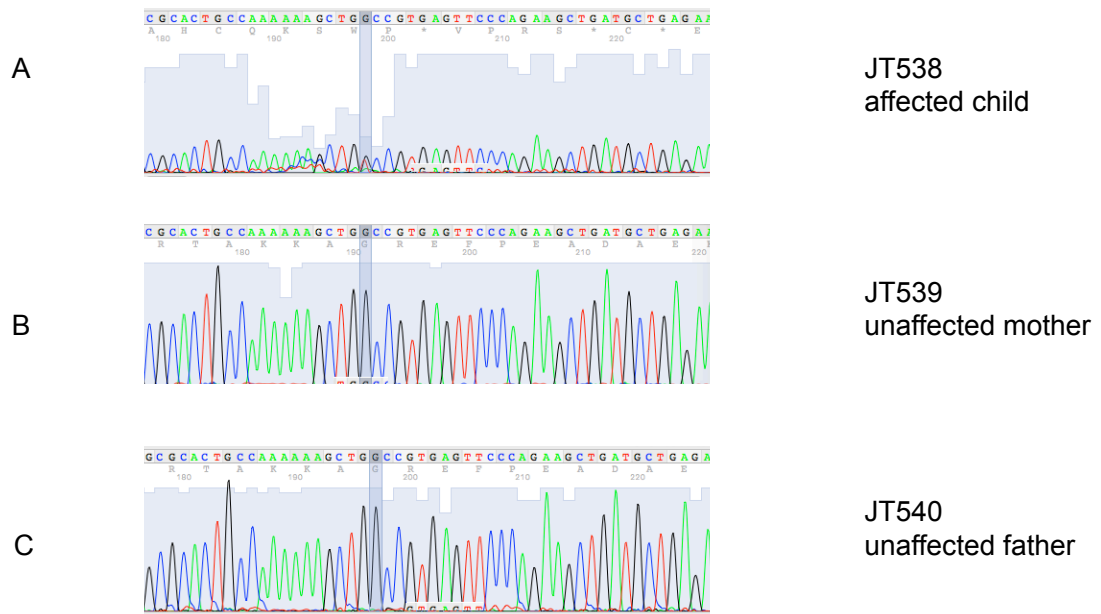


Figure 6-2: Sanger sequencing electropherograms to demonstrate a mutation in *KMT2D*, c.10658 G>T, in family CHD9.

The mutation is shown to be heterozygous in the affected child (A), and not present in the unaffected mother (B) or father (C). The blue shading indicates the position of the mutation.

6.2.2 Identification of a Second Family affected with Athelia and Choanal Atresia: Family CHD9b

Following the finding of a *de novo* variant in *KMT2D* in family CHD9, we conducted a search of a number of databases for this variant. It was found to be present in two individuals in a dataset containing a large number of control samples, but with some patient samples also present. On further investigation it was found that the family who carried this same variant in *KMT2D* (family CHD9b) had a very similar phenotype to that seen in family CHD9. This family had previously been described in the literature in 2002 (Al-Gazali et al., 2002) and WES had been performed at Queen Mary's Hospital in London.

6.2.2.1 Clinical Phenotype

This family is a first cousin union of Palestinian origin. Their first child had choanal atresia and died at birth. They had two further affected children. The male child had right choanal atresia, hypoplastic nipples, a PFO, a small thyroid gland and bilateral absent lacrimal duct with bilateral duct stenosis. He suffered with bilateral conductive hearing loss. His left testis was undescended. He had failure to thrive and chronic respiratory infections in the first year of life, which lead to chronic interstitial lung fibrosis. He had sparse hair and eyebrows, dystrophic nails, down-slanting palpebral fissures, low set ears and hypoplastic alai nasi. Cranial CT scan and renal ultrasound scan were normal. Thyroid function tests were normal. At the time of review he was 10 years of age.

The affected female child had widely spaced eyes, a depressed nasal bridge, a low right ear and a neck pit. She had hypoplastic nipples and bilateral choanal atresia. She had sparse hair and eyebrows. She had hypothyroidism and complete absence of the thyroid gland. An abdominal ultrasound scan and brain CT scan were normal. She died at the age of 4 months following a chest infection. Both children had growth parameters <5th centile. Both parents were unaffected and there were four unaffected children in the family (Al-Gazali et al., 2002).

6.2.2.2 Data analysis

.vcf files from Queen Mary's Hospital, London were sent to our research team for analysis from the affected female child and both parents. The variant in *KMT2D*, c.10658C>A, was present in the affected female and her unaffected father. The number of reads for each call at this position were 30 wildtype to 33 mutant in the affected girl, and 33 wildtype to 17 mutant in her father, allowing for the possibility of

mosaicism in the father. The mother had 40 reads aligning to the wildtype sequence. The data was also filtered for other possible pathogenic mutations, or further overlap with family CHD9, but no likely candidates were identified.

6.2.2.3 Sanger Sequencing

Confirmation of the variants in *KMT2D* was performed by Dr Laura Southgate, Research Fellow of Queen Mary's, University of London, by conventional Sanger sequencing methods. The variant was confirmed to be heterozygous in the father and heterozygous in the two affected siblings in this family. Mosaicism was supported by the finding of a reduced mutant peak on Sanger sequencing of the father, compared to the two affected children (Figure 6-3).

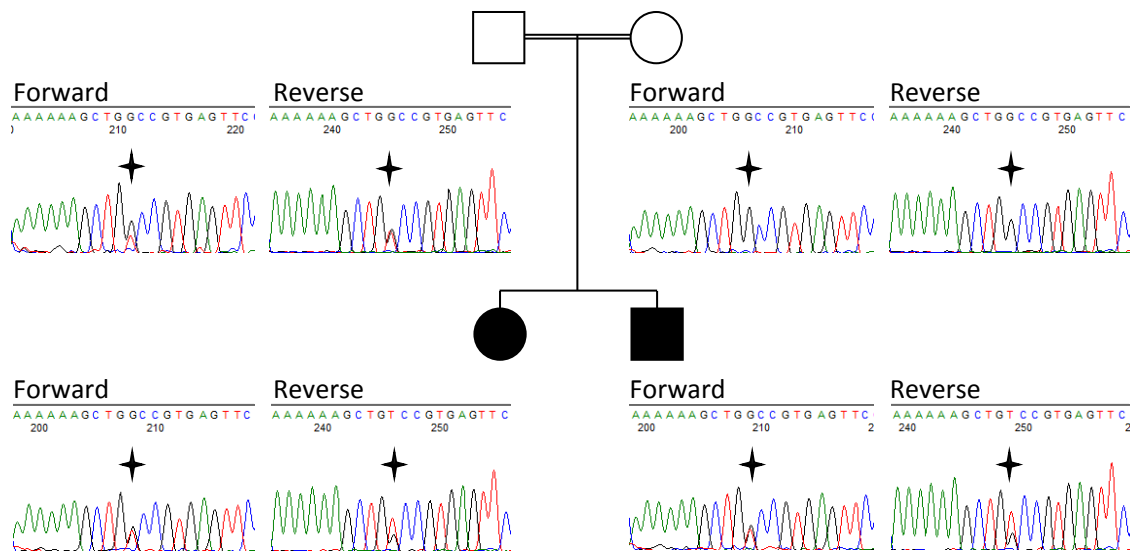


Figure 6-3: Sanger sequencing electropherograms showing the presence of *KMT2D* c.10658G>T (p.G3553V) in family CHD9b.

The position of the variant is indicated by a star. It is shown to be heterozygous in the father and both affected siblings and wildtype sequence is seen in the mother. The sequence peak height is reduced for the mutant allele in the father, suggesting mosaicism. This analysis was performed by Dr L. Southgate, University of London.

6.2.3 Family CHD10

Further to the findings in family CHD9 and CHD9b, we were sent data from a family affected with isolated athelia, to identify whether *KMT2D* was a consistent cause associated with athelia.

6.2.3.1 Clinical Phenotype

This is a family consisting of a mother and daughter affected with athelia, but with no other congenital anomalies. They were referred from Dr Meenakshi Bhat, Department of Fetal Medicine, Bangalore.

6.2.3.2 Whole Exome Sequencing Data Analysis

WES data was received from the collaborating centre for the affected mother and child in the form of .fastq files. Data filtering was performed by following the standardized pipeline. The inheritance pattern in this family was assumed to be autosomal dominant inheritance, and so variants were filtered to identify rare heterozygous variants present in the mother and child. There were no unaffected samples to filter against, making the removal of variants extremely difficult. Variants were removed that were present in the EXaC, EVS and dbSNP control datasets. Even following this strict filtering criteria, 25 variants with a CADD score >15, predicted to be damaging by *in silico* tools and passing hard filters remained (see Table 6-4). The remaining variants are listed in appendix N. The most likely candidate was a variant in *PTCH1* (see Table 6-5).

Table 6-4: Number of variants remaining following filtering assuming a dominant inheritance pattern in family CHD10.

Stage in Analysis	Number of variants
Called with GATK	264327
Filter on	
Exac	166436
dbSNP (b129 and prev)	56271
EVS	56218
Functional variants	461
Hard filter	48
'Damaging'	35
CADD >15	23

Table 6-5: Showing the variant in *PTCH1* identified in family CHD10.

D= Deleterious, PD=Probably damaging.

Gene	Vari-ant	Protein change	Poly-phen2	SIFT	Condel	CADD	MAF	Known protein function
<i>PTCH1</i>	1324 C>T	V442M; NP_000255	PD (0.921)	D (0)	D (0.816)	26.2	N/a	Shh signalling

The data was checked for candidate variants in the following genes; *KMT2D*, *PTPRM*, *TXNL4A*, *PTPN13*, *PTPRO*, *FOXE1*, *CHD7*, *TBX5*, *TBX3*, *LEF1*, *NRG3*, *ERBB4* and the known interactants of *PTPRM* (*NHS*, *CDKL5*, *PVRL3*, *PIP5K1C*, *CTNNB1*, *GNB2L1*). No variant was identified that was present with a MAF <1% in control datasets.

6.2.4 Family CHD11

The team who cared for family CHD11 were contacted due to the overlap in phenotype between these patients and those of families CHD9 and CHD9b. The only available

material was in the form of FFPE embedded slides, which were sent to our laboratory with informed consent.

6.2.4.1 Clinical Phenotype

Paraffin embedded slides were sent from collaborators in Yale-New Haven Hospital (CT, USA) from a patient seen there in 1998. This was a member of a family described in the literature with a lethal syndrome involving renal tubular dysgenesis and absent nipples (Hisama et al., 1998).

This is a non-consanguineous family: the mother was of German/Danish descent and the father was of Norwegian/Irish descent. In the first affected child, IUGR was recognized at 30 weeks gestation. The baby was born at 38 weeks gestation with dysmorphic features including hypertelorism, down-slanting palpebral fissures, micrognathia, a short beaked nose with hypoplastic alae nasi and a coloboma of the left nostril. His ears were low set and there was athelia, broad thumbs and an imperforate anus. The autopsy showed an abnormal right-sided lung lobe, ancillary spleen and hypoplastic kidneys with renal tubular dysgenesis on histology. The karyotype was normal except for the presence of a Y-heterochromatin marker chromosome. FISH for 22q11 deletion was normal. He died at 12 days of life.

The second child was delivered at 37 weeks. A Chorionic Villus Sampling (CVS) had been performed and was normal. There were similar dysmorphic features, with a unilateral neck pit, absent nipples and undescended testes. He had cardiac defects including a large VSD, bicuspid aortic valve, distal aortic arch and large ductus arteriosus. His kidneys were hypoplastic. This patient died at 25 hours of life.

Their 3rd child was affected with a similar syndrome. He was born at 36 weeks gestation. He had similar dysmorphic features, with a simple left ear and a pre-auricular skin tag and pit. He had bilateral neck pits, choanal atresia, athelia and 5th finger clinodactyly. He had hypoplastic kidneys and a normal echocardiogram. He died at 35 hours of life. Parathyroid, thymus and gallbladder were absent. There was an accessory spleen, focal renal dysplasia in the left kidney and neuroblastoma in the left adrenal gland. There was a family history of neck cysts.

6.2.4.2 DNA Extraction

DNA was extracted from five kidney sections of paraffin embedded slides from patient JT619, the 3rd child in this family. The resulting DNA concentration was 7.68ng/μl, measured using the Qubit system.

6.2.4.3 Whole Exome Sequencing

Due to the low concentration of DNA extracted from paraffin-embedded slides, sample JT619 was vacuum concentrated to give a total of 70ng in 2 μ l. WES library preparation was initially attempted using the Illumina QXT method, but initial Bioanalyser results following shearing gave a flat trace. When the sample was analysed using a high-sensitivity Bioanalyser chip, the DNA was over-fragmented with the majority of DNA strands at 100-150bp and the DNA was at very low concentration (Fig. 6-4).

Library preparation was re-attempted using the Illumina XT protocol, without shearing the DNA, as the DNA was already fragmented and degraded. End repair, addition of 'A' bases, ligation of adapters and amplification using two 6-cycle PCRs were performed. Figure 6-4 shows the high sensitivity Bioanalyser trace from this sample. Fragments were seen within the correct range at 200-400bp. The sample was vacuum concentrated to give 26ng/ μ l and hybridization was performed as recommended by the manufacturer's protocol. Unfortunately the hybridization was unsuccessful and the second Bioanalyser trace, following hybridization, showed no usable remaining DNA (Fig. 6-4). Attempts at WES library preparation were therefore discontinued.

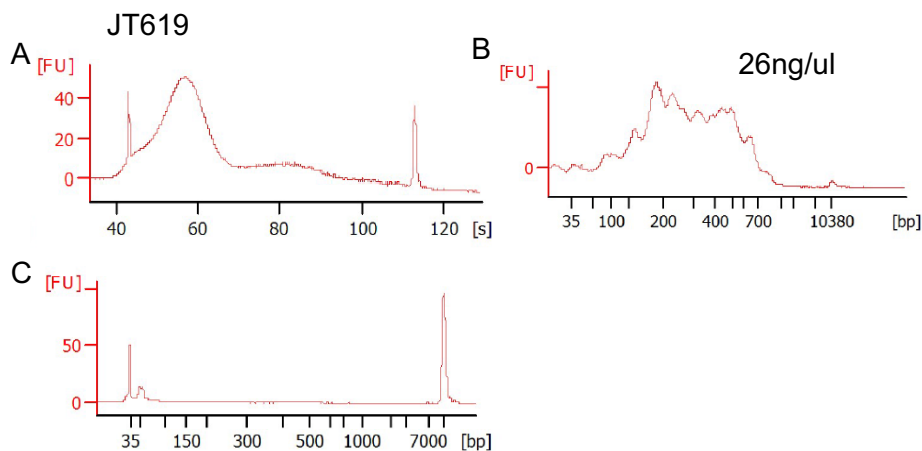


Figure 6-4: High-sensitivity bioanalyser traces from sample JT619.

A: following shearing using the QXT library preparation kit, B: following amplification using the XT library preparation kit, C: following hybridization using the XT library preparation kit.

6.2.4.4 Sanger sequencing

DNA was re-extracted from the remaining five FFPE slides from patient JT619 for Sanger sequencing. DNA concentration was measured on Nanodrop at 23.0ng/ μ l. Whole Genome Amplification (WGA) was performed on this sample. Initial PCR reactions failed to amplify the required product to sequence exon 38 of *KMT2D*, and so primers were designed to amplify a shorter PCR product (180bp) due to the low quality

of DNA. This was successful and Sanger sequencing was performed. The variant in *KMT2D*, c.10658 C>A, was not present in sample JT619, as shown in Fig. 6-5.

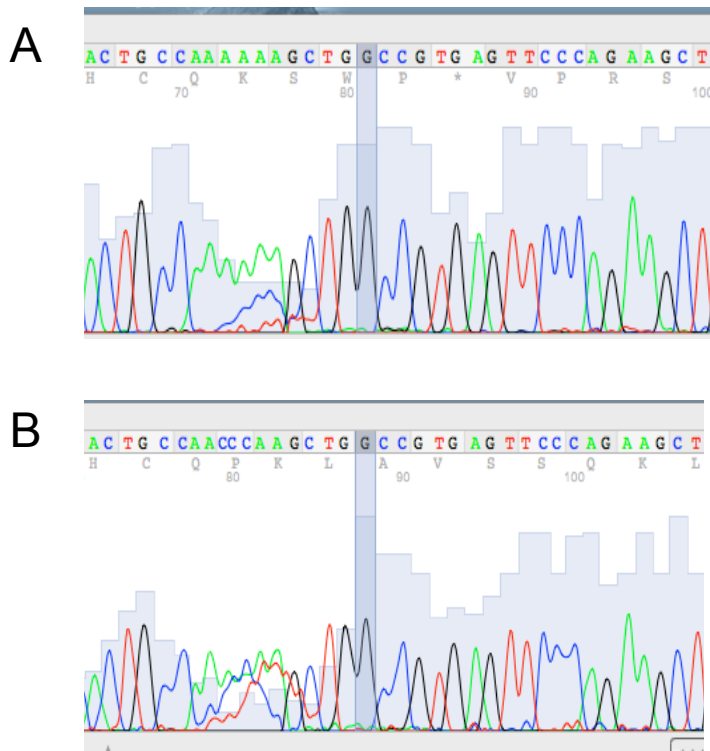


Figure 6-5: Demonstration of absence of c.10658G>T in patient JT619 from family CHD11.

A and B show two biological replicates of Sanger sequencing using the same primer pair, but different aliquots of input DNA.

6.3 Discussion

6.3.1.1 Clinical Reports of Syndromes Consisting of Athelia and Choanal Atresia

We have investigated four families affected with athelia (absent nipples), three of which have a rare developmental anomaly, associated with choanal atresia and other congenital anomalies. Table 6-6 shows a comparison of the phenotypic overlap of the investigated families, alongside other reported patients from the literature.

Table 6-6: Summary of clinical features of patients with athelia syndromes from this study and the literature.

DD= developmental delay, PVL= periventricular leukomalacia, PH= pulmonary hypertension.

Family number/ Author & year	Assumed inheritance	Athelia	Choanal atresia	FTT	Thyroid disease	Cardiac defects	Renal anomalies	Other organ absence	Dysmorph-ism	Other	KMT2D variant
CHD9	<i>De novo</i>	yes	yes	yes	Hypo-thyroidism	ASD, PH	-	Thymus, gall bladder	Neonatal teeth	Interstitial lung disease, branchial cysts, caudothalamic cyst, pre-auricular pits, hearing loss.	<i>De novo</i> c. 10658C>A
CHD9b/ (Al-Gazali et al., 2002)	Two affected sibs, consanguineous.	Hypoplastic nipples	yes	yes	Small thyroid gland/ absence of thyroid gland	PFO	-		Sparse hair, ds palpebral fissures, low-set ears, hypoplastic alai nasi, neck pits	Absent lacrimal duct, hearing loss, undescended testis, chronic infections	c. 10658C>A heterozygous in affected child and unaffected father
CHD10	Mother and daughter affected	yes	-	-	-	-	-	-	-	-	c. 10658C>A not present

Family number/ Author & year	Assumed inheritance	Athelia	Choanal atresia	FTT	Thyroid disease	Cardiac defects	Renal anomalies	Other organ absence	Dysmorphism	Other	KMT2D variant
CHD11/ (Hisama et al., 1998)	Three affected sibs	yes	yes	yes	-	VSD, BAV, absent distal aortic arch, PDA.	Hypoplastic kidneys, focal renal dysplasia.	Thymus, parathyroid, gall-bladder	Prominent forehead, flat midface, ds palpebral fissures, broad digits, low set ears, small alai nasi, coloboma left naris.	Preauricular pits, branchial remnants, accessory spleen, imperforate anus, undescended testes, neck pit. Neuroblastoma.	c. 10658C>A not present
(Horvath and Armstrong, 2007)	Single affected	yes	yes	yes	Hypothyroidism	Right aortic arch with vascular ring.	?echogenic cortex, repeat scan normal.	-	Widened, infra-orbital creases, short upturned nose,	PH, PVL, preauricular pit	unknown
(Dumic et al., 2002)	Single affected	yes	Hypothelia	yes	-	-	-	-	-	Hearing loss, neck fistulas,	unknown
(Qazi et al., 1982)	Con-sanguineous, 2 affected children, aunt also affected	yes	Hypothelia	yes	-	-	-	-	Narrow high-arched palate	DD, pectus excavatum	unknown

Family number/ Author & year	Assumed inheritance	Athelia	Choanal atresia	FTT	Thyroid disease	Cardiac defects	Renal anomalies	Other organ absence	Dysmorphism	Other	KMT2D variant
(Uchida et al., 2006)	Single affected	yes	yes	yes	-	-	-	-	-	Cleft palate, congenital jejunal atresia, mild DD, mild atrophy left cerebrum.	unknown
(Glustein et al., 1996)	Single affected	Yes	Hypoplasia left nipple			PDA, pulmonary stenosis, pulmonary hypertension.	Bilateral hydro-nephrosis			Abnormal left ear lobe, cleft palate, choroid coloboma, cerebral atrophy	unknown
Methimazole embryopathy (Greenberg, 1987)	-	yes	yes	yes	-	yes	-	-	Upward slanting palpebral fissures, broad nasal bridge.	Deafness, imperforate anus, hypospadias scalp defects, DD, retinal coloboma	

As can be seen from Table 6-6, there is significant phenotypic overlap between family CHD9 and family CHD9b, including athelia, choanal atresia, CHD, thyroid gland abnormalities, dysmorphic features, hearing loss and failure to thrive. There are many similarities with the phenotype of family CHD11, but who have the additional features of hypoplastic kidneys and focal renal dysplasia. The affected individuals from family CHD11 died in the neonatal period, whereas this was not a feature in families CHD9 and CHD9b.

There have been a number of reports in the literature of patients with a similar phenotype to these families. The individual described by Horvath *et al.* was thought to have the same syndrome as that seen in family CHD11. This child had periventricular leukomalacia, and a more wizened appearance than had been described in previous reports, but otherwise had consistent features (Horvath and Armstrong, 2007). An attempt was made by our team to source genetic material from this family, but the parents did not give consent for further investigation. The reports of Qazi *et al.*, 1982, and Uchida *et al.*, 2006, describe a similar phenotype, but do not report thyroid disease, cardiac defects or renal anomalies and no other organ absence is reported (see Table 6-6) (Qazi *et al.*, 1982, Uchida *et al.*, 2006). These reports may or may not represent the same syndrome.

There is overlap of the phenotype seen in our families with methimazole embryopathy. Methimazole is the active metabolite of carbimazole, a drug that, in the past, was used to treat thyrotoxicosis during pregnancy. However, it has been linked to teratogenicity in a number of different reports. The major described features are choanal atresia, oesophageal atresia, dysmorphic facial features, athelia, CHD, learning difficulties, scalp defects and retinal coloboma (Clementi *et al.*, 1999, Valdez *et al.*, 2007, Greenberg, 1987, Wilson *et al.*, 1998). Greenberg, 1987, published the initial report of a child with choanal atresia, athelia and learning difficulties following exposure to methimazole in pregnancy (Greenberg, 1987) and a number of consistent reports have been published since, although athelia is not always a consistent phenotype (Clementi *et al.*, 1999). There is no history of this medication being used in the families we describe and the pathophysiology of the link between this drug and the described features is not known.

The condition described in our families also overlaps with CHARGE syndrome, a syndrome characterized by Coloboma (80-90%), Heart defects (75-85%), Choanal atresia (50-60%), Mental Retardation, Genital hypoplasia (50-60%) and Ear anomalies

(90%). Other features include cranial nerve dysfunction (70-90%), growth delay (70-80%), cleft palate (15-20%), distinctive facial features (70-80%) and trachea-oesophageal fistulae (15-20%) (Aramaki et al., 2006). Renal anomalies are also occasionally associated with CHARGE syndrome. It is caused by mutations in *CHD7*, a transcriptional regulator (Vissers et al., 2004). Nipple abnormalities are only occasionally described as a feature of CHARGE syndrome; Glustein *et al.*, 1996, describe a syndrome they classified as CHARGE syndrome, but with hypoplastic nipples (see Table 6-6) (Glustein et al., 1996). Molecular testing was not available at the time of this report.

The other major differential diagnosis for our families is Bamforth-Lazarus syndrome. This is a condition with the features of congenital hypothyroidism due to thyroid dysplasia, cleft palate and spiky hair, with or without choanal atresia or bifid epiglottis (Bamforth et al., 1989). This condition therefore shares the overlapping features of thyroid dysplasia and choanal atresia with our families, but is not known to be associated with athelia. It is due to homozygous loss of function mutations in *FOXE1* (*TTF2*). *Foxe1* knockout mice had an ectopic or absent thyroid gland (Castanet and Polak, 2010). *FOXE1* acts as a transcriptional activator of thyroglobulin and thyroperoxidase, is expressed in thyroid follicular cells, and plays a key role in thyroid development (Castanet and Polak, 2010).

We suggest that our families display a distinct phenotype, with a specific set of features that consistently includes athelia, choanal atresia, cardiac defects, dysmorphic features, failure to thrive and thyroid gland abnormalities. It is often associated with gallbladder, thymus or parathyroid absence, and deafness. This phenotype appears to be caused by a specific mutation in exon 38 of *KMT2D*.

6.3.1.2 *KMT2D* is a Known Cause of Kabuki Syndrome

Kabuki syndrome (KS) is caused by mutations in *KMT2D* (also known as *MLL2*) and *KDM6A*, and is a disorder characterized by dysmorphic facial features, growth retardation, learning difficulties, skeletal and craniofacial abnormalities (Ng et al., 2010). Dysmorphic features include long palpebral fissures, short columella with a thin upper lip, large cupped ears, sparse arched eyebrows and the presence of fetal fingertip pads. Associated anomalies include cleft palate, CHD, renal anomalies, feeding problems, gastrointestinal anomalies, seizures, congenital hypothyroidism and susceptibility to infections due to immune dysfunction (Matsumoto and Niikawa, 2003, Armstrong et al., 2005).

KMT2D is reported to cause around 60% of Kabuki syndrome cases and is involved in chromatin regulation. It encodes a methyltransferase that di- and tri-methylates histone H3 lysine-4 (H3K4), associated with transcriptional activation (Lintas and Persico, 2017). This protein contains a C-terminal SET domain, which is associated with its enzymatic activity (Ruthenburg et al., 2007). *KMT2D* appears to affect the expression of retinoic acid-responsive genes, so is likely to be important in anterior-posterior patterning in the brain (Guo et al., 2012). *KMT2D* is known to form a complex with MLL3 and other subunits such as WDR5, ASH2L, RbBP5 and DPY30 (also known as the WRAD complex) (Issaeva et al., 2007, Ernst and Vakoc, 2012). This complex is involved in the regulation of HOX gene expression (Ansari et al., 2011) and is a co-activator of oestrogen receptor alpha (Mo et al., 2006). A recent study showed *KMT2D* loss to be associated with specific DNA methylation signatures and for this gene to regulate methylation of *HOXA4*, *HOXA5* and *SLITRK5* (Butcher et al., 2017). This is perhaps not surprising as HOX genes are important in axial patterning in the developing embryo and *hoxa5* deficient mice have growth deficiencies, skeletal, and limb abnormalities, as well as defects in neural development (Lizen et al., 2017). Murine loss of *Kmt2d* slows growth and development, and causes early embryonic lethality, in part due to mis-regulation of homeobox gene expression (Ng et al., 2010). The MLL2/3 complex is somatically mutated in a number of different cancers, for example medulloblastoma and non-Hodgkin B-cell lymphomas (Morin et al., 2011).

6.3.1.3 Mutations in *KMT2D* May Cause a Phenotype of Choanal Atresia, Athelia and Other Congenital Abnormalities, that is Distinct to Kabuki Syndrome

We have described a potentially pathogenic missense variant in *KMT2D* in two families with a distinctive phenotype of choanal atresia, athelia and other associated congenital anomalies. This variant was predicted to be pathogenic by *in silico* tools and was not present in control datasets. It was confirmed to be *de novo* in the affected child in family CHD9, but present in both affected siblings and the unaffected father in family CHD9b. This suggests that, if pathogenic, this variant is either non-penetrant in the father in family CHD9b, or that he is mosaic for the variant. This possibility is supported by the reduced number of WES reads showing the mutant allele in the father, compared to the affected child, and the reduced mutant peak heights on Sanger sequencing electropherograms in the father. However, further investigation is required to confirm this, such as sampling buccal cells or fibroblasts from the father. Unfortunately, the father in this family has been approached by the family's clinician and has declined to give further samples, so further evidence for mosaicism is not

available. If the father is mosaic, this is an example of gonadal mosaicism acting as a pheno-copy of recessive disease in a consanguineous union, as previous analysis in this family had always assumed a recessive inheritance pattern.

Figure 6-6 shows the known domains of the KMT2D protein. There are 7 plant homeodomains (PHD) for protein binding, one high-mobility group domain (HMG), five LXXLL motifs (for nuclear receptor interaction), an FYRC and FRYN domain (for heterodimerisation) and a SET domain (for histone lysine methylation) (Banka et al., 2012). Mutations known to cause KS are located throughout the protein, including exon 38. The majority of KS mutations are protein truncating, but around 12-22% are missense (Hannibal et al., 2011) and these mutations are presumed to be loss-of-function. G3553V falls within the LXXLL domain (Banka et al., 2012) (see Fig. 6-6), but does not affect the LXXLL motif. We cannot differentiate between this variant and mutations known to cause KS based on our knowledge of the protein structure of KMT2D.

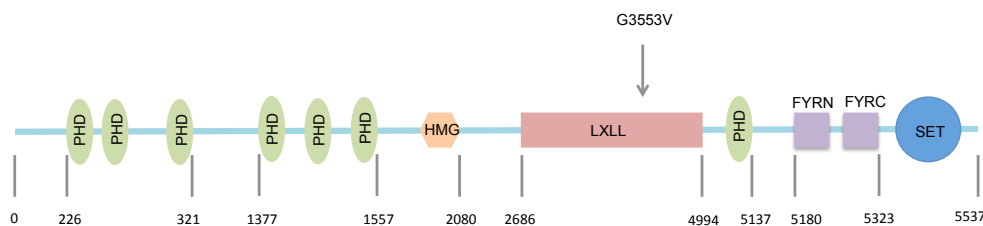


Figure 6-6: KMT2D protein showing domains and position of the G3553V variant.

Figure adapted from (Banka et al., 2012).

The mutation identified in family CHD9 and CHD9b, c.10658 C>A p.G3553V, is not present in control datasets, or in any of the previous publications of KS. It was identified in the COSMIC database (<http://cancer.sanger.ac.uk/cosmic/mutation/>; COSM940022 & COSM940021) as a somatic mutation identified in endometrial cancer and coded as pathogenic. This may give relevance to the finding of neuroblastoma in one of our described patients.

As molecular testing has become more accessible, the spectrum of phenotypes of patients with KS has expanded. The overlap of KS with CHARGE syndrome has recently been delineated (Patel and Alkuraya, 2015). A number of reports have described coloboma in Kabuki syndrome, a feature more typically related to CHARGE syndrome (Ming et al., 2003, Genevieve et al., 2004, Schulz et al., 2014, Patel and Alkuraya, 2015, Verhagen et al., 2014). A recent report described a family with autosomal dominant choanal atresia, hearing loss, dysmorphic features, learning

difficulties and dental anomalies. One family member had an ASD, but none of the other features of KS were present. This family had a heterozygous mutation in *KMT2D*, p.Gln3575His, which segregated with disease in the family. Interestingly this variant, which causes an atypical KS, also falls into exon 38 of the gene (Badalato et al., 2017), as seen in our families. CHARGE syndrome is caused by mutations in *CHD7*, an ATP-dependent helicase that remodels chromatin, which appears to remodel methylated H3K4. There is a known molecular link between CHD7 and KMT2D function, as they are both known to interact with members of the WRAD complex (Zhang et al., 2012). This may explain the overlap in phenotypes in these two conditions (Van Laarhoven et al., 2015).

The phenotype described in our families appears different from that of classical KS. On review of clinical photographs of the family members, a dysmorphologist who specializes in Kabuki syndrome did not classify these patients as having typical KS (Dr S. Banka; unpublished communication). We would therefore describe this phenotype as either a 'novel *KMT2D*-related disorder' or an 'atypical Kabuki syndrome' which expands the clinical spectrum of known KS. Further investigation is required to define a genotype-phenotype correlation within *KMT2D*-related disorders and to delineate classification for these families. We would recommend *KMT2D* sequencing in patients presenting with a similar phenotype.

6.3.1.4 *PTPRM* as a Candidate Gene for Athelia

In the analysis of family CHD9, a variant in *PTPRM* was also identified as *de novo* in the affected child. This was initially investigated as a possible candidate gene as the variant passed all filters, had a high CADD score and was not present in control datasets. *PTPRM* encodes a receptor protein-tyrosine phosphatase, important in cell-cell adhesion (Gebbink et al., 1993) and has been shown to bind to and dephosphorylate p120 catenin (Zondag et al., 2000). *PTPRM* is reported to control organization of the actin cytoskeleton at sites of cell-cell interaction through p120 phosphorylation (Koop et al., 2005).

PTPRF, another receptor protein-tyrosine phosphatase, has previously been associated with athelia. In a recent paper, a homozygous truncating mutation in *PTPRF* was shown to cause athelia/hypothelia in a consanguineous Israeli-Arab family. The affected patients had hypoplasia or absence of the nipple and breast tissue and dysmorphic facial features, but no other abnormalities (Borck et al., 2014). In a further report, a balanced translocation, found to disrupt *PTPRF*, was associated with

bilateral amastia, ectodermal dysplasia, unilateral renal agenesis and dysmorphic features in a Thai woman (Ausavarat et al., 2011). *Ptprf*^{-/-} mice have impaired mammary duct differentiation in the late stages of pregnancy (Schaapveld et al., 1997). Furthermore, *PTPRF* is expressed in human breast tissue (Du et al., 2013).

This link between *PTPRF* and athelia led us to further investigate the potential pathogenicity of a variant in *PTPRM*. However, *Ptprm*-null mice have a diminished flow-induced dilation in mesenteric resistance arteries, but there is no effect on breast development or lactation (Koop et al., 2005). Expression of *PTPRM* was identified in the endothelial cells of the arteries, cardiac muscle and neuronal tissues (Koop et al., 2005). Furthermore, no further mutations in *PTPRM* or other receptor protein-tyrosine phosphatases were identified in the other families investigated, including family CHD10 who had a phenotype of isolated athelia. There was no further evidence of a role of *PTPRM* in breast or nipple development, and this variant was thought less likely to be the cause of the phenotype than *KMT2D*. However, it is possible that the variant in *PTPRM* is playing a modifier role in the phenotype seen in this patient.

6.4 Conclusions

We have demonstrated a specific variant in *KMT2D*, c.10658 C>A, to be associated with a phenotype of athelia, choanal atresia, CHD, thyroid dysgenesis and other congenital anomalies in two families. We suggest that the variant is present in a mosaic state in the father of the affected children in family CHD9b, explaining the apparent recessive inheritance pattern in this family as due to gonadal mosaicism. The reason for the association of this specific phenotype with the described variant is not currently known. It would be of interest to further investigate this finding, perhaps by studying methylation patterns of known *KMT2D* targets, such as *HOXA5* (Butcher et al., 2017), but this is work currently beyond the scope of this project.

7 Discussion

7.1 Summary of Key Findings

This thesis describes a study investigating the genetic causes of congenital heart disease (CHD) using whole exome sequencing. 16 families with CHD were recruited to the study. For the appropriate family members, a full clinical assessment was performed, and blood or saliva samples were taken for DNA extraction. The methods of WES used in this study are shown to provide a good and consistent depth of sequencing coverage across the exomic regions in the assessed samples, and have been successful in identifying known and candidate genes in CHD.

In chapter 3, a family with a mutation identified in a known disease-causing gene, *PIGV*, was described. Children in this family presented with a range of structural abnormalities including congenital diaphragmatic hernia (CDH) and nail dysplasia. These were suggestive of a diagnosis of Fryns syndrome (MIM:229850). Whole exome analysis revealed variants in the *PIGV* gene (MIM:610274). Variants at this locus had previously been described in Hyperphosphatasia and Mental Retardation syndrome (HPMRS MIM:239300). Subsequent review confirmed the key finding of hyperphosphatasia in one affected child, consistent with the molecular findings. This represents an atypical presentation of HPMRS. In particular, the hallmark feature of CDH has not been previously described in HPMRS, nor has the absent bladder seen in one of the affected individuals. These features considerably extend the phenotype of HPMRS.

Ten further families with CHD diagnoses and putative candidate variants identified by WES were investigated and discussed in chapter 3. A number of the candidate genes have functional links to CHD, for example variants in *TMEM260* were identified in a family with Truncus Arteriosus. *TMEM260* has recently been described to cause a syndrome associated with Truncus Arteriosus, renal anomalies and neurodevelopmental delay. Further studies will be required to delineate the true pathogenicity of these and the other candidate variants identified, possibly through collaboration with other groups. Eleven patients with laterality disorders from the DDD project were also analyzed, four with likely disease-causing mutations and a further seven with no clear overlapping candidate gene.

In chapter 4, two children from a consanguineous family affected with sudden death in infancy, associated with cardiomyopathy, were identified to have homozygous mutations in *PPA2*. Through collaboration with other centers, this gene was identified to be the cause of disease in three further families, two with a very similar phenotype to the first, and one with a later onset of disease. Functional characterization of the identified mutations shows them to affect the pyrophosphatase activity of the PPA2 enzyme. Defects in PPA2 enzyme activity are hypothesized to lead to insufficient inorganic phosphate for ATP synthesis, affecting cardiac muscle function, and leading to the phenotype of cardiomyopathy and arrhythmia.

In chapter 5, two children from a consanguineous family affected with heterotaxy were shown to have homozygous mutations in *DNAAF1*. Mutations in this gene are a known cause of PCD, but there was an absence of clinical signs or symptoms of PCD in this family. Support for the pathogenicity of the identified missense mutation was presented using genetic complementation studies in a zebrafish mutant model. Three further families with mutations in *DNAAF1* presenting with clinical PCD and CHD were discussed. Cellular studies showed DNAAF1 to interact with Pontin, Reptin and IFT88, suggesting DNAAF1 acts as part of a novel dynein arm assembly complex in motile ciliated cells. An assessment of Pontin expression at the embryonic node in mice and zebrafish showed an unexpected left-sided expression pattern, which was abolished in the *dnaaf1*^{-/-} mutant fish embryos, placing DNAAF1 and Pontin in the pathways involved in the development of laterality in the early embryo.

In two families with an unusual phenotype of athelia, choanal atresia, thyroid abnormalities and CHD, described in Chapter 6, a putative pathogenic variant in *KMT2D* was identified. This phenotype is distinct from Kabuki syndrome, usually caused by mutations in this gene. This is likely to represent a novel *KMT2D*-related disorder.

WES has been shown to be a successful tool in gene identification in CHD and, coupled with functional studies to enable variant interpretation, this study has provided novel insights into cardiac developmental biology.

7.2 The Difficulties of Candidate Gene Identification in CHD

A clear causative variant was not identified in all of the CHD families studied. In 5 of 16 families (31%) a variant in a known disease-causing gene or a highly plausible candidate variant was identified. In the remaining 11 families, the most likely candidate was less clear-cut. These results are consistent with the published literature describing gene discovery rates in CHD using WES of around 30% (LaHaye et al., 2016).

The difficulties in identifying candidates in this cohort may be due to a number of reasons. It may be that one of the identified candidate variants is the cause of disease in the family studied, but the link to cardiac disease is not yet clear. For example, there may be a paucity of previous literature or known animal models to give an important clue to gene function. In this situation, it is difficult to prioritize the likely disease-causing variant from a short-list. One method to address this is to work collaboratively with other researchers in an attempt to identify multiple families with mutations in the same gene. To address this need, websites such as Matchmaker Exchange (<http://www.matchmakerexchange.org>) are playing an increasingly important role. However, it is important to verify that high quality and consistent phenotypic matching also occurs between families with candidate pathogenic variants in the same gene, in order to avoid false positive results.

In this study, variants have been filtered to identify those that are likely to be 'functional', and both synonymous and non-coding variants are filtered out in the analysis pipeline. It is possible that a true disease-causing variant could be missed due to the filtering stringency. For example, synonymous variants are known to affect splicing or codon usage (Scott et al., 2012). However, reducing the filtering stringency might leave the researcher with an unworkably long list of candidate variants (Cooper and Shendure, 2011). One of the other major difficulties of WES data analysis is in filtering potentially pathogenic mutations from general variation in the population. Publically-available databases, such as EXaC, dbSNP and EVS are useful tools for filtering variants, but none provide a comprehensive survey of normal variation and care needs to be taken in selecting a minor allele frequency on which to filter, to ensure the causative variant is not filtered out due to a comparatively high carrier frequency in the general population (MacArthur et al., 2014).

The mutations causing the phenotype in these unsolved CHD families may be present in deep intronic regions or non-coding DNA, and these regions will not be captured by WES techniques. WGS may be a useful tool to further investigate the cause of disease in these families, but reliable filtering techniques will need to be developed. Large intragenic deletions could also have been missed in WES analysis. In some families, WGS at low read depth was utilized to assess for CNVs. However, due to cost restraints, this was not possible in all of the unsolved cases. Programs such as ExomeDepth, which assess for CNVs from WES data, can be useful in identifying CNVs (Plagnol et al., 2012), but intragenic deletions or duplications may be missed and we have found their use to be complicated by noise and false positive results.

As previously discussed, the genetic causes of CHD are thought to be complex, so in this project we have focused on the assessment of families with an apparent Mendelian pattern of inheritance of CHD. However, we cannot discount the effect of environmental factors, modifier alleles and more complex genetics in these families. It may even be that levels of transcription of certain genes and their chaperones are the source of the observed phenotypes, or epigenetic factors, such as variants in DNA methylation patterns or histone modifications. Families with known disease-causing mutations in *GATA4* and *NKX2-5* are known to show incomplete penetrance and variable expressivity, even in family members who carry a known pathogenic mutation (Schott et al., 1998, Garg et al., 2003). This does suggest that some other effect, such as the epistatic effect of modifier alleles or environmental influences, may be playing a role in these conditions. This hypothesis is supported by the work of Granados-Riveron *et al.*, 2012, who identified multiple variants in *NKX2-5*, *MYH6* and *GATA4* in single patients with CHD, perhaps indicating an additive effect of these variants in the pathogenesis of disease (Granados-Riveron et al., 2012). The cause of CHD in our unsolved families may therefore not be due to a single gene mutation, but perhaps the additive effect of many variants. With such a small number of analyzed families, genetic heterogeneity and epistatic effects would be difficult to interpret.

7.3 WES Filtering Strategies

An average dataset from a WES experiment will provide the researcher with more than 20,000 variants for analysis, dependent upon the reagents and methods used (Stitzel et al., 2011). The methods described in this study for disease filtering have been shown to be robust and to have successfully identified candidate genes in Mendelian disease in numerous previous cases (Parry et al., 2013). However, as with all types of

WES analysis, it is important to understand these methods of filtering fully, as their use may require adjustment according to an individual case.

Variants were filtered according to the observed inheritance pattern in the family. For a recessive disease, variants were retained that were biallelic in the affected individuals, but not biallelic in the parents or unaffected siblings. This methodology requires confidence in the clinical phenotyping, to ensure that 'unaffected' family members do not in fact harbor a sub-clinical disease, for example an undiagnosed VSD. To exclude this possibility, we have aimed to conduct or review echocardiogram investigations in unaffected family members in the families discussed in this project. This is particularly important in apparently unaffected parents who have had more than one child with CHD.

In this project, variant filtering was performed on databases of known variation, for example dbSNP, where a minor allele frequency was selected and variants present above this frequency were removed. This assumes that the mutation causing a rare, recessive disease will not be present at high frequency in a database of known variation. This seemed an appropriate method in this study because, although CHD is relatively common in the population, Mendelian forms and specifically recessive forms of CHD are uncommon. However, difficulties can occur if patient samples are present in these databases, due to a late onset of disease, variable penetrance or variable expressivity as is the case for cardiac arrhythmias (Kaab and Schulze-Bahr, 2005). Databases, such as dbSNP have also been criticized for the quality of data submitted due to genotyping and base-calling errors. One review suggested that 15-17% of SNPs in the dbSNP database are false positives and that around half of the included SNPs are not validated (Mitchell et al., 2004).

The *in silico* tools SIFT, Polyphen2 and CONDEL were further utilized in this project to score variants. Variants with low *in silico* pathogenicity predictions were not automatically filtered out, but always assessed by eye and pathogenicity scores were taken into account alongside functional data. These tools use information from a variety of sources, including sequence conservation, the observed range of variation at a nucleotide position and structural features of the new protein compared to wildtype (Thompson et al., 2013). These tools do have a limited specificity and sensitivity, but it is thought that the use of more than one tool concurrently increases accuracy (Chan et al., 2007). The combined annotation-dependent depletion (CADD) score was used in this project to prioritize variants. CADD is a framework that combines many diverse

functional annotations and combines them into one score. These functional annotations include conservation scores, DNase hypersensitivity sites, transcription factor binding, distance to intron-exon boundaries, expression levels in commonly studied cell-lines, and pathogenicity scores such as SIFT and Polyphen (Kircher et al., 2014). The CADD score is derived from a support vector machine that has been trained to differentiate 'fixed' SNPs, differences between the human genome and an inferred human-chimpanzee ancestral genome, which are tolerated, and simulated model polymorphisms. This machine was then used to calculate a score for every potential single nucleotide variant or short indel across the genome (Kircher et al., 2014). These scores are then "PHRED-scaled" (ranking the variant relative to all possible substitutions in the genome in a logarithmic manner) so only the top 10% of variants will achieve a CADD score >10 and only the top 1% a score >20. The authors of the scoring system suggest using a CADD score of 15 as a cut-off for pathogenicity, but admit that this is arbitrary and should be adjusted according to the project (Kircher et al., 2014). CADD scores were shown to be effective in assessing variants from a number of cohorts, for example in cohorts of children with autistic spectrum disorders (Kircher et al., 2014). However, the clinical utility of CADD score in non-coding regions of the genome has recently been questioned (Mather et al., 2016).

Further selection of candidate variants relies on traditional methods such as review of the literature on encoded protein function and phenotypes of animal models. Ideally, these methods are combined in an automated pipeline, but attention is required by the research team to ensure that these tools are used appropriately, so that candidate variants are not missed.

7.4 The Impact of Understanding the Genetic Cause of Disease

The main aim of this project was the identification of genetic causes of CHD in families with an apparent Mendelian inheritance. The importance of identifying a genetic cause for disease is clear. For individual families, the existence of a definite molecular genetic diagnosis provides important information that can be used in lifestyle decisions and family planning. It allows families to understand the inheritance pattern of the disease, the risks to relatives and the likely recurrence risks to offspring, and permits an informed decision on genetic testing in the wider family. Genetic testing might affect life-style decisions, such as diet, smoking or undertaking risk-associated activities. Genetic testing can also improve the confidence in a clinical diagnosis. Those

individuals in a family requiring health surveillance can be identified, and those without the genetic mutation can be reassured (Arndt and MacRae, 2014). A genetic diagnosis may be utilized for interventions such as prenatal diagnosis or Pre-implantation Genetic Diagnosis (PGD).

To illustrate this translational benefit, in one family recruited to this project, pathogenic homozygous mutations in *PIGV* were identified leading to a diagnosis of Hyperphosphatasia With Mental Retardation Syndrome Type 1. This finding gives the family precise information on recurrence risks, and could be used in prenatal testing. It also allows other family members to come forward for carrier testing. The parents of these children now have a name for the condition from which their children suffered, which may help in obtaining emotional support, for example through support groups such as Rare Disease UK (<https://www.raredisease.org.uk>). A recent publication from Rare Disease UK describes how patient organizations are often a main source of information for rare disease families and patients (<https://www.raredisease.org.uk/media/1594/rduk-family-report.pdf>). For future pregnancies, they can also now access more detailed prognostic information, including the likely level of developmental delay or learning difficulties in an affected child, which can be moderate to severe. Any future affected child could be provided with disease-specific support, without the need for further investigation of raised Alkaline Phosphatase levels, and clinicians could monitor for known complications of this condition such as seizures and hypotonia.

In other types of CHD, a genetic diagnosis may allow for counseling on disease prognosis, likely associated co-morbidities and appropriate management strategies. An example of this is in the case of syndromic CHD, which might be caused by a number of different conditions that include Down's syndrome, Di George syndrome, Turner's syndrome, Williams syndrome or Noonan syndrome. All have extra-cardiac anomalies and different potential managements, follow-up and screening. Noonan syndrome, for example, is associated with the risk of cardiomyopathy (LaHaye et al., 2016), which requires regular echocardiographic screening. In certain types of non-syndromic CHD, for example BAV, echocardiographic screening is recommended for first-degree relatives due to the risk of aortic dilatation (Nistri et al., 2005). Identification of a genetic mutation may therefore enable targeted screening in a family. This is further highlighted by the identification of *PPA2* mutations in the families described in chapter 4 of this thesis. The identification of the genetic cause in a family affected by this disease will allow the screening of family members for carrier status,

echocardiographic screening for cardiomyopathy, lifestyle modification (for example, the avoidance of alcohol) and ICD implantation in at-risk family members which is a potentially life-saving intervention.

In the past, clinical diagnosis and management was purely based on clinical signs and symptoms. In the modern era, a genetic diagnosis can affect many aspects of clinical care, including recruitment to clinical trials, and access to therapeutic interventions, which may, in the near future, be stratified according to genetic aetiology. This so-called 'personalized medicine' is already entering clinical practice. Warfarin is a commonly used anticoagulant in cardiovascular diseases. SNPs in the *CYP2C9* gene can affect a patient's metabolism of this drug, so undertaking genetic analysis can help to advise the clinician on dosage regimes (Voora et al., 2005).

The genetic diagnosis of disease will allow researchers to follow-up a cohort of molecularly-defined patients to determine prognosis, therapeutic outcomes and co-morbidities (as, for example, in the field of human cardiomyopathies) (Lakdawala, 2013). Identification of the genetic causes of disease can lead to an increased scientific understanding of normal developmental pathways, the factors affected in disease processes, and ways in which these factors may be corrected. This leads towards potential therapeutics, the ultimate aim of disease research. An example of this is in the identification of mutations in *CEP290* in patients with Leber's Congenital Amaurosis (LCA), an inherited eye disease (den Hollander et al., 2006). This finding has provided novel insights into ciliary biology, and has led the way towards antisense oligonucleotide therapy to redirect normal splicing in patient cell lines (Collin et al., 2012), and lentivirus transfer of full-length *CEP290* into patient-derived iPSC photoreceptor precursor cells (Burnight et al., 2014), revealing the potential for this type of therapy to treat *CEP290*-associated LCA.

The advantages of genetic testing should be weighed against the potential negative impacts, including altered personal or social perceptions, such as stigmatization or discrimination, anxiety, the possibility of false-positive or false-negative results, incidental findings and uncertain outcomes (Katsanis and Katsanis, 2013). Genetic testing should therefore always be performed in the context of support from a qualified health-care professional, such as a clinical geneticist or genetic counselor (Knoppers et al., 2015).

7.5 WES in the Clinical Setting

NGS sequencing is now a routine genetic test in the healthcare setting (Katsanis and Katsanis, 2013). In the UK, NHS testing is currently mainly focused on gene panels, which are tailored to particular phenotypes or diseases. Compared to WES, targeted reagent panels have higher throughput and lower costs, and can provide a greater depth of coverage of the selected genes. Furthermore, certain centres can develop expertise in variant interpretation, reducing the identification of variants of unknown significance (Thomas et al., 2017).

However, NGS panels may miss deep intronic, regulatory and splice variants, and developing a focused clinical panel of genes can be difficult owing to the variability of clinical presentations. This has led to the increased use of WES in the clinical setting, mainly in the form of clinical and research collaborations, for example the DDD project (Firth et al., 2011), which focused on WES in a trio-based approach alongside microarray analysis.

For WES to be used appropriately in the clinical setting, robust filtering strategies and quality control should be in place, to reduce the number of incidental findings and variants of uncertain pathogenicity identified (Bertier et al., 2016). One of the main issues is in differentiating benign polymorphisms from disease-causing variants. Advances in analytical software tools and bioinformatics programs can go some way to support these decisions, but ultimately the development of high-throughput functional experiments to assess variants may be required if WES is to be in main-stream use (for example (Thouvenot et al., 2016)). Examples might include genome-wide phenotype screens, high-throughput massively-parallel reporter assays or *in vitro* splicing assays (Blaby-Haas and de Crecy-Lagard, 2011, Ipe et al., 2017).

Guidelines have been created to standardize the assessment of variants, with the use of five reporting categories, but this assessment still contains many challenges (Richards et al., 2015). In order to address the potential ethical and moral issues associated with the reporting of incidental or secondary findings in the clinical setting a number of guidelines have been published. The most influential guidelines have been the American College of Genetics and Genomics Guidelines, published in 2013 (Green et al., 2013), which were extensively debated (Allyse and Michie, 2013) and then subsequently updated in 2014 (ACMG Board of Directors, 2015). The ACMG guidelines initially suggested a list of 56 genes should be searched for “actionable

variants” and reported back to patients, but this is a very contentious issue. The European Society of Human Genetics issued its own statement in 2015 (Hehir-Kwa et al., 2015), which highlights the key challenges in reaching a consensus on the reporting of incidental findings. It is generally thought that only “clinically actionable” incidental findings should be reported back to patients, or those that are likely to have serious consequences for future generations (Green et al., 2013). However, these guidelines remove individual patient choice (Holtzman, 2013), and have received considerable criticism due to the possibility of the identification of variants of unknown significance even in well-studied genes. For example, variants in *BRCA1* and *BRCA2*, genes that are very well-studied and associated with a high risk of breast and ovarian cancers, can still be difficult to interpret (Rodriguez-Balada et al., 2016).

Other issues for the use of WES in a clinical setting include the importance of developing secure systems for data storage, the development of robust reporting standards and guidelines on data ownership and privacy (Bertier et al., 2016). Furthermore, a key factor in the implementation of WES in the clinic is the standardized communication of results to patients (Bertier et al., 2016).

In late 2012, the UK government launched the 100,000 Genomes Project. Run by Genomics England, this project aims to sequence the whole genomes of 100,000 NHS patients by 2018. The project will focus on two main groups of patients: those with cancer and those with rare diseases. The rare disease group will focus on WGS of trios. Syndromic CHD, familial CHD and young-onset cardiomyopathy are included in the list for rare disease recruitment. This project demonstrates a huge collaboration between clinicians, industry workers, researchers and patients and will, no doubt, change the face of clinical genetic testing in the UK. However, the project is likely to face many challenges, including many of those described for WES, such as delivering a consistent quality of clinical phenotype, sequencing, data annotation and filtering, and data storage and security. The vast amounts of data will need to be interpreted in a manner that is suitable for feedback of results to patients, including clear guidance on the appropriate reporting of incidental findings and variants of unknown significance (<https://www.genomicsengland.co.uk>).

7.6 The Future of CHD Genetics

The current known genetic causes of CHD only account for around 30% of cases, leaving a large proportion of CHD unsolved (Homsy et al., 2015). These unsolved cases may include those with variants in genes that are poorly annotated or not

covered by conventional sequencing techniques. These include variants in regulatory regions outside of exons, variants in unidentified enhancers or promoters, or synonymous variants that act as splice enhancers or silencers (Triedman and Newburger, 2016). Intragenic CNVs may also be currently under-identified in these families. WGS studies may help to delineate the causes in some of these cases.

However, it may be that a large proportion of CHD is not monogenetic, but caused by the action of multiple genetic modifiers, and could therefore be characterized as a complex disease. Identification of these modifiers and the genetic backgrounds on which they work will require large cohorts of patients to create the required statistical power. For example, in the case of Cystic Fibrosis (CF), studies of more than 500 patients are required for sufficient power to identify modifier genes, and there is often poor replication between studies of this size. One review suggests that modifiers with “probable effect” are those identified in at least three independent populations of >1000 participants (Cutting, 2010). This size of CHD cohort is now being collated, for example the NHLBI Pediatric Cardiac Genomics Consortium (Pediatric Cardiac Genomics Consortium et al., 2013). GWAS studies have already been performed in some CHD cohorts, for example in left-sided lesions (Hanchard et al., 2016), but further studies will be required to delineate genetic causes in this manner.

Transcriptome analysis of cardiac tissues, using techniques such as RNA-seq, may help to identify transcriptional changes that are associated with changes in haemodynamic load and that are implicated in post-transcriptional RNA processing during early cardiac development (Wang et al., 2016b). However, these techniques also depend on the development of robust data analysis pipelines.

One of the main challenges in genetics in the modern era is not just the identification of genetic variants, but in their functional interpretation. Improved cellular models of cardiac disease, through the use of induced pluripotent stem cell lines and DNA manipulation technologies, such as CRISPR, offer exciting new resources to study the functional effects of mutations on cardiac development.

Patient-derived induced pluripotent stem cells (iPSC) are an important new tool in the study of human disease. Using a combination of reprogramming factors, adult somatic cells, such as skin fibroblasts, can now be converted into pluripotent stem cells (Takahashi et al., 2007). These can then be directed into other cell types, such as cardiomyocytes (Lian et al., 2014, Zhang et al., 2009). iPSC-derived cardiomyocytes

demonstrate the immunohistochemistry, cardiac gene expression and action potential profiles of normal cardiomyocytes (Zhang et al., 2009). Granata *et al.*, 2017, recently demonstrated the development of patient-derived hiPSC smooth muscle cell lines as a model of Marfan syndrome. These cells modeled the aortic pathology observed in patients with this condition. Through experiments on these cells lines, they were able to further delineate the role of TGF-beta in the development of the aortic pathology and identified a role for KLF4 in the Marfan's syndrome disease process (Granata et al., 2017). Patient-derived cell lines such as these can provide huge insights into disease physiology. However, patient-derived cardiomyocytes do display delayed maturation when compared to embryonic stem cells and it is unclear whether this is for technical or biological reasons (Blazeski et al., 2012). Other issues with iPSC-derived cells include the creation of mixed and variable populations of differentiated cell lineages, and interclonal variation even for cell lines derived from the same patient. These are presumably due to epigenetic effects (Gopalakrishnan et al., 2017).

CRISPR-Cas9 is a system of immunity in bacteria and *Archaea*, used to respond to invading genetic material. CRISPRs are Clustered Regularly Interspaced Short Palindromic Repeats and Cas genes are CRISPR-associated nucleases. CRISPR repeats were initially identified in *E. coli* in the 1980s (Ishino et al., 1987), but it wasn't until recently that their function was confirmed. In the CRISPR system, invading genetic material is identified, cut into small fragments and incorporated into the genome within the repeat sections of the CRISPR locus. When the sequence is transcribed, a CRISPR-RNA (cr-RNA) is generated (Jinek et al., 2012). This is used to guide effector endonucleases that target the invading DNA. The Cas9 endonuclease is a key player in this type of CRISPR system and participates in the destruction of invading DNA by the creation of double-stranded breaks and in processing the cr-RNAs. For DNA cleavage, the Cas9 enzyme must be in the presence of crRNA and a separate trans-activating crRNA (trRNA) (Jinek et al., 2012). Also, a protospacer-associated motif (PAM), which is a 2-5 nucleotide motif, should be present immediately 3' to the cr-RNA sequence (Swarts et al., 2012). This system has been adapted for use as a molecular tool, using a synthetic single guide RNA (sgRNA, a combination of cr-RNA and tr-RNA sequence), along with Cas9, to create double-strand breaks at any specified location in the genome (Jinek et al., 2012). These double-strand breaks are either repaired by non-homologous end joining (NHEJ), resulting in insertions and deletions that disrupt gene function, or homology-directed repair (HDR), which can be exploited if a donor template is also provided, resulting in a specified substitution (Gong et al., 2005).

More recently, a mutant Cas9 enzyme was created, Cas9D10A (nickase), which only cleaves one DNA strand, reducing NHEJ and increasing HDR, resulting in a reduced number of indels (Cong et al., 2013). Using this technology, loci can be targeted by paired Cas-9 nickases to create targeted double-strand breaks with improved specificity and a decrease in off-target effects (Ran et al., 2013). Furthermore, a nuclease-deficient Cas9 enzyme (dCas9) without cleavage activity has been created, which can bind to specific sequences, but does not disrupt the gene sequence. This can then be used to manipulate gene expression, also called CRISPR interference (CRISPR-i) (Qi et al., 2013).

The targeting efficiency of CRISPR is much higher than that of previous technologies such as Zinc Finger Nucleases (ZNFs) and transcription activator-like effectors (TALENs) (Mussolino et al., 2011). However, a major challenge exists in the creation of off-target effects, as the Cas9 enzyme can tolerate a number of mismatches in the target sequence (Fu et al., 2013). A recent paper highlighted the concerns regarding off-target effects in CRISPR-Cas9-edited mice. By using WGS, the authors showed a high number of off-target indels and SNVs in these mice, raising fears regarding the fidelity of the CRISPR-Cas9 system in an *in vivo* model (Schaefer et al., 2017). These off-target mutations can be difficult to detect and new developments, such as the use of paired nickases, are being used to reduce their effect.

CRISPR-Cas9 technology is now being used in exciting new ways such as epigenetic modulation (Liu et al., 2016) and microscopic dynamic imaging (Chen et al., 2013). Perhaps one of the most exciting lines of work will be in the creation of animal models of disease. For example, the CRISPR-Cas9 system was recently used in zebrafish embryos to model *MMP21* mutations, which caused cardiac looping defects, consistent with the observed phenotype in human patients (Perles et al., 2015). CRISPR-Cas9 can also be used in genome-wide screening, similar to previous RNAi-based screens, but with the advantages of gene-level effects and the ability to assess non-transcribed elements (Koike-Yusa et al., 2014).

Animal models, such as those studied in this project, including zebrafish and mice, can aid in the assessment of variants *in vivo*. Future work is likely to continue to rely on these animal models, possibly combined with CRISPR technologies to study the pathogenicity of variants in their effect on normal development. Zebrafish are often used in CHD research due to the transparency of the embryo, and their ease of use in

genetic manipulation, which combined with CRISPR-Cas9 technologies provides an excellent research tool for cardiac development studies (Wilkinson et al., 2014). They also have a rapid embryonic development, high reproductive capacity, and a short generation time meaning a high-throughput of experimental results. There is good conservation between the human and zebrafish genomes, but teleost fish have undergone genome-wide duplication, so it is important to consider the physiological differences between species in the assessment of pathogenic variants. Since zebrafish are evolutionarily distinct from humans, disease phenotypes may not always be consistent (Guyon et al., 2007).

7.7 Potential Therapeutics in Genetic Disease

At the current time, patients affected by genetic disease have options to prevent passing a disease-causing mutation on to subsequent generations. Prenatal testing can be performed at an early stage of pregnancy to allow for parental choice regarding the continuation of the pregnancy. Testing for aneuploidies and single gene disorders can now be performed by Non Invasive Prenatal Testing (NIPT), analyzing cell-free DNA from a maternal blood sample, as early as 7 weeks gestation (Skrzypek and Hui, 2017). However, to prevent passing a disease-causing gene on to offspring would involve the termination of an affected pregnancy. Pre-implantation Genetic Diagnosis (PGD) allows genetic testing of an embryo, created by in vitro fertilization (IVF), prior to implantation in the womb. A number of structural, muscular and arrhythmic cardiac disorders are included on the Human Fertilization and Embryology Authority's list of conditions permissible for embryo selection, for example Cardiac Valvular Dysplasia and Dilated Cardiomyopathy and Left Ventricular Noncompaction Type 5. Perhaps unsurprisingly, conditions that do not affect life expectancy, such as VSD, are not included (<http://guide.hfea.gov.uk/pgd/>). However, these techniques do carry risks such as potential IVF failure, increased risk of multiple births and ovarian hyper-stimulation syndrome (Kumar et al., 2011).

Theoretically, in the future, CRISPR-Cas9 technologies could be used to manipulate embryos prior to implantation. In fact, a Chinese group have already demonstrated that this could be a possibility. By injecting one-cell human embryos with Cas9 and sgRNA, they show correction of point mutations in the *HBB* and *G6PD* genes (Tang et al., 2017). However the group met with many challenges: there was a poor mutation correction rate and many off-target effects were observed, higher even than those seen in other animals such as mice, creating many safety concerns (Schaefer et al., 2017). There are strong moral, ethical and legal concerns regarding the use of genetic

manipulation in human embryos, as any mutations created would be passed on to future generations. Off-target effects will have an unknown clinical outcome, which may not be apparent for many years. Furthermore, even if these techniques become feasible, for which diseases should they be used? Should parents be allowed to make such decisions for their unborn children? Do such technologies open the door to eugenic manipulation? (Lanphier et al., 2015) Current UK laws do not allow the culture of human embryos past 14 days gestation (Warnick, 1984) but the Human Fertilization and Embryology Authority has recently approved a license to allow CRIPSR gene editing in early human embryos (<http://www.hfea.gov.uk/10187.html>). It remains to be seen to what extent CRISPR technologies will be used in genetic editing of human embryos, and in the targeting of adult disease. In adult disease, much of the current work is focused on genetic manipulation of somatic cells, such as T-cells, in the treatment of viral infections such as HIV (Patel et al., 2016).

Other current research focuses not only on the prevention of genetic disease in offspring but also on treatment in affected patients. Many genetic diseases, including some types of CHD, are caused by premature stop codons that terminate the amino acid sequence leading to premature protein truncation. These premature stop codons are not as efficient as normal stop codons and sometimes 'readthrough' can occur, meaning that the stop codon is ignored and the sequence is translated (Keeling et al., 2014). This discovery has led to the development of nonsense suppression therapy, which encourages ribosomal read-through of the premature stop codon to create a normal protein sequence. Therapeutic agents that can cause nonsense suppression include aminoglycoside antibiotics, such as gentamicin, non-aminoglycoside antibiotics such as negamycin (Arakawa et al., 2003) and other agents such as PTC124 (Ataluren). Aminoglycosides are associated with toxic effects in long-term use, such as hearing loss and kidney damage, and clinical trials have produced variable results, restoring protein function in only a fraction of patients (Clancy et al., 2001). Researchers are now attempting to create aminoglycosides with reduced toxicity, for example using encapsulation in liposomes (Schiffelers et al., 2001). Ataluren is an oxadiazole compound that acts as a nonsense suppressor. It has been shown to be safe in human trials and to have minimal side effects (Hirawat et al., 2007). Numerous *in vitro* studies have confirmed its ability to suppress nonsense mutations, for example in the *CFTR* gene (Gonzalez-Hilarion et al., 2012). Recent double-blind, placebo-controlled phase 3 clinical trials have been performed in Cystic Fibrosis (CF) and Duchenne Muscular Dystrophy (DMD). Some symptom relief was recognized in the CF group, although not to a significant level (Kerem et al., 2014) and a recent Cochrane

review of the use of Ataluren in CF shows insufficient evidence to support its use (Aslam et al., 2017). In DMD trials, Ataluren has shown some promising results, including an improvement in the 6-minute walk test (Bushby et al., 2014). However, further research will be needed to identify a role for these agents in the treatment of human disease. Other approaches to mediate premature stop codon readthrough include the use of suppressor tRNAs (Temple et al., 1982) and by using pseudouridylation modification of the first base of a stop codon (Karijolic and Yu, 2011).

The concept of 'gene therapy' or the introduction of normal genes to human cells has been around for over 20 years (Rosenberg et al., 1990). Vectors for gene delivery include viral vectors such as lentivirus or adeno-associated virus (AAV), or non-viral vectors such as nanoparticles (Nance et al., 2017). Some of the earliest human trials of gene therapy were performed in X-linked Severe Combined Immunodeficiency Disorder (SCID), using a retroviral vector, but suffered a major set-back when a number of the treated patients developed leukaemia, raising safety concerns. However at the 10 year review, 8 of the 9 treated patients were still alive (Hacein-Bey-Abina et al., 2003). AAVs do not integrate into the host genome and so are not associated with the same safety risks. Recent successes using viral vectors in animal models has led the way for gene therapy trials in a number of human diseases, for example primary immunodeficiencies, leukodystrophies, thalassaemia, haemophilia and retinal dystrophies such as Leber's Congenital Amaurosis (LCA) (Weleber et al., 2016). In these trials, genetic material is either transferred *ex vivo* (for example to T-lymphocytes) and the cells are then returned to the body, or targeting is *in vivo* (for example in retinal disease) (Naldini, 2015). For *in vivo* work, eye diseases are often targeted due to the small, enclosed nature of the globe, with a tight blood-ocular barrier that can be easily accessed by injection and surgery (Samiy, 2014). The development of gene therapy in cardiac disease is less straightforward. For myocardial disease, therapy could conceivably be delivered directly to the heart using myocardial injection (Jones and Koch, 2005), but to prevent a structural disease, such as CHD, the developing heart would need to be targeted *in utero* at a very early developmental stage. Gene therapy still faces many challenges, for example in the development of efficient and safe vectors and in improving specificity (Samiy, 2014).

iPSC cells are also being targeted as a therapeutic agent, for example patient-derived iPSC cells from a patient with cardiovascular disease could be used to generate functional cardiomyocytes to replace diseased heart tissue. Immune rejection would be

minimized as the cell lines originate from the patient's own tissue. As proof of principle, iPSC cells from patients with Amyotrophic Lateral Sclerosis (ALS) have been differentiated into motor neurons, which could be used in autologous cell replacement therapies (Dimos et al., 2008).

7.8 Plans for Future Research

There are a number of ways in which the work from this project could be continued in further studies. There remain a number of unsolved cases of CHD that could initiate further investigation. WGS would be a good next step in investigation of families where no clear candidate is identified, since this would assess non-coding regions of the genome and enable further detailed structural analysis. All recruited unsolved families will be offered recruitment to the 100,000 Genomes Project, via the consenting clinician, to ensure that WGS is available to them if so desired.

Some of the candidate variants identified in this project could be further assessed to gain insights into likely pathogenicity, perhaps in collaboration with other groups. Examples include the variants identified in *PKHD1L1* for family CHD3 and in *RBM48* for family CHD16, which have been uploaded to Matchmaker Exchange (<http://www.matchmakerexchange.org>). It may be that identification of additional families with a similar phenotype and variants in these genes would provide the necessary support of pathogenicity to endorse further functional work.

The functional assessment of mutations in *DNAAF1*, as detailed in chapter 5, has led to a number of questions that could be further investigated. The exact role of *DNAAF1* within the Pontin/Reptin complex is not clear. Furthermore, the exciting finding of an asymmetric distribution of Pontin at the embryonic node requires further investigation. This could be taken forward using immunofluorescence labeling of other asymmetrically expressed genes, for example *Nodal* and *Pitx2*, at the Kupffer's Vesicle in zebrafish to classify the hierarchy of Pontin expression in early development. It would also be very interesting to model the effect of loss of Pontin in the developing zebrafish, perhaps using a CRISPR model. This could be used to identify the effect on heart looping and expression of downstream laterality markers at the KV.

The variant identified in *KMT2D* in chapter 6 in two families affected with athelia, choanal atresia and CHD deserves further investigation. This is a fascinating phenotype, but is very different to the usual phenotype caused by mutations in the gene (Kabuki Syndrome). Functional studies could be used to assess the pathogenicity

of this variant and to investigate the reason for the phenotype observed. *KMT2D* is a very large protein, making functional studies, such as over-expression experiments to model the mutation, very challenging. Therefore, having access to patient fibroblasts or creating a lymphoblastoid cell line from patient blood would be a key next step in the analysis of these families. A recent study assessed gene-specific histone methylation signatures of patients with *KMT2D* mutations, demonstrating *KMT2D* to be involved in the histone methylation of gene targets such as *HOXA5* (Butcher et al., 2017). One suggestion would be to investigate the methylation of these gene targets in cell-lines from our patients and to identify whether a similar methylation signature is present. This would help to delineate the pathogenicity of this variant and may help to explain the effect of the variant on protein function.

7.9 Overall Conclusions

The hypothesis-free approach of WES has taken this project in a number of exciting and unexpected directions. New disease genes have been identified for early-onset cardiomyopathy and heterotaxy, and a number of candidate genes for structural CHD have been described. Further work is required to fully understand the complex genetic aetiology of CHD, but this study demonstrates that novel genes can be identified in a family-based approach. The field of genetics is now moving very rapidly and it is likely that CHD genetics will take huge steps forward with the introduction of new cellular and animal models, methods of genome modification, larger cohorts of patients and with the introduction of WGS approaches.

References

- Abu-Tair, T., Wiethoff, C. M., Kehr, J., Kuroczynski, W. & Kampmann, C. 2016. Transcatheter Closure of Atrial Septal Defects using the GORE((R)) Septal Occluder in Children Less Than 10 kg of Body Weight. *Pediatr Cardiol*, 37, 778-83.
- Acmg Board of Directors 2015. ACMG policy statement: updated recommendations regarding analysis and reporting of secondary findings in clinical genome-scale sequencing. *Genet Med*, 17, 68-9.
- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., et al. 2010. A method and server for predicting damaging missense mutations. *Nat Methods*, 7, 248-9.
- Agergaard, P., Olesen, C., Ostergaard, J. R., Christiansen, M. & Sorensen, K. M. 2012. The prevalence of chromosome 22q11.2 deletions in 2,478 children with cardiovascular malformations. A population-based study. *Am J Med Genet A*, 158A, 498-508.
- Ahokas, K., Lohi, J., Lohi, H., Elomaa, O., Karjalainen-Lindsberg, M. L., et al. 2002. Matrix metalloproteinase-21, the human orthologue for XMMP, is expressed during fetal development and in cancer. *Gene*, 301, 31-41.
- Ahrens-Nicklas, R. C., Khan, S., Garbarini, J., Woyciechowski, S., D'alessandro, L., et al. 2016. Utility of genetic evaluation in infants with congenital heart defects admitted to the cardiac intensive care unit. *Am J Med Genet A*, 170, 3090-3097.
- Akawi, N., Mcrae, J., Ansari, M., Balasubramanian, M., Blyth, M., et al. 2015. Discovery of four recessive developmental disorders using probabilistic genotype and phenotype matching among 4,125 families. *Nat Genet*, 47, 1363-9.
- Al Turki, S., Manickaraj, A. K., Mercer, C. L., Gerety, S. S., Hitz, M. P., et al. 2014. Rare variants in NR2F2 cause congenital heart defects in humans. *Am J Hum Genet*, 94, 574-85.
- Al-Gazali, L. I., Hamid, Z., Hertecant, J., Bakir, M., Nath, D., et al. 2002. An autosomal recessive syndrome of choanal atresia, hypothelia/athelia and thyroid gland anomalies overlapping bamforth syndrome, ANOTHER syndrome and methimazole embryopathy. *Clin Dysmorphol*, 11, 79-85.
- Allyse, M. & Michie, M. 2013. Not-so-incident findings: the ACMG recommendations on the reporting of incidental findings in clinical whole genome and whole exome sequencing. *Trends Biotechnol*, 31, 439-41.
- Amieux, P. S., Howe, D. G., Knickerbocker, H., Lee, D. C., Su, T., et al. 2002. Increased basal cAMP-dependent protein kinase activity inhibits the formation of mesoderm-derived structures in the developing mouse embryo. *J Biol Chem*, 277, 27294-304.
- Andersen, T. A., Troelsen, K. D. & Larsen, L. A. 2013. Of mice and men: molecular genetics of congenital heart disease. *Cell Mol Life Sci*.
- Ang, S. Y., Uebersohn, A., Spencer, C. I., Huang, Y., Lee, J. E., et al. 2016. KMT2D regulates specific programs in heart development via histone H3 lysine 4 dimethylation. *Development*, 143, 810-21.
- Angers, M., Uldry, M., Kong, D., Gimble, J. M. & Jetten, A. M. 2008. Mfsd2a encodes a novel major facilitator superfamily domain-containing protein highly induced in brown adipose tissue during fasting and adaptive thermogenesis. *Biochem J*, 416, 347-55.
- Ansari, K. I., Hussain, I., Shrestha, B., Kasiri, S. & Mandal, S. S. 2011. HOXC6 Is transcriptionally regulated via coordination of MLL histone methylase and estrogen receptor in an estrogen environment. *J Mol Biol*, 411, 334-49.

- Arakawa, M., Shiozuka, M., Nakayama, Y., Hara, T., Hamada, M., et al. 2003. Negamycin restores dystrophin expression in skeletal and cardiac muscles of mdx mice. *J Biochem*, 134, 751-8.
- Aramaki, M., Udaka, T., Kosaki, R., Makita, Y., Okamoto, N., et al. 2006. Phenotypic spectrum of CHARGE syndrome with CHD7 mutations. *J Pediatr*, 148, 410-4.
- Armstrong, L., Abd El Moneim, A., Aleck, K., Aughton, D. J., Baumann, C., et al. 2005. Further delineation of Kabuki syndrome in 48 well-defined new individuals. *Am J Med Genet A*, 132A, 265-72.
- Arndt, A. K. & Macrae, C. A. 2014. Genetic testing in cardiovascular diseases. *Curr Opin Cardiol*, 29, 235-40.
- Aslam, A. A., Higgins, C., Sinha, I. P. & Southern, K. W. 2017. Ataluren and similar compounds (specific therapies for premature termination codon class I mutations) for cystic fibrosis. *Cochrane Database Syst Rev*, 1, CD012040.
- Ausavarat, S., Tongkobpetch, S., Praphanphoj, V., Mahatumarat, C., Rojvachiranonda, N., et al. 2011. PTPRF is disrupted in a patient with syndromic amastia. *BMC Med Genet*, 12, 46.
- Baban, A., Postma, A. V., Marini, M., Trocchio, G., Santilli, A., et al. 2014. Identification of TBX5 mutations in a series of 94 patients with Tetralogy of Fallot. *Am J Med Genet A*, 164A, 3100-7.
- Badalato, L., Farhan, S. M., Dillio, A. A., Care4rare Canada, C., Bulman, D. E., et al. 2017. KMT2D p.Gln3575His segregating in a family with autosomal dominant choanal atresia strengthens the Kabuki/CHARGE connection. *Am J Med Genet A*, 173, 183-189.
- Bajpai, R., Chen, D. A., Rada-Iglesias, A., Zhang, J., Xiong, Y., et al. 2010. CHD7 cooperates with PBAF to control multipotent neural crest formation. *Nature*, 463, 958-62.
- Bamforth, J. S., Hughes, I. A., Lazarus, J. H., Weaver, C. M. & Harper, P. S. 1989. Congenital hypothyroidism, spiky hair, and cleft palate. *J Med Genet*, 26, 49-51.
- Banka, S., Veeramachaneni, R., Reardon, W., Howard, E., Bunstone, S., et al. 2012. How genetically heterogeneous is Kabuki syndrome?: MLL2 testing in 116 patients, review and analyses of mutation and phenotypic spectrum. *Eur J Hum Genet*, 20, 381-8.
- Basten, S. G., Davis, E. E., Gillis, A. J., Van Rooijen, E., Stoop, H., et al. 2013. Mutations in LRRC50 predispose zebrafish and humans to seminomas. *PLoS Genet*, 9, e1003384.
- Bauer, A., Chauvet, S., Huber, O., Usseglio, F., Rothbacher, U., et al. 2000. Pontin52 and reptin52 function as antagonistic regulators of beta-catenin signalling activity. *EMBO J*, 19, 6121-30.
- Baumgartner, H., Bonhoeffer, P., De Groot, N. M., De Haan, F., Deanfield, J. E., et al. 2010. ESC Guidelines for the management of grown-up congenital heart disease (new version 2010). *Eur Heart J*, 31, 2915-57.
- Becker, S. & Al Halees, Z. 1999. First-cousin matings and congenital heart disease in Saudi Arabia. *Community Genet*, 2, 69-73.
- Belkadi, A., Bolze, A., Itan, Y., Cobat, A., Vincent, Q. B., et al. 2015. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A*, 112, 5473-8.
- Benson, D. W., Silberbach, G. M., Kavanaugh-Mchugh, A., Cottrill, C., Zhang, Y., et al. 1999. Mutations in the cardiac transcription factor NKX2.5 affect diverse cardiac developmental pathways. *J Clin Invest*, 104, 1567-73.
- Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456, 53-9.
- Bertier, G., Hetu, M. & Joly, Y. 2016. Unsolved challenges of clinical whole-exome sequencing: a systematic literature review of end-users' views. *BMC Med Genomics*, 9, 52.

- Bettinelli, A. L., Mulder, T. J., Funke, B. H., Lafferty, K. A., Longo, S. A., et al. 2013. Familial ebstein anomaly, left ventricular hypertrabeculation, and ventricular septal defect associated with a MYH7 mutation. *Am J Med Genet A*.
- Bevilacqua, A., Willis, M. S. & Bultman, S. J. 2013. SWI/SNF chromatin-remodeling complexes in cardiovascular development and disease. *Cardiovasc Pathol*.
- Bichet, D., Peters, D., Patel, A. J., Delmas, P. & Honore, E. 2006. Cardiovascular polycystins: insights from autosomal dominant polycystic kidney disease and transgenic animal models. *Trends Cardiovasc Med*, 16, 292-8.
- Blaby-Haas, C. E. & De Crecy-Lagard, V. 2011. Mining high-throughput experimental data to link gene and function. *Trends Biotechnol*, 29, 174-82.
- Blainey, P., Krzywinski, M. & Altman, N. 2014. Points of significance: replication. *Nat Methods*, 11, 879-80.
- Blazeski, A., Zhu, R., Hunter, D. W., Weinberg, S. H., Zambidis, E. T., et al. 2012. Cardiomyocytes derived from human induced pluripotent stem cells as models for normal and diseased cardiac electrophysiology and contractility. *Prog Biophys Mol Biol*, 110, 166-77.
- Blue, G. M., Humphreys, D., Szot, J., Major, J., Chapman, G., et al. 2017. The promises and challenges of exome sequencing in familial, non-syndromic congenital heart disease. *Int J Cardiol*, 230, 155-163.
- Blue, G. M., Kirk, E. P., Giannoulatou, E., Dunwoodie, S. L., Ho, J. W., et al. 2014. Targeted next-generation sequencing identifies pathogenic variants in familial congenital heart disease. *J Am Coll Cardiol*, 64, 2498-506.
- Boogerd, C. J., Moorman, A. F. & Barnett, P. 2009. Protein interactions at the heart of cardiac chamber formation. *Ann Anat*, 191, 505-17.
- Borck, G., De Vries, L., Wu, H. J., Smirin-Yosef, P., Nurnberg, G., et al. 2014. Homozygous truncating PTPRF mutation causes athelia. *Hum Genet*, 133, 1041-7.
- Boulon, S., Pradet-Balade, B., Verheggen, C., Molle, D., Boireau, S., et al. 2010. HSP90 and its R2TP/Prefoldin-like cochaperone are involved in the cytoplasmic assembly of RNA polymerase II. *Mol Cell*, 39, 912-24.
- Boycott, K. M., Vanstone, M. R., Bulman, D. E. & Mackenzie, A. E. 2013. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet*, 14, 681-91.
- Brambilla, R., Gnesutta, N., Minichiello, L., White, G., Roylance, A. J., et al. 1997. A role for the Ras signalling pathway in synaptic transmission and long-term memory. *Nature*, 390, 281-6.
- Braun, D. A. & Hildebrandt, F. 2017. Ciliopathies. *Cold Spring Harb Perspect Biol*, 9.
- Braun, D. A., Schueler, M., Halbritter, J., Gee, H. Y., Porath, J. D., et al. 2016. Whole exome sequencing identifies causative mutations in the majority of consanguineous or familial cases with childhood-onset increased renal echogenicity. *Kidney Int*, 89, 468-75.
- Breckpot, J., Thienpont, B., Arens, Y., Tranchevent, L. C., Vermeesch, J. R., et al. 2011. Challenges of interpreting copy number variation in syndromic and non-syndromic congenital heart defects. *Cytogenet Genome Res*, 135, 251-9.
- Breig, O., Bras, S., Martinez Soria, N., Osman, D., Heidenreich, O., et al. 2014. Pontin is a critical regulator for AML1-ETO-induced leukemia. *Leukemia*, 28, 1271-9.
- Breuss, M., Heng, J. I., Poirier, K., Tian, G., Jaglin, X. H., et al. 2012. Mutations in the beta-tubulin gene TUBB5 cause microcephaly with structural brain abnormalities. *Cell Rep*, 2, 1554-62.
- Bruneau, B. G., Nemer, G., Schmitt, J. P., Charron, F., Robitaille, L., et al. 2001. A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. *Cell*, 106, 709-21.
- Burgers, P. P., Ma, Y., Margarucci, L., Mackey, M., Van Der Heyden, M. A., et al. 2012. A small novel A-kinase anchoring protein (AKAP) that localizes specifically

- protein kinase A-regulatory subunit I (PKA-RI) to the plasma membrane. *J Biol Chem*, 287, 43789-97.
- Burnight, E. R., Wiley, L. A., Drack, A. V., Braun, T. A., Anfinson, K. R., et al. 2014. CEP290 gene transfer rescues Leber congenital amaurosis cellular phenotype. *Gene Ther*, 21, 662-72.
- Bushby, K., Finkel, R., Wong, B., Barohn, R., Campbell, C., et al. 2014. Ataluren treatment of patients with nonsense mutation dystrophinopathy. *Muscle Nerve*, 50, 477-87.
- Butcher, D. T., Cytrynbaum, C., Turinsky, A. L., Siu, M. T., Inbar-Feigenberg, M., et al. 2017. CHARGE and Kabuki Syndromes: Gene-Specific DNA Methylation Signatures Identify Epigenetic Mechanisms Linking These Clinically Overlapping Conditions. *Am J Hum Genet*, 100, 773-788.
- C. Yuen, R., Merico, D., Bookman, M., J. L. H., Thiruvahindrapuram, B., et al. 2017. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat Neurosci*, 20, 602-611.
- Cao, Y. & Aypar, U. 2016. A novel Xq22.1 deletion in a male with multiple congenital abnormalities and respiratory failure. *Eur J Med Genet*, 59, 274-7.
- Castanet, M. & Polak, M. 2010. Spectrum of Human Foxe1/TF2 Mutations. *Horm Res Paediatr*, 73, 423-9.
- Cecchi, F., Tomberli, B. & Olivetto, I. 2012. Clinical and molecular classification of cardiomyopathies. *Glob Cardiol Sci Pract*, 2012, 4.
- Chaix, M. A., Andelfinger, G. & Khairy, P. 2016. Genetic testing in congenital heart disease: A clinical approach. *World J Cardiol*, 8, 180-91.
- Chalhoub, V., Abi-Rafeh, L., Hachem, K., Ayoub, E. & Yazbeck, P. 2013. Intracranial aneurysm and recessive polycystic kidney disease: the third reported case. *JAMA Neurol*, 70, 114-6.
- Chan, P. A., Duraisamy, S., Miller, P. J., Newell, J. A., McBride, C., et al. 2007. Interpreting missense variants: comparing computational methods in human disease genes CDKN2A, MLH1, MSH2, MECP2, and tyrosinase (TYR). *Hum Mutat*, 28, 683-93.
- Chang, C. P. & Bruneau, B. G. 2012. Epigenetics and cardiovascular development. *Annu Rev Physiol*, 74, 41-68.
- Chapman, A. B., Rubinstein, D., Hughes, R., Stears, J. C., Earnest, M. P., et al. 1992. Intracranial aneurysms in autosomal dominant polycystic kidney disease. *N Engl J Med*, 327, 916-20.
- Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., et al. 2013. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, 155, 1479-91.
- Chen, J. & Wang, D. Z. 2012. microRNAs in cardiovascular development. *J Mol Cell Cardiol*, 52, 949-57.
- Chen, L., Liu, T., Tran, A., Lu, X., Tomilov, A. A., et al. 2012. OPA1 mutation and late-onset cardiomyopathy: mitochondrial dysfunction and mtDNA instability. *J Am Heart Assoc*, 1, e003012.
- Chen, T., Shan, G., Ma, J. & Zhong, Y. 2015. Polymorphism in the RASGRF1 gene with high myopia: A meta-analysis. *Mol Vis*, 21, 1272-80.
- Ching, Y. H., Ghosh, T. K., Cross, S. J., Packham, E. A., Honeyman, L., et al. 2005. Mutation in myosin heavy chain 6 causes atrial septal defect. *Nat Genet*, 37, 423-8.
- Chinnery, P. F. 1993. Mitochondrial Disorders Overview. In: PAGON, R. A., ADAM, M. P., ARDINGER, H. H., WALLACE, S. E., AMEMIYA, A., BEAN, L. J. H., BIRD, T. D., LEDBETTER, N., MEFFORD, H. C., SMITH, R. J. H. & STEPHENS, K. (eds.) *GeneReviews(R)*. Seattle (WA).
- Chiyonobu, T., Inoue, N., Morimoto, M., Kinoshita, T. & Murakami, Y. 2014. Glycosylphosphatidylinositol (GPI) anchor deficiency caused by mutations in

- PIGW is associated with West syndrome and hyperphosphatasia with mental retardation syndrome. *J Med Genet*, 51, 203-7.
- Chowdhury, S., Bandholz, A. M., Parkash, S., Dyack, S., Rideout, A. L., et al. 2014. Phenotypic and molecular characterization of 19q12q13.1 deletions: a report of five patients. *Am J Med Genet A*, 164A, 62-9.
- Chung, Y. W. & Kang, S. M. 2015. An experimental approach to study the function of mitochondria in cardiomyopathy. *BMB Rep*, 48, 541-8.
- Cizkova, A., Stranecky, V., Mayr, J. A., Tesarova, M., Havlickova, V., et al. 2008. TMEM70 mutations cause isolated ATP synthase deficiency and neonatal mitochondrial encephalocardiomyopathy. *Nat Genet*, 40, 1288-90.
- Clancy, J. P., Bebok, Z., Ruiz, F., King, C., Jones, J., et al. 2001. Evidence that systemic gentamicin suppresses premature stop mutations in patients with cystic fibrosis. *Am J Respir Crit Care Med*, 163, 1683-92.
- Clementi, M., Di Gianantonio, E., Pelo, E., Mammi, I., Basile, R. T., et al. 1999. Methimazole embryopathy: delineation of the phenotype. *Am J Med Genet*, 83, 43-6.
- Coe, B. P., Ylstra, B., Carvalho, B., Meijer, G. A., Macaulay, C., et al. 2007. Resolving the resolution of array CGH. *Genomics*, 89, 647-53.
- Collin, R. W., Den Hollander, A. I., Van Der Velde-Visser, S. D., Bennicelli, J., Bennett, J., et al. 2012. Antisense Oligonucleotide (AON)-based Therapy for Leber Congenital Amaurosis Caused by a Frequent Mutation in CEP290. *Mol Ther Nucleic Acids*, 1, e14.
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339, 819-23.
- Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., et al. 2010. Origins and functional impact of copy number variation in the human genome. *Nature*, 464, 704-12.
- Cooper, G. M. & Hausman, R. E. 2016. *The Cell : a molecular approach*, Sunderland, Mass., Sinauer Associates.
- Cooper, G. M. & Shendure, J. 2011. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat Rev Genet*, 12, 628-40.
- Cordell, H. J., Bentham, J., Topf, A., Zelenika, D., Heath, S., et al. 2013. Genome-wide association study of multiple congenital heart disease phenotypes identifies a susceptibility locus for atrial septal defect at chromosome 4p16. *Nat Genet*, 45, 822-4.
- Cripe, L., Andelfinger, G., Martin, L. J., Shoener, K. & Benson, D. W. 2004. Bicuspid aortic valve is heritable. *J Am Coll Cardiol*, 44, 138-43.
- Curbo, S., Lagier-Tourenne, C., Carozzo, R., Palenzuela, L., Luciola, S., et al. 2006. Human mitochondrial pyrophosphatase: cDNA cloning and analysis of the gene in patients with mtDNA depletion syndromes. *Genomics*, 87, 410-6.
- Cutting, G. R. 2010. Modifier genes in Mendelian disorders: the example of cystic fibrosis. *Ann N Y Acad Sci*, 1214, 57-69.
- Das, G., Yu, Q., Hui, R., Reuhl, K., Gale, N. W., et al. 2016. EphA5 and EphA6: regulation of neuronal and spine morphology. *Cell Biosci*, 6, 48.
- Den Hollander, A. I., Koenekoop, R. K., Yzer, S., Lopez, I., Arends, M. L., et al. 2006. Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. *Am J Hum Genet*, 79, 556-61.
- Dentici, M. L., Di Pede, A., Lepri, F. R., Gnazzo, M., Lombardi, M. H., et al. 2015. Kabuki syndrome: clinical and molecular diagnosis in the first year of life. *Arch Dis Child*, 100, 158-64.
- Dhanoa, B. S., Cogliati, T., Satish, A. G., Bruford, E. A. & Friedman, J. S. 2013. Update on the Kelch-like (KLHL) gene family. *Hum Genomics*, 7, 13.
- Dimitrov, A., Paupe, V., Gueudry, C., Sibarita, J. B., Raposo, G., et al. 2009. The gene responsible for Dyggve-Melchior-Clausen syndrome encodes a novel peripheral

- membrane protein dynamically associated with the Golgi apparatus. *Hum Mol Genet*, 18, 440-53.
- Dimos, J. T., Rodolfa, K. T., Niakan, K. K., Weisenthal, L. M., Mitsumoto, H., et al. 2008. Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. *Science*, 321, 1218-21.
- Doenst, T., Nguyen, T. D. & Abel, E. D. 2013. Cardiac metabolism in heart failure: implications beyond ATP production. *Circ Res*, 113, 709-24.
- Dong, F., Shinohara, K., Botilde, Y., Nabeshima, R., Asai, Y., et al. 2014. Pih1d3 is required for cytoplasmic preassembly of axonemal dynein in mouse sperm. *J Cell Biol*, 204, 203-13.
- Dong, Z., Zhang, J., Hu, P., Chen, H., Xu, J., et al. 2016. Low-pass whole-genome sequencing in clinical cytogenetics: a validated approach. *Genet Med*, 18, 940-8.
- Du, W. W., Fang, L., Li, M., Yang, X., Liang, Y., et al. 2013. MicroRNA miR-24 enhances tumor invasion and metastasis by targeting PTPN9 and PTPRF to promote EGF signaling. *J Cell Sci*, 126, 1440-53.
- Dumic, M., Cvitanovic, M., Saric, B., Spehar, A. & Batinica, S. 2002. Choanal stenosis, hypothelia, deafness, recurrent dacryocystitis, neck fistulas, short stature, and microcephaly: report of a case. *Am J Med Genet*, 113, 295-7.
- Duquesnoy, P., Escudier, E., Vincensini, L., Freshour, J., Bridoux, A. M., et al. 2009. Loss-of-function mutations in the human ortholog of *Chlamydomonas reinhardtii* ODA7 disrupt dynein arm assembly and cause primary ciliary dyskinesia. *Am J Hum Genet*, 85, 890-6.
- Durocher, D. & Nemer, M. 1998. Combinatorial interactions regulating cardiac transcription. *Dev Genet*, 22, 250-62.
- El Ghouzzi, V., Dagoneau, N., Kinning, E., Thauvin-Robinet, C., Chemaitilly, W., et al. 2003. Mutations in a novel gene *Dymeclin* (FLJ20071) are responsible for Dyggve-Melchior-Clausen syndrome. *Hum Mol Genet*, 12, 357-64.
- Eldadah, Z. A., Hamosh, A., Biery, N. J., Montgomery, R. A., Duke, M., et al. 2001. Familial Tetralogy of Fallot caused by mutation in the *jagged1* gene. *Hum Mol Genet*, 10, 163-9.
- Elliott, A. M., Simard, L. R., Coghlan, G., Chudley, A. E., Chodirker, B. N., et al. 2013. A novel mutation in *KIAA0196*: identification of a gene involved in Ritscher-Schinzel/3C syndrome in a First Nations cohort. *J Med Genet*, 50, 819-22.
- Erdogan, F., Larsen, L. A., Zhang, L., Tumer, Z., Tommerup, N., et al. 2008. High frequency of submicroscopic genomic aberrations detected by tiling path array comparative genome hybridisation in patients with isolated congenital heart disease. *J Med Genet*, 45, 704-9.
- Ernst, P. & Vakoc, C. R. 2012. WRAD: enabler of the SET1-family of H3K4 methyltransferases. *Brief Funct Genomics*, 11, 217-26.
- Fakhro, K. A., Choi, M., Ware, S. M., Belmont, J. W., Towbin, J. A., et al. 2011. Rare copy number variations in congenital heart disease patients identify unique genes in left-right patterning. *Proc Natl Acad Sci U S A*, 108, 2915-20.
- Fallet-Bianco, C., Laquerriere, A., Poirier, K., Razavi, F., Guimiot, F., et al. 2014. Mutations in tubulin genes are frequent causes of various foetal malformations of cortical development including microlissencephaly. *Acta Neuropathol Commun*, 2, 69.
- Feldheim, D. A., Nakamoto, M., Osterfield, M., Gale, N. W., Dechiara, T. M., et al. 2004. Loss-of-function analysis of EphA receptors in retinotectal mapping. *J Neurosci*, 24, 2542-50.
- Fesslova, V., Brankovic, J., Lalatta, F., Villa, L., Meli, V., et al. 2011. Recurrence of congenital heart disease in cases with familial risk screened prenatally by echocardiography. *J Pregnancy*, 2011, 368067.
- Firth, H. V., Wright, C. F. & Study, D. D. D. 2011. The Deciphering Developmental Disorders (DDD) study. *Dev Med Child Neurol*, 53, 702-3.

- Fleming, P. J., Blair, P. S. & Pease, A. 2015. Sudden unexpected death in infancy: aetiology, pathophysiology, epidemiology and prevention in 2015. *Arch Dis Child*, 100, 984-8.
- Font De Mora, J., Esteban, L. M., Burks, D. J., Nunez, A., Garces, C., et al. 2003. Ras-GRF1 signaling is required for normal beta-cell development and glucose homeostasis. *EMBO J*, 22, 3039-49.
- Fowkes, M. E. & Mitchell, D. R. 1998. The role of preassembled cytoplasmic complexes in assembly of flagellar dynein subunits. *Mol Biol Cell*, 9, 2337-47.
- Francis, R. J., Christopher, A., Devine, W. A., Ostrowski, L. & Lo, C. 2012. Congenital heart disease and the specification of left-right asymmetry. *Am J Physiol Heart Circ Physiol*, 302, H2102-11.
- Frank, V., Habbig, S., Bartram, M. P., Eisenberger, T., Veenstra-Knol, H. E., et al. 2013. Mutations in NEK8 link multiple organ dysplasia with altered Hippo signalling and increased c-MYC expression. *Hum Mol Genet*, 22, 2177-85.
- Freshour, J., Yokoyama, R. & Mitchell, D. R. 2007. Chlamydomonas flagellar outer row dynein assembly protein ODA7 interacts with both outer row and I1 inner row dyneins. *J Biol Chem*, 282, 5404-12.
- Freson, K., Thys, C., Wittewrongel, C., Vermylen, J., Hoylaerts, M. F., et al. 2003. Molecular cloning and characterization of the GATA1 cofactor human FOG1 and assessment of its binding to GATA1 proteins carrying D218 substitutions. *Hum Genet*, 112, 42-9.
- Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., et al. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol*, 31, 822-6.
- Fujiwara, I., Murakami, Y., Niihori, T., Kanno, J., Hakoda, A., et al. 2015. Mutations in PIGL in a patient with Mabry syndrome. *Am J Med Genet A*, 167A, 777-85.
- Fulcoli, F. G., Franzese, M., Liu, X., Zhang, Z., Angelini, C., et al. 2016. Rebalancing gene haploinsufficiency in vivo by targeting chromatin. *Nat Commun*, 7, 11688.
- Gall, J. C., Jr., Stern, A. M., Cohen, M. M., Adams, M. S. & Davidson, R. T. 1966. Holt-Oram syndrome: clinical and genetic study of a large family. *Am J Hum Genet*, 18, 187-200.
- Gallant, P. 2007. Control of transcription by Pontin and Reptin. *Trends Cell Biol*, 17, 187-92.
- Gana, S., Veggiotti, P., Sciacca, G., Fedeli, C., Bersano, A., et al. 2012. 19q13.11 cryptic deletion: description of two new cases and indication for a role of WTIP haploinsufficiency in hypospadias. *Eur J Hum Genet*, 20, 852-6.
- Garg, V., Kathiriya, I. S., Barnes, R., Schluterman, M. K., King, I. N., et al. 2003. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature*, 424, 443-7.
- Gebbink, M. F., Zondag, G. C., Wubbolts, R. W., Beijersbergen, R. L., Van Etten, I., et al. 1993. Cell-cell adhesion mediated by a receptor-like protein tyrosine phosphatase. *J Biol Chem*, 268, 16101-4.
- Geiersbach, K. B., Gardiner, A. E., Wilson, A., Shetty, S., Bruyere, H., et al. 2014. Subjectivity in chromosome band-level estimation: a multicenter study. *Genet Med*, 16, 170-5.
- Genevieve, D., Amiel, J., Viot, G., Le Merrer, M., Sanlaville, D., et al. 2004. Atypical findings in Kabuki syndrome: report of 8 patients in a series of 20 and review of the literature. *Am J Med Genet A*, 129A, 64-8.
- Ghebranious, N., Giampietro, P. F., Westbrook, F. P. & Rezkalla, S. H. 2007. A novel microdeletion at 16p11.2 harbors candidate genes for aortic valve development, seizure disorder, and mild mental retardation. *Am J Med Genet A*, 143A, 1462-71.
- Gibson, W. T., Hood, R. L., Zhan, S. H., Bulman, D. E., Fejes, A. P., et al. 2012. Mutations in EZH2 cause Weaver syndrome. *Am J Hum Genet*, 90, 110-8.

- Giese, K. P., Friedman, E., Telliez, J. B., Fedorov, N. B., Wines, M., et al. 2001. Hippocampus-dependent learning and memory is impaired in mice lacking the Ras-guanine-nucleotide releasing factor 1 (Ras-GRF1). *Neuropharmacology*, 41, 791-800.
- Glanz, A. & Fraser, F. C. 1982. Spectrum of anomalies in Fanconi anaemia. *J Med Genet*, 19, 412-6.
- Glessner, J. T., Bick, A. G., Ito, K., Homsy, J. G., Rodriguez-Murillo, L., et al. 2014. Increased frequency of de novo copy number variants in congenital heart disease by integrative analysis of single nucleotide polymorphism array and exome sequence data. *Circ Res*, 115, 884-96.
- Glustein, J. Z., Weill, M. & Steinberg, A. 1996. Anomalies of the nipple: an additional finding in CHARGE syndrome. *Am J Med Genet*, 61, 201.
- Goldmuntz, E., Bamford, R., Karkera, J. D., Dela Cruz, J., Roessler, E., et al. 2002. CFC1 mutations in patients with transposition of the great arteries and double-outlet right ventricle. *Am J Hum Genet*, 70, 776-80.
- Gong, C., Bongiorno, P., Martins, A., Stephanou, N. C., Zhu, H., et al. 2005. Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. *Nat Struct Mol Biol*, 12, 304-12.
- Gonzalez-Hilarion, S., Beghyn, T., Jia, J., Debreuck, N., Berte, G., et al. 2012. Rescue of nonsense mutations by amlexanox in human cells. *Orphanet J Rare Dis*, 7, 58.
- Goodship, J. A., Hall, D., Topf, A., Mamasoula, C., Griffin, H., et al. 2012. A common variant in the PTPN11 gene contributes to the risk of tetralogy of Fallot. *Circ Cardiovasc Genet*, 5, 287-92.
- Gopalakrishnan, S., Hor, P. & Ichida, J. K. 2017. New approaches for direct conversion of patient fibroblasts into neural cells. *Brain Res*, 1656, 2-13.
- Grampa, V., Delous, M., Zaidan, M., Ody, G., Thomas, S., et al. 2016. Novel NEK8 Mutations Cause Severe Syndromic Renal Cystic Dysplasia through YAP Dysregulation. *PLoS Genet*, 12, e1005894.
- Granados-Riveron, J. T., Ghosh, T. K., Pope, M., Bu'lock, F., Thornborough, C., et al. 2010. Alpha-cardiac myosin heavy chain (MYH6) mutations affecting myofibril formation are associated with congenital heart defects. *Hum Mol Genet*, 19, 4007-16.
- Granados-Riveron, J. T., Pope, M., Bu'lock, F. A., Thornborough, C., Eason, J., et al. 2012. Combined mutation screening of NKX2-5, GATA4, and TBX5 in congenital heart disease: multiple heterozygosity and novel mutations. *Congenit Heart Dis*, 7, 151-9.
- Granata, A., Serrano, F., Bernard, W. G., Mcnamara, M., Low, L., et al. 2017. An iPSC-derived vascular model of Marfan syndrome identifies key mediators of smooth muscle cell death. *Nat Genet*, 49, 97-109.
- Green, R. C., Berg, J. S., Grody, W. W., Kalia, S. S., Korf, B. R., et al. 2013. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med*, 15, 565-74.
- Greenberg, F. 1987. Choanal atresia and athelia: methimazole teratogenicity or a new syndrome? *Am J Med Genet*, 28, 931-4.
- Greenway, S. C., Mcleod, R., Hume, S., Roslin, N. M., Alvarez, N., et al. 2014. Exome Sequencing Identifies a Novel Variant in ACTC1 Associated With Familial Atrial Septal Defect. *Can J Cardiol*, 30, 181-7.
- Griffiths, E. J. & Halestrap, A. P. 1993. Pyrophosphate metabolism in the perfused heart and isolated heart mitochondria and its role in regulation of mitochondrial function by calcium. *Biochem J*, 290 (Pt 2), 489-95.
- Guest, S. S., Evans, C. D. & Winter, R. M. 1999. The Online London Dysmorphology Database. *Genet Med*, 1, 207-12.

- Guimier, A., Gabriel, G. C., Bajolle, F., Tsang, M., Liu, H., et al. 2015. MMP21 is mutated in human heterotaxy and is required for normal left-right asymmetry in vertebrates. *Nat Genet*, 47, 1260-3.
- Guimier, A., Gordon, C. T., Godard, F., Ravenscroft, G., Oufadem, M., et al. 2016. Biallelic PPA2 Mutations Cause Sudden Unexpected Cardiac Arrest in Infancy. *Am J Hum Genet*, 99, 666-73.
- Guo, C., Chang, C. C., Wortham, M., Chen, L. H., Kernagis, D. N., et al. 2012. Global identification of MLL2-targeted loci reveals MLL2's role in diverse signaling pathways. *Proc Natl Acad Sci U S A*, 109, 17603-8.
- Guo, T., Chung, J. H., Wang, T., McDonald-McGinn, D. M., Kates, W. R., et al. 2015. Histone Modifier Genes Alter Conotruncal Heart Phenotypes in 22q11.2 Deletion Syndrome. *Am J Hum Genet*, 97, 869-77.
- Guyon, J. R., Steffen, L. S., Howell, M. H., Pusack, T. J., Lawrence, C., et al. 2007. Modeling human muscle disease in zebrafish. *Biochim Biophys Acta*, 1772, 205-15.
- Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., Le Deist, F., Wulffraat, N., et al. 2003. A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N Engl J Med*, 348, 255-6.
- Hanchard, N. A., Swaminathan, S., Bucayas, K., Furthner, D., Fernbach, S., et al. 2016. A genome-wide association study of congenital cardiovascular left-sided lesions shows association with a locus on chromosome 20. *Hum Mol Genet*, 25, 2331-2341.
- Hang, C. T., Yang, J., Han, P., Cheng, H. L., Shang, C., et al. 2010. Chromatin regulation by Brg1 underlies heart muscle development and disease. *Nature*, 466, 62-7.
- Hannibal, M. C., Buckingham, K. J., Ng, S. B., Ming, J. E., Beck, A. E., et al. 2011. Spectrum of MLL2 (ALR) mutations in 110 cases of Kabuki syndrome. *Am J Med Genet A*, 155A, 1511-6.
- Harris, P. C. & Rossetti, S. 2004. Molecular genetics of autosomal recessive polycystic kidney disease. *Mol Genet Metab*, 81, 75-85.
- Hehir-Kwa, J. Y., Claustres, M., Hastings, R. J., Van Ravenswaaij-Arts, C., Christenhusz, G., et al. 2015. Towards a European consensus for reporting incidental findings during clinical NGS testing. *Eur J Hum Genet*, 23, 1601-6.
- Heinonen, J. 2013. *Biological Role of Inorganic Pyrophosphate*, Springer Verlag.
- Helgason, C. D., Damen, J. E., Rosten, P., Grewal, R., Sorensen, P., et al. 1998. Targeted disruption of SHIP leads to hemopoietic perturbations, lung pathology, and a shortened life span. *Genes Dev*, 12, 1610-20.
- Herman, M. A., Peroni, O. D., Villoria, J., Schon, M. R., Abumrad, N. A., et al. 2012. A novel ChREBP isoform in adipose tissue regulates systemic glucose metabolism. *Nature*, 484, 333-8.
- Hildebrandt, F., Benzing, T. & Katsanis, N. 2011. Ciliopathies. *N Engl J Med*, 364, 1533-43.
- Hinton, R. B., Jr., Martin, L. J., Tabangin, M. E., Mazwi, M. L., Cripe, L. H., et al. 2007. Hypoplastic left heart syndrome is heritable. *J Am Coll Cardiol*, 50, 1590-5.
- Hirawat, S., Welch, E. M., Elfring, G. L., Northcutt, V. J., Paushkin, S., et al. 2007. Safety, tolerability, and pharmacokinetics of PTC124, a nonaminoglycoside nonsense mutation suppressor, following single- and multiple-dose administration to healthy male and female adult volunteers. *J Clin Pharmacol*, 47, 430-44.
- Hiroi, Y., Kudoh, S., Monzen, K., Ikeda, Y., Yazaki, Y., et al. 2001. Tbx5 associates with Nkx2-5 and synergistically promotes cardiomyocyte differentiation. *Nat Genet*, 28, 276-80.
- Hirokawa, N., Tanaka, Y. & Okada, Y. 2009. Left-right determination: involvement of molecular motor KIF3, cilia, and nodal flow. *Cold Spring Harb Perspect Biol*, 1, a000802.

- Hisama, F. M., Reyes-Mugica, M., Wargowski, D. S., Thompson, K. J. & Mahoney, M. J. 1998. Renal tubular dysgenesis, absent nipples, and multiple malformations in three brothers: a new, lethal syndrome. *Am J Med Genet*, 80, 335-42.
- Hitz, M. P., Lemieux-Perreault, L. P., Marshall, C., Feroz-Zada, Y., Davies, R., et al. 2012. Rare copy number variants contribute to congenital left-sided heart disease. *PLoS Genet*, 8, e1002903.
- Hoffman, J. I. & Kaplan, S. 2002. The incidence of congenital heart disease. *J Am Coll Cardiol*, 39, 1890-900.
- Hogan, M. C., Griffin, M. D., Rossetti, S., Torres, V. E., Ward, C. J., et al. 2003. PKHDL1, a homolog of the autosomal recessive polycystic kidney disease gene, encodes a receptor with inducible T lymphocyte expression. *Hum Mol Genet*, 12, 685-98.
- Holtzman, N. A. 2013. ACMG recommendations on incidental findings are flawed scientifically and ethically. *Genet Med*, 15, 750-1.
- Holzer, R. J., Sallehuddin, A. & Hijazi, Z. M. 2016. Surgical strategies and novel alternatives for the closure of ventricular septal defects. *Expert Rev Cardiovasc Ther*, 14, 831-41.
- Homsy, J., Zaidi, S., Shen, Y., Ware, J. S., Samocha, K. E., et al. 2015. De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science*, 350, 1262-6.
- Horani, A., Brody, S. L. & Ferkol, T. W. 2014. Picking up speed: advances in the genetics of primary ciliary dyskinesia. *Pediatr Res*, 75, 158-64.
- Horani, A., Ferkol, T. W., Dutcher, S. K. & Brody, S. L. 2016. Genetics and biology of primary ciliary dyskinesia. *Paediatr Respir Rev*, 18, 18-24.
- Horn, D., Krawitz, P., Mannhardt, A., Korenke, G. C. & Meinecke, P. 2011. Hyperphosphatasia-mental retardation syndrome due to PIGV mutations: expanded clinical spectrum. *Am J Med Genet A*, 155A, 1917-22.
- Horn, D., Wieczorek, D., Metcalfe, K., Baric, I., Palezac, L., et al. 2014. Delineation of PIGV mutation spectrum and associated phenotypes in hyperphosphatasia with mental retardation syndrome. *Eur J Hum Genet*, 22, 762-7.
- Horvath, G. A. & Armstrong, L. 2007. Report of a fourth individual with a lethal syndrome of choanal atresia, athelia, evidence of renal tubulopathy, and family history of neck cysts. *Am J Med Genet A*, 143A, 1231-5.
- Howard, M. F., Murakami, Y., Pagnamenta, A. T., Daumer-Haas, C., Fischer, B., et al. 2014. Mutations in PGAP3 impair GPI-anchor maturation, causing a subtype of hyperphosphatasia with mental retardation. *Am J Hum Genet*, 94, 278-87.
- Hu, Z., Shi, Y., Mo, X., Xu, J., Zhao, B., et al. 2013. A genome-wide association study identifies two risk loci for congenital heart malformations in Han Chinese populations. *Nat Genet*, 45, 818-21.
- Iascone, M., Ciccone, R., Galletti, L., Marchetti, D., Seddio, F., et al. 2012. Identification of de novo mutations and rare variants in hypoplastic left heart syndrome. *Clin Genet*, 81, 542-54.
- Ikura, T., Ogryzko, V. V., Grigoriev, M., Groisman, R., Wang, J., et al. 2000. Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. *Cell*, 102, 463-73.
- Ilkovski, B., Pagnamenta, A. T., O'grady, G. L., Kinoshita, T., Howard, M. F., et al. 2015. Mutations in PIGY: expanding the phenotype of inherited glycosylphosphatidylinositol deficiencies. *Hum Mol Genet*, 24, 6146-59.
- Inaba, Y., Shinohara, K., Botilde, Y., Nabeshima, R., Takaoka, K., et al. 2016. Transport of the outer dynein arm complex to cilia requires a cytoplasmic protein Lrrc6. *Genes Cells*, 21, 728-39.
- Ipe, J., Swart, M., Burgess, K. S. & Skaar, T. C. 2017. High-Throughput Assays to Assess the Functional Impact of Genetic Variants: A Road Towards Genomic-Driven Medicine. *Clin Transl Sci*, 10, 67-77.

- Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. 1987. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol*, 169, 5429-33.
- Issaeva, I., Zonis, Y., Rozovskaia, T., Orlovsky, K., Croce, C. M., et al. 2007. Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. *Mol Cell Biol*, 27, 1889-903.
- Jenkins, K. J., Correa, A., Feinstein, J. A., Botto, L., Britt, A. E., et al. 2007. Noninherited risk factors and congenital cardiovascular defects: current knowledge: a scientific statement from the American Heart Association Council on Cardiovascular Disease in the Young: endorsed by the American Academy of Pediatrics. *Circulation*, 115, 2995-3014.
- Jerome, L. A. & Papaioannou, V. E. 2001. DiGeorge syndrome phenotype in mice mutant for the T-box gene, *Tbx1*. *Nat Genet*, 27, 286-91.
- Jha, S. & Dutta, A. 2009. RVB1/RVB2: running rings around molecular biology. *Mol Cell*, 34, 521-33.
- Jia, Y., Louw, J. J., Breckpot, J., Callewaert, B., Barrea, C., et al. 2015. The diagnostic value of next generation sequencing in familial nonsyndromic congenital heart defects. *Am J Med Genet A*, 167A, 1822-9.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., et al. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337, 816-21.
- Jones, C. A., London, N. R., Chen, H., Park, K. W., Sauvaget, D., et al. 2008. Robo4 stabilizes the vascular network by inhibiting pathologic angiogenesis and endothelial hyperpermeability. *Nat Med*, 14, 448-53.
- Jones, J. M. & Koch, W. J. 2005. Gene therapy approaches to cardiovascular disease. *Methods Mol Med*, 112, 15-35.
- Jonsson, Z. O., Dhar, S. K., Narlikar, G. J., Auty, R., Wagle, N., et al. 2001. Rvb1p and Rvb2p are essential components of a chromatin remodeling complex that regulates transcription of over 5% of yeast genes. *J Biol Chem*, 276, 16279-88.
- Jose, T., Gdynia, H. J., Mahrholdt, H., Vohringer, M., Klingel, K., et al. 2011. CMR gives clue to "ragged red fibers" in the heart in a patient with mitochondrial myopathy. *Int J Cardiol*, 149, e24-7.
- Kaab, S. & Schulze-Bahr, E. 2005. Susceptibility genes and modifiers for cardiac arrhythmias. *Cardiovasc Res*, 67, 397-413.
- Karijolich, J. & Yu, Y. T. 2011. Converting nonsense codons into sense codons by targeted pseudouridylation. *Nature*, 474, 395-8.
- Karp, N., Grosse-Wortmann, L. & Bowdin, S. 2012. Severe aortic stenosis, bicuspid aortic valve and atrial septal defect in a child with Joubert Syndrome and Related Disorders (JSRD) - a case report and review of congenital heart defects reported in the human ciliopathies. *Eur J Med Genet*, 55, 605-10.
- Katsanis, S. H. & Katsanis, N. 2013. Molecular genetic testing and the future of clinical genomics. *Nat Rev Genet*, 14, 415-26.
- Katsuhara, K., Kawamoto, S., Wakabayashi, T. & Belsky, J. L. 1972. Situs inversus totalis and Kartagener's syndrome in a Japanese population. *Chest*, 61, 56-61.
- Keeling, K. M., Xue, X., Gunn, G. & Bedwell, D. M. 2014. Therapeutics based on stop codon readthrough. *Annu Rev Genomics Hum Genet*, 15, 371-94.
- Keller, P. J., Schmidt, A. D., Santella, A., Khairy, K., Bao, Z., et al. 2010. Fast, high-contrast imaging of animal development with scanned light sheet-based structured-illumination microscopy. *Nat Meth*, 7, 637-642.
- Kennedy, H., Haack, T. B., Hartill, V., Matakovic, L., Baumgartner, E. R., et al. 2016. Sudden Cardiac Death Due to Deficiency of the Mitochondrial Inorganic Pyrophosphatase PPA2. *Am J Hum Genet*, 99, 674-82.

- Kennedy, M. P., Omran, H., Leigh, M. W., Dell, S., Morgan, L., et al. 2007. Congenital heart disease and other heterotaxic defects in a large cohort of patients with primary ciliary dyskinesia. *Circulation*, 115, 2814-21.
- Kerem, E., Konstan, M. W., De Boeck, K., Accurso, F. J., Sermet-Gaudelus, I., et al. 2014. Ataluren for the treatment of nonsense-mutation cystic fibrosis: a randomised, double-blind, placebo-controlled phase 3 trial. *Lancet Respir Med*, 2, 539-47.
- Kerstjens-Frederikse, W. S., Van De Laar, I. M., Vos, Y. J., Verhagen, J. M., Berger, R. M., et al. 2016. Cardiovascular malformations caused by NOTCH1 mutations do not keep left: data on 428 probands with left-sided CHD and their families. *Genet Med*, 18, 914-23.
- Kim, T. G., Chen, J., Sadoshima, J. & Lee, Y. 2004. Jumonji represses atrial natriuretic factor gene expression by inhibiting transcriptional activities of cardiac transcription factors. *Mol Cell Biol*, 24, 10151-60.
- Kircher, M., Witten, D. M., Jain, P., O'roak, B. J., Cooper, G. M., et al. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*, 46, 310-5.
- Kirshbom, P. M., Mahle, W. T., Joyner, R. W., Leong, T., Wilson, M., et al. 2008. The endothelin-1 G5665T polymorphism impacts transplant-free survival for single ventricle patients. *J Thorac Cardiovasc Surg*, 136, 117-22.
- Klattenhoff, C. A., Scheuermann, J. C., Surface, L. E., Bradley, R. K., Fields, P. A., et al. 2013. Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell*, 152, 570-83.
- Klaver, E. C., Versluijs, G. M. & Wilders, R. 2011. Cardiac ion channel mutations in the sudden infant death syndrome. *Int J Cardiol*, 152, 162-70.
- Knoppers, B. M., Zawati, M. H. & Senecal, K. 2015. Return of genetic testing results in the era of whole-genome sequencing. *Nat Rev Genet*, 16, 553-9.
- Kodo, K., Nishizawa, T., Furutani, M., Arai, S., Yamamura, E., et al. 2009. GATA6 mutations cause human cardiac outflow tract defects by disrupting semaphorin-plexin signaling. *Proc Natl Acad Sci U S A*, 106, 13933-8.
- Koike-Yusa, H., Li, Y., Tan, E. P., Velasco-Herrera Mdel, C. & Yusa, K. 2014. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol*, 32, 267-73.
- Koop, E. A., Gebbink, M. F., Sweeney, T. E., Mathy, M. J., Heijnen, H. F., et al. 2005. Impaired flow-induced dilation in mesenteric resistance arteries from receptor protein tyrosine phosphatase-mu-deficient mice. *Am J Physiol Heart Circ Physiol*, 288, H1218-23.
- Krawitz, P. M., Murakami, Y., Hecht, J., Kruger, U., Holder, S. E., et al. 2012. Mutations in PIGO, a member of the GPI-anchor-synthesis pathway, cause hyperphosphatasia with mental retardation. *Am J Hum Genet*, 91, 146-51.
- Krawitz, P. M., Murakami, Y., Riess, A., Hietala, M., Kruger, U., et al. 2013. PGAP2 mutations, affecting the GPI-anchor-synthesis pathway, cause hyperphosphatasia with mental retardation syndrome. *Am J Hum Genet*, 92, 584-9.
- Krawitz, P. M., Schweiger, M. R., Rodelsperger, C., Marcelis, C., Kolsch, U., et al. 2010. Identity-by-descent filtering of exome sequence data identifies PIGV mutations in hyperphosphatasia mental retardation syndrome. *Nat Genet*, 42, 827-9.
- Kumar, P., Sait, S. F., Sharma, A. & Kumar, M. 2011. Ovarian hyperstimulation syndrome. *J Hum Reprod Sci*, 4, 70-5.
- Kurzrock, R., Kantarjian, H. M., Druker, B. J. & Talpaz, M. 2003. Philadelphia chromosome-positive leukemias: from basic mechanisms to molecular therapeutics. *Ann Intern Med*, 138, 819-30.

- Lahaye, S., Corsmeier, D., Basu, M., Bowman, J. L., Fitzgerald-Butt, S., et al. 2016. Utilization of Whole Exome Sequencing to Identify Causative Mutations in Familial Congenital Heart Disease. *Circ Cardiovasc Genet*, 9, 320-9.
- Laity, J. H., Lee, B. M. & Wright, P. E. 2001. Zinc finger proteins: new insights into structural and functional diversity. *Curr Opin Struct Biol*, 11, 39-46.
- Lakdawala, N. K. 2013. Using genetic testing to guide therapeutic decisions in cardiomyopathy. *Curr Treat Options Cardiovasc Med*, 15, 387-96.
- Lakshminarasimhan, M., Boanca, G., Banks, C. A., Hattem, G. L., Gabriel, A. E., et al. 2016. Proteomic and Genomic Analyses of the Rvb1 and Rvb2 Interaction Network upon Deletion of R2TP Complex Components. *Mol Cell Proteomics*, 15, 960-74.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., et al. 2001. Initial sequencing and analysis of the human genome. *Nature*, 409, 860-921.
- Landis, B. J. & Ware, S. M. 2016. The Current Landscape of Genetic Testing in Cardiovascular Malformations: Opportunities and Challenges. *Front Cardiovasc Med*, 3, 22.
- Lanphier, E., Urnov, F., Haecker, S. E., Werner, M. & Smolenski, J. 2015. Don't edit the human germ line. *Nature*, 519, 410-1.
- Lee, Y., Song, A. J., Baker, R., Micales, B., Conway, S. J., et al. 2000. Jumonji, a nuclear protein that is necessary for normal heart development. *Circ Res*, 86, 932-8.
- Leonardi, M. L., Pai, G. S., Wilkes, B. & Lebel, R. R. 2001. Ritscher-Schinzel cranio-cerebello-cardiac (3C) syndrome: report of four new cases and review. *Am J Med Genet*, 102, 237-42.
- Li, H. & Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25, 1754-60.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25, 2078-9.
- Li, J., Cao, Y., Ma, X. J., Wang, H. J., Zhang, J., et al. 2013. Roles of miR-1-1 and miR-181c in ventricular septal defects. *Int J Cardiol*, 168, 1441-6.
- Li, Y., Klena, N. T., Gabriel, G. C., Liu, X., Kim, A. J., et al. 2015. Global genetic analysis in mice unveils central role for cilia in congenital heart disease. *Nature*, 521, 520-4.
- Lian, X., Xu, J., Li, J. & Chien, K. R. 2014. Next-generation models of human cardiogenesis via genome editing. *Cold Spring Harb Perspect Med*, 4, a013920.
- Lickert, H., Takeuchi, J. K., Von Both, I., Walls, J. R., Mcauliffe, F., et al. 2004. Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. *Nature*, 432, 107-12.
- Lien, C. L., Wu, C., Mercer, B., Webb, R., Richardson, J. A., et al. 1999. Control of early cardiac-specific transcription of Nkx2-5 by a GATA-dependent enhancer. *Development*, 126, 75-84.
- Lin, C. J., Lin, C. Y., Chen, C. H., Zhou, B. & Chang, C. P. 2012. Partitioning the heart: mechanisms of cardiac septation and valve development. *Development*, 139, 3277-99.
- Lintas, C. & Persico, A. M. 2017. Unraveling molecular pathways shared by Kabuki and Kabuki-like syndromes. *Clin Genet*.
- Liu, N., Bezprozvannaya, S., Williams, A. H., Qi, X., Richardson, J. A., et al. 2008. microRNA-133a regulates cardiomyocyte proliferation and suppresses smooth muscle gene expression in the heart. *Genes Dev*, 22, 3242-54.
- Liu, Q., Shalaby, F., Jones, J., Bouchard, D. & Dumont, D. J. 1998. The SH2-containing inositol polyphosphate 5-phosphatase, ship, is expressed during hematopoiesis and spermatogenesis. *Blood*, 91, 2753-9.

- Liu, S., Lu, W., Obara, T., Kuida, S., Lehoczky, J., et al. 2002. A defect in a novel Nek-family kinase causes cystic kidney disease in the mouse and in zebrafish. *Development*, 129, 5839-46.
- Liu, X. S., Wu, H., Ji, X., Stelzer, Y., Wu, X., et al. 2016. Editing DNA Methylation in the Mammalian Genome. *Cell*, 167, 233-247 e17.
- Lizen, B., Hutlet, B., Bissen, D., Sauvegarde, D., Hermant, M., et al. 2017. HOXA5 localization in postnatal and adult mouse brain is suggestive of regulatory roles in postmitotic neurons. *J Comp Neurol*, 525, 1155-1175.
- Loges, N. T., Olbrich, H., Becker-Heck, A., Haffner, K., Heer, A., et al. 2009. Deletions and point mutations of LRRC50 cause primary ciliary dyskinesia due to dynein arm defects. *Am J Hum Genet*, 85, 883-9.
- Lough, J. W. 2002. Transient expression of TIP60 protein during early chick heart development. *Dev Dyn*, 223, 419-25.
- Lundin, M., Baltscheffsky, H. & Ronne, H. 1991. Yeast PPA2 gene encodes a mitochondrial inorganic pyrophosphatase that is essential for mitochondrial function. *J Biol Chem*, 266, 12168-72.
- Macarthur, D. G., Manolio, T. A., Dimmock, D. P., Rehm, H. L., Shendure, J., et al. 2014. Guidelines for investigating causality of sequence variants in human disease. *Nature*, 508, 469-76.
- Mainwaring, R. D., Lamberti, J. J. & Uzark, K. 1994. The bidirectional Glenn procedure: palliation of the univentricular heart. *Adv Card Surg*, 5, 115-40.
- Malicki, J. J. & Johnson, C. A. 2017. The Cilium: Cellular Antenna and Central Processing Unit. *Trends Cell Biol*, 27, 126-140.
- Mather, C. A., Mooney, S. D., Salipante, S. J., Scroggins, S., Wu, D., et al. 2016. CADD score has limited clinical validity for the identification of pathogenic variants in noncoding regions in a hereditary cancer panel. *Genet Med*, 18, 1269-1275.
- Matsson, H., Eason, J., Bookwalter, C. S., Klar, J., Gustavsson, P., et al. 2008. Alpha-cardiac actin mutations produce atrial septal defects. *Hum Mol Genet*, 17, 256-65.
- Matsumoto, N. & Niikawa, N. 2003. Kabuki make-up syndrome: a review. *Am J Med Genet C Semin Med Genet*, 117C, 57-65.
- Mausner, W. S. & Bonnemeier, H. 2012. Cardiomyopathy and sudden cardiac death in Williams-Beuren-syndrome. *Int J Cardiol*, 156, e53-4.
- Mayr, J. A., Havlickova, V., Zimmermann, F., Magler, I., Kaplanova, V., et al. 2010. Mitochondrial ATP synthase deficiency due to a mutation in the ATP5E gene for the F1 epsilon subunit. *Hum Mol Genet*, 19, 3430-9.
- Mcbride, K. L., Pignatelli, R., Lewin, M., Ho, T., Fernbach, S., et al. 2005. Inheritance analysis of congenital left ventricular outflow tract obstruction malformations: Segregation, multiplex relative risk, and heritability. *Am J Med Genet A*, 134A, 180-6.
- Mcbride, K. L., Riley, M. F., Zender, G. A., Fitzgerald-Butt, S. M., Towbin, J. A., et al. 2008. NOTCH1 mutations in individuals with left ventricular outflow tract malformations reduce ligand-induced signaling. *Hum Mol Genet*, 17, 2886-93.
- Mccarroll, S. A. & Altshuler, D. M. 2007. Copy-number variation and association studies of human disease. *Nat Genet*, 39, S37-42.
- Mcculley, D. J. & Black, B. L. 2012. Transcription factor pathways and congenital heart disease. *Curr Top Dev Biol*, 100, 253-77.
- Mcinerney-Leo, A. M., Harris, J. E., Gattas, M., Peach, E. E., Sinnott, S., et al. 2016. Fryns Syndrome Associated with Recessive Mutations in PIGN in two Separate Families. *Hum Mutat*, 37, 695-702.
- Mcintyre, J. C., Davis, E. E., Joiner, A., Williams, C. L., Tsai, I. C., et al. 2012. Gene therapy rescues cilia defects and restores olfactory function in a mammalian ciliopathy model. *Nat Med*, 18, 1423-8.

- Mckellar, S. H., Tester, D. J., Yagubyan, M., Majumdar, R., Ackerman, M. J., et al. 2007. Novel NOTCH1 mutations in patients with bicuspid aortic valve disease and thoracic aortic aneurysms. *J Thorac Cardiovasc Surg*, 134, 290-6.
- Mclaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P., et al. 2010. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*, 26, 2069-70.
- Merla, G., Howald, C., Antonarakis, S. E. & Reymond, A. 2004. The subcellular localization of the ChoRE-binding protein, encoded by the Williams-Beuren syndrome critical region gene 14, is regulated by 14-3-3. *Hum Mol Genet*, 13, 1505-14.
- Ming, J. E., Russell, K. L., Bason, L., Mcdonald-Mcginn, D. M. & Zackai, E. H. 2003. Coloboma and other ophthalmologic anomalies in Kabuki syndrome: distinction from charge association. *Am J Med Genet A*, 123A, 249-52.
- Ming, L., Wilk, R., Reed, B. H. & Lipshitz, H. D. 2013. Drosophila Hindsight and mammalian RREB-1 are evolutionarily conserved DNA-binding transcriptional attenuators. *Differentiation*, 86, 159-70.
- Mitchell, A. A., Zwick, M. E., Chakravarti, A. & Cutler, D. J. 2004. Discrepancies in dbSNP confirmation rates and allele frequency distributions from varying genotyping error rates and patterns. *Bioinformatics*, 20, 1022-32.
- Mitchell, M. E., Sander, T. L., Klinkner, D. B. & Tomita-Mitchell, A. 2007. The molecular basis of congenital heart disease. *Semin Thorac Cardiovasc Surg*, 19, 228-37.
- Mitchison, H. M., Schmidts, M., Loges, N. T., Freshour, J., Dritsoula, A., et al. 2012. Mutations in axonemal dynein assembly factor DNAAF3 cause primary ciliary dyskinesia. *Nat Genet*, 44, 381-9, S1-2.
- Miyague, N. I., Cardoso, S. M., Meyer, F., Ultramari, F. T., Araujo, F. H., et al. 2003. Epidemiological study of congenital heart defects in children and adolescents. Analysis of 4,538 cases. *Arq Bras Cardiol*, 80, 269-78.
- Mo, R., Rao, S. M. & Zhu, Y. J. 2006. Identification of the MLL2 complex as a coactivator for estrogen receptor alpha. *J Biol Chem*, 281, 15714-20.
- Moller, J. H. & Anderson, R. C. 2013. A 43- to 54-year follow-up of 1,000 patients with congenital heart disease. *Am J Cardiol*, 111, 1496-500.
- Mommersteeg, M. T., Andrews, W. D., Ypsilanti, A. R., Zelina, P., Yeh, M. L., et al. 2013. Slit-roundabout signaling regulates the development of the cardiac systemic venous return and pericardium. *Circ Res*, 112, 465-75.
- Moniz, L., Dutt, P., Haider, N. & Stambolic, V. 2011. Nek family of kinases in cell cycle, checkpoint control and cancer. *Cell Div*, 6, 18.
- Morin, R. D., Mendez-Lago, M., Mungall, A. J., Goya, R., Mungall, K. L., et al. 2011. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature*, 476, 298-303.
- Morton, S. U., Scherz, P. J., Cordes, K. R., Ivey, K. N., Stainier, D. Y., et al. 2008. microRNA-138 modulates cardiac patterning during embryonic development. *Proc Natl Acad Sci U S A*, 105, 17830-5.
- Moser, E., Kargl, J., Whistler, J. L., Waldhoer, M. & Tschische, P. 2010. G protein-coupled receptor-associated sorting protein 1 regulates the postendocytic sorting of seven-transmembrane-spanning G protein-coupled receptors. *Pharmacology*, 86, 22-9.
- Moskowitz, I. P., Wang, J., Peterson, M. A., Pu, W. T., Mackinnon, A. C., et al. 2011. Transcription factor genes Smad4 and Gata4 cooperatively regulate cardiac valve development. [corrected]. *Proc Natl Acad Sci U S A*, 108, 4006-11.
- Moustafa, G. A., Kolokythas, A., Charitakis, K. & Avgerinos, D. V. 2016. Therapeutic Utilities of Pediatric Cardiac Catheterization. *Curr Cardiol Rev*, 12, 258-269.
- Murakami, Y., Kanzawa, N., Saito, K., Krawitz, P. M., Mundlos, S., et al. 2012. Mechanism for release of alkaline phosphatase caused by glycosylphosphatidylinositol deficiency in patients with hyperphosphatasia mental retardation syndrome. *J Biol Chem*, 287, 6318-25.

- Mussolino, C., Morbitzer, R., Lutge, F., Dannemann, N., Lahaye, T., et al. 2011. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res*, 39, 9283-93.
- Nakamura, T., Saito, D., Kawasumi, A., Shinohara, K., Asai, Y., et al. 2012. Fluid flow and interlinked feedback loops establish left-right asymmetric decay of *Cer12* mRNA. *Nat Commun*, 3, 1322.
- Nakhleh, N., Francis, R., Giese, R. A., Tian, X., Li, Y., et al. 2012. High prevalence of respiratory ciliary dysfunction in congenital heart disease patients with heterotaxy. *Circulation*, 125, 2232-42.
- Naldini, L. 2015. Gene therapy returns to centre stage. *Nature*, 526, 351-60.
- Nance, M. E., Hakim, C. H., Yang, N. N. & Duan, D. 2017. Nanotherapy for Duchenne muscular dystrophy. *Wiley Interdiscip Rev Nanomed Nanobiotechnol*.
- Ng, P. C. & Henikoff, S. 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*, 31, 3812-4.
- Ng, S. B., Bigham, A. W., Buckingham, K. J., Hannibal, M. C., Mcmillin, M. J., et al. 2010. Exome sequencing identifies *MLL2* mutations as a cause of Kabuki syndrome. *Nat Genet*, 42, 790-3.
- Ni, W., Liu, T., Wang, H. Y., Liu, L. H. & Chen, G. X. 2016. [Expression of Slit3/Robo signal pathway in mouse aortic smooth muscle cell and its impact on proliferation and migration]. *Zhonghua Xin Xue Guan Bing Za Zhi*, 44, 542-7.
- Nicastro, D., Schwartz, C., Pierson, J., Gaudette, R., Porter, M. E., et al. 2006. The molecular architecture of axonemes revealed by cryoelectron tomography. *Science*, 313, 944-8.
- Niessen, K. & Karsan, A. 2008. Notch signaling in cardiac development. *Circ Res*, 102, 1169-81.
- Nistri, S., Basso, C., Marzari, C., Mormino, P. & Thiene, G. 2005. Frequency of bicuspid aortic valve in young male conscripts by echocardiogram. *Am J Cardiol*, 96, 718-21.
- Nolasco, S., Bellido, J., Goncalves, J., Zabala, J. C. & Soares, H. 2005. Tubulin cofactor A gene silencing in mammalian cells induces changes in microtubule cytoskeleton, cell cycle arrest and cell death. *FEBS Lett*, 579, 3515-24.
- Nonaka, S., Shiratori, H., Saijoh, Y. & Hamada, H. 2002. Determination of left-right patterning of the mouse embryo by artificial nodal flow. *Nature*, 418, 96-9.
- Norris, D. P. 2012. Cilia, calcium and the basis of left-right asymmetry. *BMC Biol*, 10, 102.
- O'Brien, J. E., Jr., Kibiryeva, N., Zhou, X. G., Marshall, J. A., Lofland, G. K., et al. 2012. Noncoding RNA expression in myocardium from infants with tetralogy of Fallot. *Circ Cardiovasc Genet*, 5, 279-86.
- Omran, H., Kobayashi, D., Olbrich, H., Tsukahara, T., Loges, N. T., et al. 2008. *Ktu/PF13* is required for cytoplasmic pre-assembly of axonemal dyneins. *Nature*, 456, 611-6.
- Ono, K., Kuwabara, Y. & Han, J. 2011. MicroRNAs and cardiovascular diseases. *FEBS J*, 278, 1619-33.
- Oshima, Y., Yamamoto, T., Ishikawa, T., Mishima, H., Matsusue, A., et al. 2017. Postmortem genetic analysis of sudden unexpected death in infancy: neonatal genetic screening may enable the prevention of sudden infant death. *J Hum Genet*.
- Otsuka, F., Sakakura, K., Yahagi, K., Joner, M. & Virmani, R. 2014. Has our understanding of calcification in human coronary atherosclerosis progressed? *Arterioscler Thromb Vasc Biol*, 34, 724-36.
- Ow, Y. P., Green, D. R., Hao, Z. & Mak, T. W. 2008. Cytochrome c: functions beyond respiration. *Nat Rev Mol Cell Biol*, 9, 532-42.
- Paci, A., Liu, X. H., Huang, H., Lim, A., Houry, W. A., et al. 2012. The stability of the small nucleolar ribonucleoprotein (snoRNP) assembly protein Pih1 in

- Saccharomyces cerevisiae* is modulated by its C terminus. *J Biol Chem*, 287, 43205-14.
- Paff, T., Loges, N. T., Aprea, I., Wu, K., Bakey, Z., et al. 2017. Mutations in PIH1D3 Cause X-Linked Primary Ciliary Dyskinesia with Outer and Inner Dynein Arm Defects. *Am J Hum Genet*, 100, 160-168.
- Paige, S. L., Thomas, S., Stoick-Cooper, C. L., Wang, H., Maves, L., et al. 2012. A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. *Cell*, 151, 221-32.
- Parry, D. A., Logan, C. V., Stegmann, A. P., Abdelhamed, Z. A., Calder, A., et al. 2013. SAMS, a syndrome of short stature, auditory-canal atresia, mandibular hypoplasia, and skeletal abnormalities is a unique neurocristopathy caused by mutations in Goosecoid. *Am J Hum Genet*, 93, 1135-42.
- Pashmforoush, M., Lu, J. T., Chen, H., Amand, T. S., Kondo, R., et al. 2004. Nkx2-5 pathways and congenital heart disease; loss of ventricular myocyte lineage specification leads to progressive cardiomyopathy and complete heart block. *Cell*, 117, 373-86.
- Patel, N. & Alkuraya, F. S. 2015. Overlap between CHARGE and Kabuki syndromes: more than an interesting clinical observation? *Am J Med Genet A*, 167A, 259-60.
- Patel, S., Jones, R. B., Nixon, D. F. & Bollard, C. M. 2016. T-cell therapies for HIV: Preclinical successes and current clinical strategies. *Cytotherapy*, 18, 931-42.
- Pazour, G. J., Baker, S. A., Deane, J. A., Cole, D. G., Dickert, B. L., et al. 2002. The intraflagellar transport protein, IFT88, is essential for vertebrate photoreceptor assembly and maintenance. *J Cell Biol*, 157, 103-13.
- Pediatric Cardiac Genomics Consortium, Gelb, B., Brueckner, M., Chung, W., Goldmuntz, E., et al. 2013. The Congenital Heart Disease Genetic Network Study: rationale, design, and early results. *Circ Res*, 112, 698-706.
- Pehlivan, T., Pober, B. R., Brueckner, M., Garrett, S., Slaugh, R., et al. 1999. GATA4 haploinsufficiency in patients with interstitial deletion of chromosome region 8p23.1 and congenital heart disease. *Am J Med Genet*, 83, 201-6.
- Peiffer, D. A., Le, J. M., Steemers, F. J., Chang, W., Jenniges, T., et al. 2006. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res*, 16, 1136-48.
- Perles, Z., Moon, S., Ta-Shma, A., Yaacov, B., Francescato, L., et al. 2015. A human laterality disorder caused by a homozygous deleterious mutation in MMP21. *J Med Genet*, 52, 840-7.
- Peterson, K. M., Franchi, F., Loeffler, D. L., Psaltis, P. J., Harris, P. C., et al. 2013. Endothelial dysfunction occurs prior to clinical evidence of polycystic kidney disease. *Am J Nephrol*, 38, 233-40.
- Pfendner, E. & Uitto, J. 2005. Plectin gene mutations can cause epidermolysis bullosa with pyloric atresia. *J Invest Dermatol*, 124, 111-5.
- Pierpont, M. E., Basson, C. T., Benson, D. W., Jr., Gelb, B. D., Giglia, T. M., et al. 2007. Genetic basis for congenital heart defects: current knowledge: a scientific statement from the American Heart Association Congenital Cardiac Defects Committee, Council on Cardiovascular Disease in the Young: endorsed by the American Academy of Pediatrics. *Circulation*, 115, 3015-38.
- Pikkarainen, S., Tokola, H., Kerkela, R. & Ruskoaho, H. 2004. GATA transcription factors in the developing and adult heart. *Cardiovasc Res*, 63, 196-207.
- Plagnol, V., Curtis, J., Epstein, M., Mok, K. Y., Stebbings, E., et al. 2012. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics*, 28, 2747-54.
- Postma, A. V., Bezzina, C. R. & Christoffels, V. M. 2016. Genetics of congenital heart disease: the contribution of the noncoding regulatory genome. *J Hum Genet*, 61, 13-9.

- Postma, A. V., Van Engelen, K., Van De Meerakker, J., Rahman, T., Probst, S., et al. 2011. Mutations in the sarcomere gene MYH7 in Ebstein anomaly. *Circ Cardiovasc Genet*, 4, 43-50.
- Prus, W. & Filipek, A. 2011. S100A6 mediates nuclear translocation of Sgt1: a heat shock-regulated protein. *Amino Acids*, 41, 781-7.
- Pulkkinen, L., Smith, F. J., Shimizu, H., Murata, S., Yaoita, H., et al. 1996. Homozygous deletion mutations in the plectin gene (PLEC1) in patients with epidermolysis bullosa simplex associated with late-onset muscular dystrophy. *Hum Mol Genet*, 5, 1539-46.
- Pytela, R. & Wiche, G. 1980. High molecular weight polypeptides (270,000-340,000) from cultured cells are related to hog brain microtubule-associated proteins but copurify with intermediate filaments. *Proc Natl Acad Sci U S A*, 77, 4808-12.
- Qazi, Q. H., Kanchanapoomi, R., Beller, E. & Collins, R. 1982. Inheritance of posterior choanal atresia. *Am J Med Genet*, 13, 413-6.
- Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., et al. 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 152, 1173-83.
- Rajagopalan, R., Grochowski, C. M., Gilbert, M. A., Falsey, A. M., Coleman, K., et al. 2016. Compound heterozygous mutations in NEK8 in siblings with end-stage renal disease with hepatic and cardiac anomalies. *Am J Med Genet A*, 170, 750-3.
- Ramsdell, A. F. 2005. Left-right asymmetry and congenital cardiac defects: getting to the heart of the matter in vertebrate left-right axis determination. *Dev Biol*, 288, 1-20.
- Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., et al. 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, 154, 1380-9.
- Randall, V., Mccue, K., Roberts, C., Kyriakopoulou, V., Beddow, S., et al. 2009. Great vessel development requires biallelic expression of Chd7 and Tbx1 in pharyngeal ectoderm in mice. *J Clin Invest*, 119, 3301-10.
- Reamon-Buettner, S. M. & Borlak, J. 2004. TBX5 mutations in non-Holt-Oram syndrome (HOS) malformed hearts. *Hum Mutat*, 24, 104.
- Reller, M. D., Strickland, M. J., Riehle-Colarusso, T., Mahle, W. T. & Correa, A. 2008. Prevalence of congenital heart defects in metropolitan Atlanta, 1998-2005. *J Pediatr*, 153, 807-13.
- Requena, T., Cabrera, S., Martin-Sierra, C., Price, S. D., Lysakowski, A., et al. 2015. Identification of two novel mutations in FAM136A and DTNA genes in autosomal-dominant familial Meniere's disease. *Hum Mol Genet*, 24, 1119-26.
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., et al. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*, 17, 405-24.
- Richter-Cook, N. J., Dever, T. E., Hensold, J. O. & Merrick, W. C. 1998. Purification and characterization of a new eukaryotic protein translation factor. Eukaryotic initiation factor 4H. *J Biol Chem*, 273, 7579-87.
- Rivera-Feliciano, J., Lee, K. H., Kong, S. W., Rajagopal, S., Ma, Q., et al. 2006. Development of heart valves requires Gata4 expression in endothelial-derived cells. *Development*, 133, 3607-18.
- Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., et al. 2011. Integrative genomics viewer. *Nat Biotech*, 29, 24-26.
- Rodriguez-Balada, M., Roig, B., Martorell, L., Mele, M., Salvat, M., et al. 2016. In silico, in vitro and case-control analyses as an effective combination for analyzing BRCA1 and BRCA2 unclassified variants in a population-based sample. *Cancer Genet*, 209, 487-492.

- Rosenbaum, J., Baek, S. H., Dutta, A., Houry, W. A., Huber, O., et al. 2013. The emergence of the conserved AAA+ ATPases Pontin and Reptin on the signaling landscape. *Sci Signal*, 6, mr1.
- Rosenberg, S. A., Aebersold, P., Cornetta, K., Kasid, A., Morgan, R. A., et al. 1990. Gene transfer into humans--immunotherapy of patients with advanced melanoma, using tumor-infiltrating lymphocytes modified by retroviral gene transduction. *N Engl J Med*, 323, 570-8.
- Rottbauer, W., Saurin, A. J., Lickert, H., Shen, X., Burns, C. G., et al. 2002. Reptin and pontin antagonistically regulate heart growth in zebrafish embryos. *Cell*, 111, 661-72.
- Rual, J. F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., et al. 2005. Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, 437, 1173-8.
- Ruthenburg, A. J., Allis, C. D. & Wysocka, J. 2007. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol Cell*, 25, 15-30.
- Sadiq, M., Stumper, O., Wright, J. G., De Giovanni, J. V., Billingham, C., et al. 1995. Influence of ethnic origin on the pattern of congenital heart defects in the first year of life. *Br Heart J*, 73, 173-6.
- Sajjadi, H. & Paparella, M. M. 2008. Meniere's disease. *Lancet*, 372, 406-14.
- Salmena, L., Poliseno, L., Tay, Y., Kats, L. & Pandolfi, P. P. 2011. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*, 146, 353-8.
- Samiy, N. 2014. Gene therapy for retinal diseases. *J Ophthalmic Vis Res*, 9, 506-9.
- Sansbury, B. E., Jones, S. P., Riggs, D. W., Darley-Usmar, V. M. & Hill, B. G. 2011. Bioenergetic function in cardiovascular cells: the importance of the reserve capacity and its biological regulation. *Chem Biol Interact*, 191, 288-95.
- Satir, P. & Christensen, S. T. 2007. Overview of structure and function of mammalian cilia. *Annu Rev Physiol*, 69, 377-400.
- Schaapveld, R. Q., Schepens, J. T., Robinson, G. W., Attema, J., Oerlemans, F. T., et al. 1997. Impaired mammary gland development and function in mice lacking LAR receptor-like tyrosine phosphatase activity. *Dev Biol*, 188, 134-46.
- Schaefer, E., Durand, M., Stoetzel, C., Doray, B., Viville, B., et al. 2011. Molecular diagnosis reveals genetic heterogeneity for the overlapping MKKS and BBS phenotypes. *Eur J Med Genet*, 54, 157-60.
- Schaefer, K. A., Wu, W. H., Colgan, D. F., Tsang, S. H., Bassuk, A. G., et al. 2017. Unexpected mutations after CRISPR-Cas9 editing in vivo. *Nat Methods*, 14, 547-548.
- Scheuermann, J. C. & Boyer, L. A. 2013. Getting to the heart of the matter: long non-coding RNAs in cardiac development and disease. *EMBO J*, 32, 1805-16.
- Schiffelers, R., Storm, G. & Bakker-Woudenberg, I. 2001. Liposome-encapsulated aminoglycosides in pre-clinical and clinical studies. *J Antimicrob Chemother*, 48, 333-44.
- Schott, J. J., Benson, D. W., Basson, C. T., Pease, W., Silberbach, G. M., et al. 1998. Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science*, 281, 108-11.
- Schulz, Y., Freese, L., Manz, J., Zoll, B., Volter, C., et al. 2014. CHARGE and Kabuki syndromes: a phenotypic and molecular link. *Hum Mol Genet*, 23, 4396-405.
- Schweighoffer, F., Faure, M., Fath, I., Chevallier-Multon, M. C., Apiou, F., et al. 1993. Identification of a human guanine nucleotide-releasing factor (H-GRF55) specific for Ras proteins. *Oncogene*, 8, 1477-85.
- Scott, A., Petrykowska, H. M., Hefferon, T., Gotea, V. & Elnitski, L. 2012. Functional analysis of synonymous substitutions predicted to affect splicing of the CFTR gene. *J Cyst Fibros*, 11, 511-7.

- Seidahmed, M. Z., Alkuraya, F. S., Shaheed, M., Al Zahrani, M., Al Manea, W., et al. 2011. Ritscher-Schinzel (cranio-cerebello-cardiac, 3C) syndrome: report of four new cases with renal involvement. *Am J Med Genet A*, 155A, 1393-7.
- Serrano-Bueno, G., Hernandez, A., Lopez-Lluch, G., Perez-Castineira, J. R., Navas, P., et al. 2013. Inorganic pyrophosphatase defects lead to cell cycle arrest and autophagic cell death through NAD⁺ depletion in fermenting yeast. *J Biol Chem*, 288, 13082-92.
- Sheridan, E., Wright, J., Small, N., Corry, P. C., Oddie, S., et al. 2013. Risk factors for congenital anomaly in a multiethnic birth cohort: an analysis of the Born in Bradford study. *Lancet*, 382, 1350-9.
- Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., et al. 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*, 29, 308-11.
- Shigemizu, D., Momozawa, Y., Abe, T., Morizono, T., Boroevich, K. A., et al. 2015. Performance comparison of four commercial human whole-exome capture platforms. *Sci Rep*, 5, 12742.
- Shimajima, K., Komoike, Y., Tohyama, J., Takahashi, S., Paez, M. T., et al. 2009. TULIP1 (RALGAPA1) haploinsufficiency with brain development delay. *Genomics*, 94, 414-22.
- Shinohara, K., Kawasumi, A., Takamatsu, A., Yoshida, S., Botilde, Y., et al. 2012. Two rotating cilia in the node cavity are sufficient to break left-right symmetry in the mouse embryo. *Nat Commun*, 3, 622.
- Shoemark, A., Dixon, M., Corrin, B. & Dewar, A. 2012. Twenty-year review of quantitative transmission electron microscopy for the diagnosis of primary ciliary dyskinesia. *J Clin Pathol*, 65, 267-71.
- Shoemark, A., Ives, A., Becker-Heck, A., Burgoyne, T., Dixon, M., et al. 2013. Inner Dynein Arm Defects in Primary Ciliary Dyskinesia. *Journal of Genetic Syndromes & Gene Therapy*, 4, 163.
- Sifrim, A., Hitz, M. P., Wilsdon, A., Breckpot, J., Turki, S. H., et al. 2016. Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nat Genet*, 48, 1060-5.
- Silversides, C. K., Lionel, A. C., Costain, G., Merico, D., Migita, O., et al. 2012. Rare copy number variations in adults with tetralogy of Fallot implicate novel risk gene pathways. *PLoS Genet*, 8, e1002843.
- Singh, V., Badheka, A. O., Patel, N. J., Chothani, A., Mehta, K., et al. 2015. Influence of hospital volume on outcomes of percutaneous atrial septal defect and patent foramen ovale closure: a 10-years US perspective. *Catheter Cardiovasc Interv*, 85, 1073-81.
- Skrzypek, H. & Hui, L. 2017. Noninvasive prenatal testing for fetal aneuploidy and single gene disorders. *Best Pract Res Clin Obstet Gynaecol*.
- Slavotinek, A. M., Schauer, G., Machin, G., Dasouki, M., Rueda-Pedraza, M. E., et al. 2005. Fryns syndrome: report of eight new cases. *Genet Med*, 7, 74-6.
- Slightom, J. L., Siemieniak, D. R., Sieu, L. C., Koop, B. F. & Hood, L. 1994. Nucleotide sequence analysis of 77.7 kb of the human V beta T-cell receptor gene locus: direct primer-walking using cosmid template DNAs. *Genomics*, 20, 149-68.
- Slough, J., Cooney, L. & Brueckner, M. 2008. Monocilia in the embryonic mouse heart suggest a direct role for cilia in cardiac morphogenesis. *Dev Dyn*, 237, 2304-14.
- Smemo, S., Campos, L. C., Moskowitz, I. P., Krieger, J. E., Pereira, A. C., et al. 2012. Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. *Hum Mol Genet*, 21, 3255-63.
- Smith, F. J., Eady, R. A., Leigh, I. M., Mcmillan, J. R., Rugg, E. L., et al. 1996. Plectin deficiency results in muscular dystrophy with epidermolysis bullosa. *Nat Genet*, 13, 450-7.
- Soemedi, R., Wilson, I. J., Bentham, J., Darlay, R., Topf, A., et al. 2012. Contribution of global rare copy-number variants to the risk of sporadic congenital heart disease. *Am J Hum Genet*, 91, 489-501.

- Southgate, L., Sukalo, M., Karountzos, A. S., Taylor, E. J., Collinson, C. S., et al. 2015. Haploinsufficiency of the NOTCH1 Receptor as a Cause of Adams-Oliver Syndrome With Variable Cardiac Anomalies. *Circ Cardiovasc Genet*, 8, 572-81.
- Srivastava, D. 2006. Making or breaking the heart: from lineage determination to morphogenesis. *Cell*, 126, 1037-48.
- Stitzel, N. O., Kiezun, A. & Sunyaev, S. 2011. Computational and statistical approaches to analyzing variants identified by exome sequencing. *Genome Biol*, 12, 227.
- Stolc, V., Samanta, M. P., Tongprasit, W. & Marshall, W. F. 2005. Genome-wide transcriptional analysis of flagellar regeneration in *Chlamydomonas reinhardtii* identifies orthologs of ciliary disease genes. *Proc Natl Acad Sci U S A*, 102, 3703-7.
- Stratakis, C. A., Kirschner, L. S. & Carney, J. A. 2001. Clinical and molecular features of the Carney complex: diagnostic criteria and recommendations for patient evaluation. *J Clin Endocrinol Metab*, 86, 4041-6.
- Sun, Z., Amsterdam, A., Pazour, G. J., Cole, D. G., Miller, M. S., et al. 2004. A genetic screen in zebrafish identifies cilia genes as a principal cause of cystic kidney. *Development*, 131, 4085-93.
- Sutherland, M. J. & Ware, S. M. 2009. Disorders of left-right asymmetry: heterotaxy and situs inversus. *Am J Med Genet C Semin Med Genet*, 151C, 307-17.
- Swarts, D. C., Mosterd, C., Van Passel, M. W. & Brouns, S. J. 2012. CRISPR interference directs strand specific spacer acquisition. *PLoS One*, 7, e35888.
- Sylva, M., Van Den Hoff, M. J. & Moorman, A. F. 2014. Development of the human heart. *Am J Med Genet A*, 164, 1347-71.
- Ta-Shma, A., Khan, T. N., Vivante, A., Willer, J. R., Matak, P., et al. 2017. Mutations in TMEM260 Cause a Pediatric Neurodevelopmental, Cardiac, and Renal Syndrome. *Am J Hum Genet*, 100, 666-675.
- Ta-Shma, A., Perles, Z., Yaacov, B., Werner, M., Frumkin, A., et al. 2015. A human laterality disorder associated with a homozygous WDR16 deletion. *Eur J Hum Genet*, 23, 1262-5.
- Taipale, M., Jarosz, D. F. & Lindquist, S. 2010. HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nat Rev Mol Cell Biol*, 11, 515-28.
- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., et al. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, 131, 861-72.
- Tang, L., Zeng, Y., Du, H., Gong, M., Peng, J., et al. 2017. CRISPR/Cas9-mediated gene editing in human zygotes using Cas9 protein. *Mol Genet Genomics*, 292, 525-533.
- Tariq, M., Belmont, J. W., Lalani, S., Smolarek, T. & Ware, S. M. 2011. SHROOM3 is a novel candidate for heterotaxy identified by whole exome sequencing. *Genome Biol*, 12, R91.
- Temple, G. F., Dozy, A. M., Roy, K. L. & Kan, Y. W. 1982. Construction of a functional human suppressor tRNA gene: an approach to gene therapy for beta-thalassaemia. *Nature*, 296, 537-40.
- Terada, H. 1990. Uncouplers of oxidative phosphorylation. *Environ Health Perspect*, 87, 213-8.
- Thiagalingam, A., De Bustros, A., Borges, M., Jasti, R., Compton, D., et al. 1996. RREB-1, a novel zinc finger protein, is involved in the differentiation response to Ras in human medullary thyroid carcinomas. *Mol Cell Biol*, 16, 5335-45.
- Thomas, M. G., Maconachie, G., Sheth, V., Mclean, R. J. & Gottlob, I. 2017. Development and clinical utility of a novel diagnostic nystagmus gene panel using targeted next-generation sequencing. *Eur J Hum Genet*, 25, 725-734.
- Thompson, B. A., Greenblatt, M. S., Vallee, M. P., Herkert, J. C., Tessereau, C., et al. 2013. Calibration of multiple in silico tools for predicting pathogenicity of mismatch repair gene missense substitutions. *Hum Mutat*, 34, 255-65.

- Thouvenot, P., Ben Yamin, B., Fourriere, L., Lescure, A., Boudier, T., et al. 2016. Functional Assessment of Genetic Variants with Outcomes Adapted to Clinical Decision-Making. *PLoS Genet*, 12, e1006096.
- Tian, Y., Yuan, L., Goss, A. M., Wang, T., Yang, J., et al. 2010. Characterization and in vivo pharmacological rescue of a Wnt2-Gata6 pathway required for cardiac inflow tract development. *Dev Cell*, 18, 275-87.
- Timmerman, L. A., Grego-Bessa, J., Raya, A., Bertran, E., Perez-Pomares, J. M., et al. 2004. Notch promotes epithelial-mesenchymal transition during cardiac development and oncogenic transformation. *Genes Dev*, 18, 99-115.
- Tomita-Mitchell, A., Mahnke, D. K., Struble, C. A., Tuffnell, M. E., Stamm, K. D., et al. 2012. Human gene copy number spectra analysis in congenital heart malformations. *Physiol Genomics*, 44, 518-41.
- Tomita-Mitchell, A., Maslen, C. L., Morris, C. D., Garg, V. & Goldmuntz, E. 2007. GATA4 sequence variants in patients with congenital heart disease. *J Med Genet*, 44, 779-83.
- Tran, P. V., Haycraft, C. J., Besschetnova, T. Y., Turbe-Doan, A., Stottmann, R. W., et al. 2008. THM1 negatively modulates mouse sonic hedgehog signal transduction and affects retrograde intraflagellar transport in cilia. *Nat Genet*, 40, 403-10.
- Trapnell, C., Pachter, L. & Salzberg, S. L. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 25, 1105-11.
- Triedman, J. K. & Newburger, J. W. 2016. Trends in Congenital Heart Disease: The Next Decade. *Circulation*, 133, 2716-33.
- Tsang, A. P., Visvader, J. E., Turner, C. A., Fujiwara, Y., Yu, C., et al. 1997. FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell*, 90, 109-19.
- Uchida, K., Konishi, N., Inoue, M., Otake, K. & Kusunoki, M. 2006. A case of congenital jejunal atresia associated with bilateral athelia and choanal atresia: new syndrome spectrum. *Clin Dysmorphol*, 15, 37-8.
- Valdez, R. M., Barbero, P. M., Liascovich, R. C., De Rosa, L. F., Aguirre, M. A., et al. 2007. Methimazole embryopathy: a contribution to defining the phenotype. *Reprod Toxicol*, 23, 253-5.
- Valdmanis, P. N., Meijer, I. A., Reynolds, A., Lei, A., Macleod, P., et al. 2007. Mutations in the KIAA0196 gene at the SPG8 locus cause hereditary spastic paraplegia. *Am J Hum Genet*, 80, 152-61.
- Van Der Bom, T., Bouma, B. J., Meijboom, F. J., Zwinderman, A. H. & Mulder, B. J. 2012. The prevalence of adult congenital heart disease, results from a systematic review and evidence based calculation. *Am Heart J*, 164, 568-75.
- Van Der Bom, T., Zomer, A. C., Zwinderman, A. H., Meijboom, F. J., Bouma, B. J., et al. 2011. The changing epidemiology of congenital heart disease. *Nat Rev Cardiol*, 8, 50-60.
- Van Laarhoven, P. M., Neitzel, L. R., Quintana, A. M., Geiger, E. A., Zackai, E. H., et al. 2015. Kabuki syndrome genes KMT2D and KDM6A: functional analyses demonstrate critical roles in craniofacial, heart and brain development. *Hum Mol Genet*, 24, 4443-53.
- Van Rooijen, E., Giles, R. H., Voest, E. E., Van Rooijen, C., Schulte-Merker, S., et al. 2008. LRRC50, a conserved ciliary protein implicated in polycystic kidney disease. *J Am Soc Nephrol*, 19, 1128-38.
- Van Veldhoven, P. P. & Mannaerts, G. P. 1987. Inorganic and organic phosphate measurements in the nanomolar range. *Anal Biochem*, 161, 45-8.
- Vandenberg, L. N. & Levin, M. 2013. A unified model for left-right asymmetry? Comparison and synthesis of molecular models of embryonic laterality. *Dev Biol*, 379, 1-15.

- Verhagen, J. M., Oostdijk, W., Terwisscha Van Scheltinga, C. E., Schalij-Delfos, N. E. & Van Bever, Y. 2014. An unusual presentation of Kabuki syndrome: clinical overlap with CHARGE syndrome. *Eur J Med Genet*, 57, 510-2.
- Villa-Bellosta, R., Hamczyk, M. R. & Andres, V. 2017. Novel phosphate-activated macrophages prevent ectopic calcification by increasing extracellular ATP and pyrophosphate. *PLoS One*, 12, e0174998.
- Villafane, J., Lantin-Hermoso, M. R., Bhatt, A. B., Tweddell, J. S., Geva, T., et al. 2014. D-transposition of the great arteries: the current era of the arterial switch operation. *J Am Coll Cardiol*, 64, 498-511.
- Vissers, L. E., Van Ravenswaaij, C. M., Admiraal, R., Hurst, J. A., De Vries, B. B., et al. 2004. Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat Genet*, 36, 955-7.
- Volk, S. E., Baykov, A. A., Kostenko, E. B. & Avaeva, S. M. 1983. Isolation, subunit structure and localization of inorganic pyrophosphatase of heart and liver mitochondria. *Biochim Biophys Acta*, 744, 127-34.
- Von Morgen, P., Horejsi, Z. & Macurek, L. 2015. Substrate recognition and function of the R2TP complex in response to cellular stress. *Front Genet*, 6, 69.
- Voora, D., Mcleod, H. L., Eby, C. & Gage, B. F. 2005. The pharmacogenetics of coumarin therapy. *Pharmacogenomics*, 6, 503-13.
- Walters, A. M., Porter, G. A., Jr. & Brookes, P. S. 2012. Mitochondria as a drug target in ischemic heart disease and cardiomyopathy. *Circ Res*, 111, 1222-36.
- Wang, C., Zhao, Y., Ming, Y., Zhao, S. & Guo, Z. 2016a. A polymorphism at the microRNA binding site in the 3'-untranslated region of C14orf101 is associated with the risk of gastric cancer development. *Exp Ther Med*, 12, 1867-1872.
- Wang, W., Niu, Z., Wang, Y., Li, Y., Zou, H., et al. 2016b. Comparative transcriptome analysis of atrial septal defect identifies dysregulated genes during heart septum morphogenesis. *Gene*, 575, 303-12.
- Wang, X., Wang, J., Zhao, P., Guo, Y., Wu, L., et al. 2013. Familial congenital heart disease. *Cardiol Young*, 23, 394-9.
- Warnes, C. A., Williams, R. G., Bashore, T. M., Child, J. S., Connolly, H. M., et al. 2008. ACC/AHA 2008 Guidelines for the Management of Adults with Congenital Heart Disease: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (writing committee to develop guidelines on the management of adults with congenital heart disease). *Circulation*, 118, e714-833.
- Warnick, M. 1984. Report of the Committee of Inquiry into Human Fertilisation and Embryology. *Her Majesty's Stationary Office*.
- Wei, D., Bao, H., Liu, X. Y., Zhou, N., Wang, Q., et al. 2013. GATA5 loss-of-function mutations underlie tetralogy of fallot. *Int J Med Sci*, 10, 34-42.
- Weleber, R. G., Pennesi, M. E., Wilson, D. J., Kaushal, S., Erker, L. R., et al. 2016. Results at 2 Years after Gene Therapy for RPE65-Deficient Leber Congenital Amaurosis and Severe Early-Childhood-Onset Retinal Dystrophy. *Ophthalmology*, 123, 1606-20.
- Wilkinson, R. N., Jopling, C. & Van Eeden, F. J. 2014. Zebrafish as a model of cardiac disease. *Prog Mol Biol Transl Sci*, 124, 65-91.
- Wilson, L. C., Kerr, B. A., Wilkinson, R., Fossard, C. & Donnai, D. 1998. Choanal atresia and hypothelia following methimazole exposure in utero: a second report. *Am J Med Genet*, 75, 220-2.
- Wren, C., Birrell, G. & Hawthorne, G. 2003. Cardiovascular malformations in infants of diabetic mothers. *Heart*, 89, 1217-20.
- Yagi, H., Furutani, Y., Hamada, H., Sasaki, T., Asakawa, S., et al. 2003. Role of TBX1 in human del22q11.2 syndrome. *Lancet*, 362, 1366-73.
- Yamada, T., Inoue, T., Nishida, T., Furuya, E. & Tagawa, K. 1988. Hepatic accumulation of pyrophosphate during acetate metabolism. *J Biochem*, 104, 847-50.

- Yamamoto, R., Hirono, M. & Kamiya, R. 2010. Discrete PIH proteins function in the cytoplasmic preassembly of different subsets of axonemal dyneins. *J Cell Biol*, 190, 65-71.
- Yamashita, H., Kaneyuki, T. & Tagawa, K. 2001. Production of acetate in the liver and its utilization in peripheral tissues. *Biochim Biophys Acta*, 1532, 79-87.
- Yang, B., Liu, C., Diao, L., Wang, C. & Guo, Z. 2014. A polymorphism at the microRNA binding site in the 3' untranslated region of C14orf101 is associated with non-Hodgkin lymphoma overall survival. *Cancer Genet*, 207, 141-6.
- Yin, Z. & Kirschner, L. S. 2009. The Carney complex gene PRKAR1A plays an essential role in cardiac development and myxomatogenesis. *Trends Cardiovasc Med*, 19, 44-9.
- Yunis, K., Mumtaz, G., Bitar, F., Chamseddine, F., Kassab, M., et al. 2006. Consanguineous marriage and congenital heart defects: a case-control study in the neonatal period. *Am J Med Genet A*, 140, 1524-30.
- Zaidi, S. & Brueckner, M. 2017. Genetics and Genomics of Congenital Heart Disease. *Circ Res*, 120, 923-940.
- Zaidi, S., Choi, M., Wakimoto, H., Ma, L., Jiang, J., et al. 2013. De novo mutations in histone-modifying genes in congenital heart disease. *Nature*, 498, 220-3.
- Zalli, D., Bayliss, R. & Fry, A. M. 2012. The Nek8 protein kinase, mutated in the human cystic kidney disease nephronophthisis, is both activated and degraded during ciliogenesis. *Hum Mol Genet*, 21, 1155-71.
- Zariwala, M. A., Knowles, M. R. & Omran, H. 2007. Genetic defects in ciliary structure and function. *Annu Rev Physiol*, 69, 423-50.
- Zhang, J., Wilson, G. F., Soerens, A. G., Koonce, C. H., Yu, J., et al. 2009. Functional cardiomyocytes derived from human induced pluripotent stem cells. *Circ Res*, 104, e30-41.
- Zhang, P., Lee, H., Brunzelle, J. S. & Couture, J. F. 2012. The plasticity of WDR5 peptide-binding cleft enables the binding of the SET1 family of histone methyltransferases. *Nucleic Acids Res*, 40, 4237-46.
- Zhang, Q. J. & Liu, Z. P. 2015. Histone methylations in heart development, congenital and adult heart diseases. *Epigenomics*, 7, 321-30.
- Zhao, L., Yuan, S., Cao, Y., Kallakuri, S., Li, Y., et al. 2013. Reptin/Ruvbl2 is a Lrrc6/Seahorse interactor essential for cilia motility. *Proc Natl Acad Sci U S A*, 110, 12697-702.
- Zhao, R., Kakihara, Y., Gribun, A., Huen, J., Yang, G., et al. 2008. Molecular chaperone Hsp90 stabilizes Pih1/Nop17 to maintain R2TP complex activity that regulates snoRNA accumulation. *J Cell Biol*, 180, 563-78.
- Zhao, Y., Samal, E. & Srivastava, D. 2005. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature*, 436, 214-20.
- Zheng, G. F., Wei, D., Zhao, H., Zhou, N., Yang, Y. Q., et al. 2012. A novel GATA6 mutation associated with congenital ventricular septal defect. *Int J Mol Med*, 29, 1065-71.
- Zhou, J., Goldberg, E. M., Leu, N. A., Zhou, L., Coulter, D. A., et al. 2014. Respiratory failure, cleft palate and epilepsy in the mouse model of human Xq22.1 deletion syndrome. *Hum Mol Genet*, 23, 3823-9.
- Zhu, H., Kartiko, S. & Finnell, R. H. 2009. Importance of gene-environment interactions in the etiology of selected birth defects. *Clin Genet*, 75, 409-23.
- Zhu, L., Vranckx, R., Khau Van Kien, P., Lalande, A., Boisset, N., et al. 2006. Mutations in myosin heavy chain 11 cause a syndrome associating thoracic aortic aneurysm/aortic dissection and patent ductus arteriosus. *Nat Genet*, 38, 343-9.
- Zhu, Q., Wang, L., Xiao, Z., Xiao, F., Luo, J., et al. 2013. Decreased expression of Ras-GRF1 in the brain tissue of the intractable epilepsy patients and experimental rats. *Brain Res*, 1493, 99-109.

Zondag, G. C., Reynolds, A. B. & Moolenaar, W. H. 2000. Receptor protein-tyrosine phosphatase RPTPmu binds to and dephosphorylates the catenin p120(ctn). *J Biol Chem*, 275, 11264-9.

Appendix A

Inclusion Criteria

This project aims to recruit families with a likely genetic cause of Congenital Heart Disease (CHD).

CHD includes all structural, valvular and outflow tract lesions. The study will also include congenital and early-onset cardiomyopathy (<2 years of age).

CHD may or may not be associated with other congenital anomalies (syndromic CHD)

Included families:

- 1) Consanguineous families with at least one child affected with CHD
- 2) Sib pairs affected with CHD in non-consanguineous union
- 3) Families with dominant or possible dominant (with reduced penetrance) pattern of CHD
- 4) Parent and child affected with CHD

AND

CHD should be concordant within the family- e.g. AVSD in all family members or same 'type' of CHD e.g. septal defects, outflow tract defects, laterality defects (i.e. potential underlying genetic basis).

The presence of a normal karyotype/ 22q11 deletion testing/ Array CGH in an affected family member

Patient able to give informed consent, or if <16 years parent/guardian able to give informed consent

Patient under the care of Consultant Cardiologist or Clinical Geneticist.

Exclusion Criteria

The presence of a recognised genetic syndrome associated with CHD eg. Noonan syndrome.

Appendix B



UNIVERSITY OF LEEDS

The Leeds Teaching Hospitals 
NHS Trust

PARTICIPANT'S INFORMATION SHEET for parents

INVITATION TO TAKE PART IN RESEARCH TO STUDY GENES THAT CAUSE RECESSIVE INHERITED CONDITIONS:

"Molecular Genetic Investigations of Autosomal Recessive Conditions"

Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. Talk to others about the study if you wish.

- **Part 1** tells you the purpose of this study and what will happen to you if you take part.
- **Part 2** gives you more detailed information about the conduct of the study.

Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

PART 1: to give you first thoughts about the project

WHAT DO WE WANT TO DO?

We are studying the inherited factors (genes) involved in recessive conditions. Recessive conditions are genetic conditions that a couple can pass on to their children, but do not affect the couple themselves. Some children might have problems with their walking or learning, and might need operations because they have structural birth defects, such as heart defects, kidney defects, or abnormalities picked up on a brain scan. In some cases the gene causing the condition isn't known and the aim of this study is to do research to find these faulty genes. This could lead to a better understanding of the condition.

WHY ARE WE ASKING FOR YOUR HELP?

You have been asked to take part in this research study because your child has a recessive condition but the genetic cause still isn't known. As most recessive conditions are very rare, we are seeking your help so that we can look at the genes of many families with exactly the same condition.

DO I HAVE TO TAKE PART?

No. It is up to you to decide whether or not to take part. If you do, you will be given this information sheet to keep and be asked to sign a consent form. If at a later date you decide you do not want to take part in the study you can change your mind for any reason. If you do not wish to take part in the study this will not affect your or your family's medical care in any way.

WHAT WILL WE BE ASKING YOU TO DO?

A researcher will discuss the study with you and answer any questions you may have. When you are satisfied you have all the information you require and if you decide to take part, we will ask you to read and complete a consent form. The researcher will ask you to sign this form to give your consent. The form will ask your permission for an extra blood test on your child, yourself and your partner. We will try to only take blood from your child at a time when blood is being taken as part of the usual clinical care. If we already have a sample of your child's blood we may ask you if we can use that instead of taking a new sample. We may also ask for a saliva sample in addition or instead of a blood sample, and we may ask your permission to take samples from your other children if you have them, including those that are healthy. We will ask some questions about your medical and family history, and we may also ask your permission to look at your medical records. You will be given a copy of this information sheet and your signed consent form to keep.

WHAT INFORMATION WILL WE HOLD?

We will hold information that is routinely collected as part of your child's usual clinical care. We will also hold information that we gain from the blood samples that we have taken. In the future, we may need to contact you again for further medical information.

ARE THERE ANY POSSIBLE DISADVANTAGES TO TAKING PART IN THE STUDY?

We do not think that taking part in this study will harm you in any way. We will be gathering information which is already routinely collected by health professionals. If at any point you would like to discuss the study and your involvement in it, you will be able to speak to a researcher. Contact details are provided at the end of this information sheet.

WILL THIS PROJECT BENEFIT MY FAMILY?

There are no immediate short term advantages to taking part in this study. We cannot guarantee to discover anything that will directly benefit you or your family. However if we do find the faulty gene causing the condition in your family then tests may be available for your relatives to find out if they are at risk of having children with the condition. We also hope that studying the genetic causes of recessive conditions will help to discover the reasons why these conditions occur. By increasing understanding in this way we may, in the long term, be able to offer better help to children and their families

WHAT IF THERE IS A PROBLEM?

We will address any complaint about the way you have been dealt with during the study or any possible harm that you might suffer. The detailed information on this is given in Part 2.

THIS COMPLETES PART 1 OF THE INFORMATION SHEET

If the information in Part 1 has interested you and you are considering participation, please continue to read the additional information in Part 2 before making any decision.

PART 2: information you need to know if you still want to take part**WHAT IF NEW INFORMATION BECOMES AVAILABLE?**

We hope that this study will uncover the genetic cause for your child's problems. If we do find the faulty gene, this information will be fed back to you in the course of the clinics run by the doctors caring for your child. We will also inform your local clinical genetics department so they can contact you. It may be possible to then offer tests to the wider family, if they wish to see if they carry the same gene or not. It may also be possible to offer tests to either you or your partner during a future pregnancy.

HOW DO WE ENSURE CONFIDENTIALITY?

All information recorded on paper and any biological samples stored will be kept under conditions of strict confidentiality as a legal requirement under the Data Protection and Human Tissue Acts. We will put information about your family onto a secure computer database. However, we will remove all personal details, such as names and addresses, so your family cannot be recognized from it. A study number, which can only be linked to you by the research team, will be the identifier of the information. This is necessary so that we can put together your information throughout the study. The results from any information or biological samples that we collect will only be used for research purposes and will not be available to anyone else.

WILL MY GENERAL PRACTITIONER/FAMILY DOCTOR KNOW THAT I AM PART OF THIS STUDY?

With your permission, we will contact your G.P. or other doctor involved in your family's care to let them know that your child is included in the study.

WILL ANYONE ELSE KNOW ABOUT MY TAKING PART?

The information collected about you during the course of the research project will be kept strictly confidential and you will not be identifiable from it. If any research results are published in medical articles as a result of this project, all personal details will be removed so that your family cannot be recognized from it.

WHAT WILL HAPPEN TO ANY SAMPLES THAT I GIVE?

We will request an extra sample bottle be filled at the time your child is having blood tests as part of their usual clinical care. We will also request a sample of blood from you and your partner. We may also ask to take samples from your other children if you have them, including those that are healthy. We could also ask you or your partner to give a saliva sample instead or in addition to a blood sample. The samples will be sent to our lab and DNA will be extracted. We will analyse the DNA to see if we can determine if there is a genetic cause for your child's problems. If you change your mind later and decide not to take part in the study then you should let us know what we should do with any DNA samples. We need to ask you because it can often take some years to find the faulty gene that causes the specific condition in your family. We can either keep the DNA for research in the future, or keep the DNA but remove any details that can identify it as yours, or destroy the sample.

WILL ANY GENETIC TESTS BE DONE?

Tests will be done to try to establish if your child's problems are caused by a faulty gene. We will not use the sample for any other genetic tests. We may use new technology (called "clonal" or next generation sequencing) to look at many genes at once rather than one after another. We want to use this new technology because we might find the faulty gene more quickly, but we will only use it for studies appropriate for your child's condition. It is possible that we may uncover other findings that are unconnected to your child's condition but might be important for your health. With your permission, we will inform your local clinical genetics department so they can contact you about this information.

WHO IS ORGANISING AND FUNDING THE RESEARCH?

The study is organized by the University of Leeds. The study is funded by a medical research charity called the Sir Jules Thorn Charitable Trust. Neither you nor the researchers involved will benefit financially from this project.

WHO HAS REVIEWED THE STUDY?

All research in the NHS is looked at by an independent group of people, called a Research Ethics Committee to protect your safety, rights, wellbeing and dignity. This study has been reviewed and given a favourable opinion by South Yorkshire Research Ethics Committee. The research has also been reviewed by the Sir Jules Thorn Charitable Trust, to ensure that the proper science is being done for this study.

WHAT IF I HAVE ANY CONCERNS?

If you have any concerns, other questions about this study or the way it has been carried out, you should contact the investigators in charge of running the study:

Prof. Colin Johnson:

Section of Ophthalmology and Neurosciences
Wellcome Trust Brenner Building
Leeds Institute of Molecular Medicine
St James's University Hospital
Beckett Street
Leeds, LS9 7TF, U.K.

tel: (+44) 0113 343 8443
fax: (+44) 0113 343 8603

e-mail: c.johnson@leeds.ac.uk

Dr. Eamonn Sheridan:

Section of Genetics
Wellcome Trust Brenner Building
Leeds Institute of Molecular Medicine
St James's University Hospital
Beckett Street
Leeds, LS9 7TF, U.K.

tel: (+44) 0113 206 5927

e-mail: e.sheridan@leeds.ac.uk

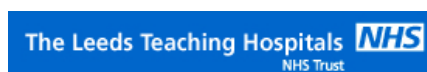
COMPLAINTS

If you have a concern about any aspect of this study, you should ask to speak with the researchers who will do their best to answer your questions. Please contact Prof. Johnson on 0113 343 8443 or Dr. Sheridan on 0113 206 5927 in the first instance. If you remain unhappy and wish to complain formally, you can do this through the NHS Complaints Procedure. Details can be obtained from the hospital.

HARM

In the event that something does go wrong and you are harmed during the research study there are no special compensation arrangements. If you are harmed and this is due to someone's negligence then you may have grounds for a legal action for compensation against Leeds Teaching Hospitals NHS Trust but you may have to pay your legal costs. The normal National Health Service complaints procedure will still be available to you.

Appendix C



Title of Project: MOLECULAR GENETIC INVESTIGATIONS OF AUTOSOMAL RECESSIVE CONDITIONS

CONSENT FORM: for patients, or parents/guardians and children in the family Participant to circle responses to all parts:

1. I confirm that I have read and understand the information sheet for the study. I have had the opportunity to consider the information, ask questions, and have had these answered satisfactorily YES/NO

 2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and without my medical care or legal rights being affected YES/NO

 3. I understand that sections of any of my/my child's medical notes may be reviewed by the researchers and other responsible individuals from regulatory authorities or from the NHS Trust, where it is relevant to my taking part in research. I give permission for these individuals to have access to my/our records

 4. i) I agree to allow medical information about me/my family to be entered on a confidential computer database. I understand that personal details (including names and addresses) about myself and my family will be accessed **only** by researcher team leaders (Johnson, Sheridan & Bonthron). Information accessed by other researchers will have personal details removed, thereby maintaining our privacy YES/NO
 - ii) If further medical information is requested by researchers, I agree to be contacted again for this purpose. YES/NO

 5. i) I agree that a blood sample/saliva sample/stored DNA sample from me may be used for genetic research studies appropriate to my family's condition. I understand that any results arising from this research work will be kept **strictly confidential**. YES/NO
 - ii) **OR** I agree that blood samples/saliva samples/stored DNA samples from the following **children** in the family can be used for the project: YES/NO
- list of **children** agreed to be part of the research:
- | | |
|----|----|
| 1. | 5. |
| 2. | 6. |
| 3. | 7. |
| 4. | 8. |
-
6. i) If a genetic test, or other genetic finding that might have health implications, becomes available as a result of medical research on my own or my family's sample(s) I would like to have the opportunity to discuss the implications of these findings with appropriate medical experts, including my local clinical genetics department YES/NO

 - ii) I understand that genetic testing may sometimes reveal information that might have health implications **unconnected** to my family's condition. I would like to have the opportunity to discuss the implications of any such findings with appropriate medical experts, including my local clinical genetics department YES/NO

 - iii) I agree that my G.P, or other healthcare professional in charge of my or my family's care, is informed that I/we are taking part in this study and of the result of any genetic test YES/NO

Participant/Family Representative
on behalf of children in the family

Date

Signature

Name of Person taking consent

Date

Signature

Researcher

Date

Signature

when completed copies: one for participant; one for researcher; signed original to be kept with hospital notes

Appendix D



National Research Ethics Service

South Yorkshire Research Ethics Committee

Millside
Mill Pond Lane
Meanwood
Leeds
LS6 4RA

Telephone: 0113 305 0128

18 February 2011

Prof Colin A Johnson
Professor of Medical & Molecular Genetics
Section of Ophthalmology and Neuroscience
Leeds Inst of Molecular Medicine
St James University Hospital
Leeds LS9 7TF

Dear Prof Johnson

Study Title: Molecular genetic investigations of autosomal recessive conditions
REC reference number: 11/H1310/1

Thank you for your letter of 8th February 2011, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Ethical review of research sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

For NHS research sites only, management permission for research ("R&D approval") should be obtained from the relevant care organisation(s) in accordance with NHS research governance arrangements. Guidance on applying for NHS permission for research is available in the Integrated Research Application System or at <http://www.rdforum.nhs.uk>.

This Research Ethics Committee is an advisory committee to the Yorkshire and The Humber Strategic Health Authority
The National Research Ethics Service (NRES) represents the NRES Directorate within
the National Patient Safety Agency and Research Ethics Committees in England

Where the only involvement of the NHS organisation is as a Participant Identification Centre (PIC), management permission for research is not required but the R&D office should be notified of the study and agree to the organisation's involvement. Guidance on procedures for PICs is available in IRAS. Further advice should be sought from the R&D office where necessary.

Sponsors are not required to notify the Committee of approvals from host organisations.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	Version	Date
Investigator CV	1	14 December 2010
Covering Letter		08 February 2011
Letter from Sponsor		
REC application		14 December 2010
Response to Request for Further Information		
Participant Information Sheet: for children, aged 8 years or younger (Mirpuri Urdu translation)	1	14 December 2010
Evidence of insurance or indemnity		
Participant Information Sheet: Children aged 12 - 15 years	2	08 February 2011
Letter of invitation to participant	1	14 December 2010
Participant Information Sheet: for parents (Mirpuri Urdu translation)	1	14 December 2010
Participant Information Sheet: Relatives that are adults or young persons	2	08 February 2011
Participant Information Sheet: for children aged 8 year or younger	1	14 December 2010
Participant Information Sheet: Parents	2	08 February 2011
Participant Information Sheet: for children aged 8-12 years	1	14 December 2010
Participant Consent Form: Assent form for older children	1	14 December 2010
Protocol	1	14 December 2010

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Now that you have completed the application process please visit the National Research Ethics Service website > After Review

You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the website.

The attached document "*After ethical review – guidance for researchers*" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Progress and safety reports
- Notifying the end of the study

The NRES website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

We would also like to inform you that we consult regularly with stakeholders to improve our service. If you would like to join our Reference Group please email referencegroup@nres.npsa.nhs.uk.

11/H1310/1	Please quote this number on all correspondence
------------	--

With the Committee's best wishes for the success of this project

Yours sincerely




MS Jo Abbott
Chair

Email: Sinead.audsley@leedspft.nhs.uk

Enclosures: "After ethical review – guidance for researchers"

Copy to: Mrs Rachel E de Souza, University of Leeds
Mrs Anne Gowing, Leeds Teaching Hospitals NHS Trust

The Leeds Teaching Hospitals 
NHS Trust

Ref: Amanda Burd

Research & Development

29/04/2011

Leeds Teaching Hospitals NHS Trust

34 Hyde Terrace
Leeds
LS2 9LN

Dr Colin A. Johnson

Tel: 0113 392 2878
Fax: 0113 392 6397

Section of Ophthalmology and Neurosciences,
Wellcome Trust Brenner Bui
Leeds Institute of Molecular Medicine,
St. James's University Hospital
LS9 7TF

r&d@leedsth.nhs.uk
www.leedsth.nhs.uk

Dear Dr Johnson

**Re: NHS Permission at LTHT for: Molecular genetic investigations of
autosomal recessive conditions
LTHT R&D Number: CG11/9764 (53788/WVY)
REC: 11/H1310/1**

I confirm that *NHS Permission for research* has been granted for this project at The Leeds Teaching Hospitals NHS Trust (LTHT). NHS Permission is granted based on the information provided in the documents listed below. All amendments (including changes to the research team) must be submitted in accordance with guidance in IRAS. Any change to the status of the project must be notified to the R&D Department.

Permission is granted on the understanding that the study is conducted in accordance with the *Research Governance Framework for Health and Social Care*, ICH GCP (if applicable) and NHS Trust policies and procedures available at http://www.leedsth.nhs.uk/sites/research_and_development/.

This permission is granted only on the understanding that you comply with the requirements of the *Framework* as listed in the attached sheet "Conditions of Approval".

If you have any queries about this approval please do not hesitate to contact the R&D Department on telephone 0113 392 2878.

Indemnity Arrangements

Chairman Mike Collier CBE Chief Executive Maggie Boyle

The Leeds Teaching Hospitals Incorporating:

Chapel Allerton Hospital Leeds Dental Institute Seacroft Hospital

St James's University Hospital The General Infirmary at Leeds Wharfedale Hospital



01/200

The Leeds Teaching Hospitals NHS Trust participates in the NHS risk pooling scheme administered by the NHS Litigation Authority 'Clinical Negligence Scheme for NHS Trusts' for: (i) medical professional and/or medical malpractice liability; and (ii) general liability. NHS Indemnity for negligent harm is extended to researchers with an employment contract (substantive or honorary) with the Trust. The Trust only accepts liability for research activity that has been managerially approved by the R&D Department.

The Trust therefore accepts liability for the above research project and extends indemnity for negligent harm to cover you as investigator and the researchers listed on the Site Specific Information form. Should there be any changes to the research team please ensure that you inform the R&D Department and that s/he obtains an appropriate contract, or letter of access, with the Trust if required.

Yours sincerely



Dr D R-Norfolk
Associate Director of R&D

Approved documents

The documents reviewed and approved are listed as follows

<i>Document</i>	<i>Version</i>	<i>Date of document</i>
NHS R&D Form	3.1	02/03/2011
SSI Form	3.1	08/03/2011
Directorate Approval		08/03/2011
Radiology Approval		N/A
Pharmacy Approval		N/A
Protocol	1.0	14/12/2010
REC Letter confirming favourable opinion		18/02/2011
Evidence of Insurance		Not dated
Patient information sheet (REC Approved) Parent	2.0	08/02/2011
Patient information sheet (REC Approved) 8/younger URDU	1.0	14/12/2010
Patient information sheet (REC Approved) Parent URDU	1.0	14/12/2010
Patient information sheet (REC Approved) 12-15 yrs	2.0	08/02/2011
PIS (REC Approved) Adult relatives/young Person	2.0	08/02/2011
PIS (REC Approved) Children aged 8 or younger	1.0	14/12/2010
PIS (REC Approved) Children aged 8 – 12 years	1.0	14/12/2010
Informed Consent (REC Approved) Assent older children	1.0	14/12/2010
Letter of Invitation to participants (REC Approved)	1.0	14/12/2010

Appendix E

Primer sequences used in this thesis

Primer	Sequence	Product Size	Annealing temp.
c2orf88 F	GAAATCAAAGCAAACCTTTCCC	396	58.9
c2orf88 R	AATGTCATGGTACTTGATGTTATT G	396	58.9
DNAAF1_central_plasmid_F	CTGGATGCTCTTAACCTCAGC		
DNAAF1_centra_lplasmid_R	TCTTCCAAGTCAGGTAGGTCA		
DNAAF1_ex1F	GTACTCTCTGGCTGGGCTGG	296	59.5
DNAAF1_ex1R	GTATGTGGTGTGGGAAGGG	296	59.5
DNAAF1_ex2F	TAACCAAGCTGAAGTTGGGG	287	58.9
DNAAF1_ex2R	TTATCCGTACAAGGGGGTGA	287	58.9
DNAAF1_ex3F	GGCAGGAATGGATGTGGTAA	268	58.9
DNAAF1_ex3R	AAAGGTGATCCCTGCAAGTAA	268	58.9
DNAAF1_ex4F	AAAAACAAGGGTGACCGTGA	396	58.9
DNAAF1_ex4R	AAACACCACTTCTTGCCCAT	396	58.9
DNAAF1_ex5F	ATTCTCCCGCTTTAGCCTTC	382	58.9
DNAAF1_ex5R	GGCAGGAAGGAGAGAAAAGA	382	58.9
DNAAF1_ex6F	CCAGGACAGGATATTGGCAC	390	58.9
DNAAF1_ex6R	GCCATCAAGCCTATTTCCAA	390	58.9
DNAAF1_ex7F	CATGTCTGATGCTCACTTTGC	364	58.9
DNAAF1_ex7R	TCCAAGATTAAGACTGGGTT	364	58.9
DNAAF1_ex8_part_1F	AGCCCTTGATGTACCCACAA	485	51
DNAAF1_ex8_part_1R	TTCACAGGCGGTGACAGTAG	485	51
DNAAF1_ex8_part_2F	CCAGCTGAGACCCTGCTACT	356	59.5
DNAAF1_ex8_part_2R	CCACAGAGACGTGAGTCGAA	356	59.5
DNAAF1_ex9F	AGCCCATCTTCACCGTAGG	295	58.9
DNAAF1_ex9R	ACCACACTTGGGGTATCGAG	295	58.9
DNAAF1_ex10F	TGCCTTTATCGTGCCTATCA	243	58.9
DNAAF1_ex10R	TGCTTTATGCACAGGGAACA	243	58.9
DNAAF1_ex11_part_1F	AACTGGCATGGTGCATTGTA	475	51
DNAAF1_ex11_part_1R	GTCTCCACTGTGCTCTCGG	475	51
DNAAF1_ex11_part_2F	ACAAGACACCAAGTCCCAA	241	58.9
DNAAF1_ex11_part_2R	TAAGGCTGAGTGACTGCCCT	241	58.9
DNAAF1_ex12F	GAATTTGGCCTGGACTGAAC	298	51
DNAAF1_ex12R	TCATACACTGACTCGCGTGG	298	51
KMT2D_exon38_F	AGGGTCGGAGAGGTCAGG	399	63
KMT2D_exon38_R	GGTTTGGGTGATATGGAGGA	399	63
KMT2D_G3553V_F	TCTTGCTCTGCTGCTCTGTA	171	58.9
KMT2D_G3553V_R	ACCCAGCACTCCTACAGATG	171	58.9
pgLAP1_F	ATCACTCTCTCGGCATGGACGAG CTGTACAAG		
pgLAP1_R	TGGCTGGCAACTAGAAGGCACAG TCGAGGC		
RBM48_exon5_F	CAAGGATTGAACTATTGGATGTAG TC	285	62
RBM48_exon5_R	AATACCAGCAGTAAAATCTCTTGA A	285	62
RREB1_F	CCACCACAGACACCAACAAG	248	63
RREB1_R	GTGTGTGAGCATGTGCCTCT	248	63
SDM_DNAAF1_1634G>T_F	AAGGAGACATTCTGCATTGATG		59
SDM_DNAAF1_1634G>T_R	TGTCTCCAGTCTAATGGTTTC		59
SDM_DNAAF1_571C>T_F	CATTGAAAACCTTCTCCTGCCTC		60
SDM_DNAAF1_571C>T_R	GTCTTGATGTAATTGTTGCTG		60
SP6	ATTTAGGTGACACTATAG		

T3	GCAATTAACCCTCACTAAAGG		
T7	TAATACGACTCACTATAGGG		
WGS breakpoint F	TGTAAAACGACGGCCAGTTGTTCT GGAGGGACGATGAG		
WGS breakpoint deletion R	CAGGAAACAGCTATGACCTTGTG TGGCAGCTGTGAAA		
WGS breakpoint wildtype R	CAGGAAACAGCTATGACCAGCCA TCAAGCCTATTTCCAA		

SiRNAs

ON-TARGETplus SMARTpool RUVBL1	AUUAAGGUGGUGAACAAGUA
	GGGAAGGACAGCAUUGAGA
	CAGGAUAAGUACAUGAAGU
	CUCAGGAGCUGGGUAGUAA
ON-TARGETplus SMARTpool RUVBL2	UAACAAGGAUUGAGCGAAU
	CGCAGUACAUGAAGGAGUA
	GAAACGCAAGGGUACAGAA
	GCGAGAAAGACACGAAGCA

Appendix F

Depth of coverage results for all samples studied in this thesis.

Sample number	Sample type	% duplication	Mean coverage (no. reads)	% bases covered by >5 reads	Instrument	Sequencing kit	Library version	Total reads (CASM)	Mapped reads	% mapped
JT67	blood	1.58	68.2	98.7	Hiseq	XT	v4	56054364	54787753	97.7404
JT377	blood	5.27	110.3	99	Hiseq	XT	v4	87809314	86101613	98.0552
JT378	blood	4.5391	121.7	99	Hiseq	XT	v4	100248326	98278482	98.035
JT379	blood	2.1309	121.2	99	Hiseq	XT	v4	93726067	91931756	98.0856
JT380	blood	5.0487	83.9	99	Hiseq	XT	v4	65257890	63977628	98.0381
JT384	blood	6.6175	74.7	99	Hiseq	XT	v4	62746406	61423573	97.8918
JT385	blood	3.9764	108.1	98.9	Hiseq	XT	v4	85057668	83362581	98.0071
JT386	blood	2.8289	94.4	98.9	Hiseq	XT	v4	74605022	73101219	97.9843
JT439	blood	1.3179	69.5	98.6	Hiseq	XT	v5	57144102	55899749	97.8224
JT440	blood	2.0791	86.2	98.6	Hiseq	XT	v5	68226938	66806953	97.9187
JT441	blood	2.8506	60.7	96.5	Hiseq	XT	v4	51057910	50369003	98.6507
JT442	blood	2.8107	57	95.3	Hiseq	XT	v4	51442802	50604420	98.3703
JT443	blood	1.6717	86.5	98.8	Hiseq	XT	v5	76613248	74813903	97.6514
JT445	blood	3.9172	78.2	98.9	Hiseq	XT	v5	71537052	69870697	97.6706
JT446	blood	1.6387	61.1	98.6	Hiseq	XT	v5	56562376	55166817	97.5327
JT447	blood	1.6253	77.4	98.8	Hiseq	XT	v5	75961880	74012166	97.4333
JT448	blood	1.6282	67.8	98.7	Hiseq	XT	v5	66111470	64390173	97.3964
JT474	saliva	4.4365	60.6	98.8	Hiseq	XT	v5	62965512	60245206	95.6797
JT476	saliva	4.2329	63.1	98.6	Hiseq	XT	v5	63793938	61012400	95.6398
JT538	blood	9.8618	50.3	96.6	Hiseq	QXT	v5	49437854	48671938	98.4507
JT539	blood	10.9173	61.55	97.5	Hiseq	QXT	v5	59262186	58301306	98.3786
JT540	blood	10.8186	55.27	97.2	Hiseq	QXT	v5	52317264	51461188	98.3637

JT541	blood	12.5238	63.87	96.6	Hiseq	QXT	v5	55944044	55072573	98.4422
JT542	blood	14.0225	69.56	98	Hiseq	QXT	v5	67251694	66211635	98.4535
JT543	blood	10.3385	54.56	97.3	Hiseq	QXT	v5	50636296	49669542	98.0908
JT576	blood	17.6766	78.56	98.4	Hiseq	QXT	v5	73769160	72104134	97.7429
JT577	blood	15.3413	64.22	98.2	Hiseq	QXT	v5	66440752	64892409	97.6696
JT578	blood	15.3273	63.98	97.9	Hiseq	QXT	v5	59183468	57848931	97.7451
JT579	blood	18.0831	73.97	98.4	Hiseq	QXT	v5	74044360	72332089	97.6875
JT584	blood	23.1056	96.63	98.7	Hiseq	QXT	v5	94390968	92277529	97.761
JT586	blood	18.3522	69.63	98.4	Hiseq	QXT	v5	67918044	65897629	97.0252
JT587	blood	18.6219	75.76	98.5	Hiseq	QXT	v5	74305842	72193713	97.1575
JT588	blood	17.1385	65.42	98.3	Hiseq	QXT	v5	64183642	62485997	97.355
JT609	fibroblasts	17.6582	75.35	98.4	Hiseq	QXT	v5	72844614	71182311	97.718
JT622	saliva	33.6593	81.53	98.6	Hiseq	QXT	v6	72403416	71386985	98.5962
JT623	saliva	29.7625	96.75	98.5	Hiseq	QXT	v6	77777666	77351173	99.4517
JT624	saliva	21.5174	64.22	98	Hiseq	QXT	v6	46340272	46113939	99.5116
JT625	saliva	29.8049	125.65	99.2	Hiseq	QXT	v6	98442630	98125155	99.6775
JT666	blood	30.3101	134.87	99.3	Hiseq	QXT	v6	105566342	105213234	99.6655
JT667	blood	31.1009	120.36	99.5	Hiseq	QXT	v6	95381764	94971101	99.5695
JT668	blood	29.8402	107.53	99.5	Hiseq	QXT	v6	83903400	83668754	99.7203
JT711	blood	27.5394	107.94	99.2	Hiseq	QXT	v6	85225814	84988938	99.7221
JT712	blood	29.7685	133.97	99.4	Hiseq	QXT	v6	103585682	103371379	99.7931
JT713	blood	29.8365	106.14	99.2	Hiseq	QXT	v6	85773756	84897237	98.9781
JT714	saliva	24.0348	107.62	99.1	Hiseq	QXT	v6	77470374	77261295	99.7301
JT720	blood	24.4923	105.47	99.4	Hiseq	QXT	v6	80002800	79647448	99.5558
JT721	blood	25.8227	121.84	99.5	Hiseq	QXT	v6	92706434	92434532	99.7067
JT722	blood	27.415	124.85	99.3	Hiseq	QXT	v6	95646452	95424422	99.7679
JT724	blood	24.8708	104.23	99.4	Hiseq	QXT	v6	75821082	75641506	99.7632
JT786	saliva	25.9555	119.85	99.4	Hiseq	QXT	v6	87447752	87188020	99.703
JT787	saliva	26.8696	99.83	99.3	Hiseq	QXT	v6	76653748	76088223	99.2622

JT788	blood	30.8861	117.04	99.6	Hiseq	QXT	v6	94706276	94482940	99.7642
-------	-------	---------	--------	------	-------	-----	----	----------	----------	---------

Appendix G

Known disease-causing genes assessed for variants in each recruited family.

Family	Associated disease	Gene
CHD1	Heterotaxy	<i>ZIC3, MMP21, CFC1, ACVR2B, PKD1L1, NODAL, CCDC11</i> , Plus genes associated with PCD*
CHD3	Familial aortic dilatation and dissection	<i>FBN1, FOXE3, TGFBR2, MYLK, LOX, TGFBR1, PRKG1, ACTA2, MFAP5, MYH11</i>
CHD4	Situs Inversus Totalis	<i>ZIC3, MMP21, CFC1, ACVR2B, PKD1L1, NODAL, CCDC11</i> , Plus genes associated with PCD*
CHD5	Dilated Cardiomyopathy	<i>PRDM16, NEXN, LMNA, TNNT2, PSEN2, ACTN2, TTN, DES, RAF1, SCN5A, TNNC1, SDHA, SGCD, LAMA4, PLN, EYA4, GATAD1, FKTN, MYPN, VCL, LDB3, RBM20, BAG3, CSRP3, MYBPC3, CRYAB, ABCC9, MYH6, MYH7, PSEN1, ACTC1, TPM1, DSG2, TNNI3, DMD</i> .
CHD6	Tetralogy of Fallot	<i>NKX2-5, GATA4, ZFPM2, GATA6, GDF1, JAG1, TBX1</i> .
CHD7	Truncus Arteriosus	<i>NKX2-5, NKX2-6, GATA6, GDF1, TBX1</i> .
CHD8	VSD/HLHS	<i>NKX2-5, GATA4, CITED2, GJA1</i>
CHD9, 10, 11	Athelia, choanal atresia	<i>KMT2D, PTPRM, TXNL4A, PTPN13, PTPRO, FOXE1, CHD7, TBX5, TBX3, LEF1, NRG3, ERBB4</i> .
CHD12	ASD	<i>TLL1, NKX2-5, CITED2, TBX20, GATA4, MYH6, ACTC1, GATA6</i> .
CHD13	3C- syndrome like	<i>KIAA0196, CCDC22</i> .
CHD15	AVSD	<i>GJA1, GATA4, GATA6</i>

CHD16	Tetralogy of fallot	<i>NKX2-5, GATA4, ZFPM2, GATA6, GDF1, JAG1, TBX1, MKKS.</i>
-------	---------------------	---

* Genes associated with PCD: *DRC1, ZMYND10, CCDC39, DNAH5, CCNO, RSPH9, RSPH4A, RSPH3, HEATR2, DNAH11, NME8, SPAG1, LRRC6, DNAI1, ARMC4, DNAJB13, CCDC65, KTU, DNAL1, DNAAF4, HYDIN, DNAAF1, GAS8, TTC25, CCDC103, DNAI2, CCDC40, CCDC151, CCDC114, DNAAF3, C21ORF59, RSPH1, PIH1D3.*

Appendix H

A list of heterozygous variant remaining following filtering in family CHD3 with a CADD score >15. PD= Probably Damaging, D= Deleterious, T= Tolerated, B= Benign, N= Neutral.

Gene	Variant	Protein change	Polyphen2	SIFT	Condel	CADD	MAF (GnomAD)	Protein function
<i>HIP1</i>	7:75168711C>T	ENSP00000336747.6, p.R998H	PD	D	D	34.0	10/246268	Actin binding
<i>HEXDC/ C17orf6 2</i>	17:80404519G>T	ENSP00000388909.2, p.Q61K	PD	D	D	25.0	33/246184	beta-N-acetylhexosaminidase activity/ integral component of membrane
<i>SLIT2</i>	4:20552506C>G	ENSP00000422591.1, p.S849C	PD	D	D	24.3	11/244776	GTPase inhibitor activity, slit2/ robo4 stabilises vascular network
<i>PSMB7</i>	9:127177647G>A	ENSP00000259457.3, p.R20C	B	T	N	16.22	10/246076	DNA damage response
<i>CCDC1 58</i>	4:77272933T>C	ENSP00000373566.2, p.I826M	PD	T	D	15.96	Not present	
<i>RCE1</i>	11:66610695C>G	ENSP00000354227.4:p.P67 5R	PD	D	D	15.62	Not present	processing of CAAX-type prenylated proteins.

Appendix I

Results from WGS analysis at low read depth to assess for CNVs in sample JT381 and JT443.

Sample JT381, calls from CNVseq

Location	Del/dup	Genes						
1-104134278-104313809	dup	AMY2A	AMY1A	AMY1B	AMY1C	RNPC3		
1-149041231-149279969	del	NBPF23						
1-149315068-149462199	del	FCGR1C	FAM72C					
3-129762478-129810356	del	TRH	ALG1L2					
4-69374731-69499460	del	TMPRSS11E	UGT2B17					
5-12515668-12744002	del	CT49						
6-258367-390083	del	DUSP22	IRF4					
8-7560068-7831482	dup	DEFB107A	SPAG11A					
8-12201691-12366392	dup	FAM66A	FAM86B2					
9-44711851-44864136	dup	none						
12-88325711-88477691	del	C12ORF50	C12ORF29	CEP290				
12-88494691-88566434	dup	TMTC3	cep290					
15-24439711-24548858	del	none						
17-34508311-34570589	del	CCL3L3	TBC1D3B	CCL4L1	TBC1D3C	TBC1D3H	CCL3L1	
17-34591351-34653629	del	TBC1D3CCCL4	TBC1D3H	CCL3L1				
19-20593054-20714875	del	ZNF626P						
22-24308696-24399168	del	GSTT2B	DDTL	DDT	GSTT2	GSTTP1	GSTT1	CABIN1

Sample JT381, calls from Read Depth

Chr	Start	Stop	Bins	Calculated copy number (2 is diploid)	Agree with CNVseq
chr1	148435101	149143500	4	0.43708931	
chr1	149143501	149194100	2	0.031369362	*

chr4	10195901	10246500	2	0.903815372	
chr4	34787501	34838100	2	1.058264312	
chr4	69482000	69499100	4	0.443405939	*
chr5	695751	721351	4	3.213621459	
chr5	12523501	12751200	9	1.009961291	*
chr7	38389800	38405400	4	1.31850222	
chr9	34559801	34661000	4	1.288506665	
chr9	135026101	135127300	4	1.307943299	
chr10	8070701	8146600	3	1.36983298	
chr10	126145801	126297600	6	1.368030125	
chr11	48297701	48341901	111	2.845929651	
chr11	48387700	48727801	111	2.845929651	
chr12	34787501	34814501	5	3.289116251	
chr12	88322301	88474100	6	1.066585781	*
chr12	93888301	93989500	4	2.706446744	
chr12	130902201	130952800	2	1.321613199	
chr15	24439801	24541000	4	0.612040913	*
chr17	41365501	41375601	4	2.823725797	
chr17	41414800	41466700	4	2.823725797	
chr17	47943501	48297700	14	1.356416275	
chr17	59505600	59581500	5	1.307399454	
chr18	77139701	77367400	9	1.317405976	
chr19	27954500	27956500	10	14.21754041	
chr21	22820601	22921800	4	1.349570179	
chr22	24313301	24313801	2	0.035275247	*
chr22	24382400	24389200	2	0.035275247	*

JT381 CNVnator calls that agree with CNVseq

CNV type	Location	Size	Read depth normalised to 1
duplication	chr1:104134001-104304000	170000	1.58337
deletion	chr1:149036001-149267000	231000	0.299106
duplication	chr1:149267001-149321000	54000	1.57572
deletion	chr1:149329001-149647000	318000	0.60251
deletion	chr3:129764001-129807000	43000	0.313157
deletion	chr4:69381001-69492000	111000	0.133311
deletion	chr5:12513001-12746000	233000	0.551357
deletion	chr6:1-202000	202000	0.492946
duplication	chr8:12231001-12462000	231000	2.04784
deletion	chr9:44452001-44727000	275000	0.396005
deletion	chr12:88324001-88482000	158000	0.546109
duplication	chr12:88482001-88566000	84000	1.55536
deletion	chr15:24442001-24542000	100000	0.224522
deletion	chr17:34510001-34550000	40000	0.287563
duplication	chr17:34576001-34592000	16000	2.27308
deletion	chr17:34592001-34651000	59000	0.315863
deletion	chr19:20503001-20718000	215000	0.419569
deletion	chr22:24308001-24398000	90000	0.281255

JT443 calls from CNVseq

Location	Del/dup	Genes			
1-196713091-196827120	del	<i>CFHR1</i>	<i>CFHR2</i>	<i>CFHR3</i>	<i>CFHR4</i>
6-162678378-162856400	del	<i>PARK2</i>			
7-8815943-8872405	del	<i>near</i>			
8-7232106-7469594	del	<i>SPAG11B</i>	<i>DEFB4B</i>		
8-7560068-7831482	del	<i>SPAG11A</i>	<i>DEFB4A</i>		
8-11955251-12106284	del	<i>DEFB130</i>	<i>USP17L7</i>		
8-39225268-39396877	del	<i>ADAM3A</i>			

13-58313491-58406893	del	none			
----------------------	-----	------	--	--	--

Sample JT443 calls from Read Depth

Chr	Start	Stop	Bins	Calculated copy number (2 is diploid)	Agree with CNVseq
chr1	842800	850200	3	2.855216216	
chr1	5689801	5708501	2	3.42332912	
chr1	16807801	16816801	5	4.339413301	
chr1	72757501	72762801	2	0.669149279	
chr1	72814000	72822900	2	0.669149279	
chr1	117131401	117139001	2	3.120765598	
chr1	117189600	117196800	2	3.120765598	
chr2	89565301	89640601	8	12.70478451	
chr4	52699900	52712400	2	26.51799879	
chr5	21483901	21486501	2	2.946086929	
chr5	46303201	46303801	7	8.580727685	
chr5	49563500	49573200	7	8.580727685	
chr6	29909600	29920500	2	1.094521277	
chr6	32438401	32449301	6	1.132883885	
chr6	32540900	32540901	6	1.132883885	
chr6	57599500	57617400	8	4.247753359	
chr7	38291701	38293501	3	1.175928221	
chr7	57519301	57524801	9	3.38674649	
chr7	62053600	62064600	13	10.2553157	
chr7	65073001	65334600	4	2.670481704	
chr7	100519801	100580151	2	3.741975207	
chr8	6801601	6899700	2	3.007365049	
chr8	46922400	46924500	2	12.13750513	

chr10	42832900	42837000	12	22.97511217	
chr10	127621700	127628100	2	5.310692512	
chr13	58336801	58402200	2	0.363442742	*
chr13	63601501	63602001	2	3.22992351	
chr13	63660800	63666900	2	3.22992351	
chr15	22674750	22677450	17	5.388560788	
chr16	32209501	32217501	12	3.680240377	
chr16	35185201	35191801	4	14.75058227	
chr17	45191401	45197601	2	6.570941053	
chr18	14126401	14715000	6	2.68328445	
chr19	24459601	24519601	3	4.07669493	
chr19	44897101	44903601	2	4.256005455	
chr20	29854000	29855100	11	7.425560869	

JT443 CNVnator calls that agree with CNVseq

CNV type	Location	Size	Read depth normalised to 1
deletion	chr6:162680001-162858000	178000	0.524078
deletion	chr8:7018001-7985000	967000	0.67404
deletion	chr8:11897001-12342000	445000	0.57844
deletion	chr8:39232001-39387000	155000	0.0328512
deletion	chr13:58321001-58395000	74000	0.049017

Appendix J

CNV calls from the Exome Depth program for all samples analysed. For brevity only CNVs absent from the Conrad database (Conrad et al., 2010) and with a Bayes Factor >10 are included. All regions were assessed for genes related to the phenotype.

Sample	Start	End	Dup/del	Position	BF	Reads expected	Reads observed	Reads ratio	Gene_exon
JT67	135080	135115	duplication	chr15:42052522-42116259	120	4113	5916	1.44	MGA_20,MGA_21,MGA_22,MGA_23,MGA_24,MAPKBP1_2,AC073657.1_2,AC073657.1_1,MAPKBP1_3,MAPKBP1_4,MAPKBP1_5,MAPKBP1_6,MAPKBP1_7,MAPKBP1_8,MAPKBP1_9,MAPKBP1_10,MAPKBP1_11,MAPKBP1_12,MAPKBP1_13,MAPKBP1_14,MAPKBP1_15,MAPKBP1_16,MAPKBP1_17,MAPKBP1_18,MAPKBP1_19,MAPKBP1_20,MAPKBP1_21,MAPKBP1_22,MAPKBP1_23,MAPKBP1_24,MAPKBP1_25,MAPKBP1_26,MAPKBP1_27,MAPKBP1_28,MAPKBP1_29,MAPKBP1_30
	71744	71749	deletion	chr7:5939916-5944900	28.2	357	162	0.454	CCZ1_2,CCZ1_3,CCZ1_4,CCZ1_5,CCZ1_6,CCZ1_7
	95380	95393	deletion	chr10:18060293-18152611	22	103	3	0.0291	TMEM236_2,TMEM236_3,TMEM236_4,MRC1_1,MRC1_2,MRC1_3,MRC1_4,MRC1_5,MRC1_6,MRC1_7,MRC1_8,MRC1_9,MRC1_10,MRC1_11
	127269	127278	duplication	chr13:114535273-114566893	20.2	519	756	1.46	GAS6_10,GAS6_9,GAS6_8,GAS6_7,GAS6_6,GAS6_5,GAS6_4,GAS6_3,GAS6_2,GAS6_1
	71936	71942	deletion	chr7:6826510-6840706	18	83	2	0.0241	RSPH10B2_16,RSPH10B2_17,RSPH10B2_18,RSPH10B2_19,RSPH10B2_20,CCZ1B_15,CCZ1B_14
	71926	71934	deletion	chr7:6805338-6821107	14.7	240	122	0.508	RSPH10B2_6,RSPH10B2_7,RSPH10B2_8,RSPH10B2_9,RSPH10B2_10,RSPH10B2_11,RSPH10B2_12,RSPH10B2_13,RSPH10B2_14
	71029	71032	deletion	chr6:167789455-167794824	12.8	161	77	0.478	TCP10_6,TCP10_5,TCP10_4,TCP10_3
	60875	60877	duplication	chr5:175386973-175388437	11.4	101	195	1.93	THOC3_6,THOC3_5,THOC3_4
	110121	110126	deletion	chr11:89768381-89774718	11.4	161	74	0.46	TRIM49C_3,TRIM49C_4,TRIM49C_5,TRIM49C_6,TRIM49C_7,TRIM49C_8
JT379	82439	82460	duplication	chr8:42939854-43054712	62.7	4226	5943	1.41	FNTA_8,FNTA_9,SGK196_4,SGK196_5,HGSNAT_1,HGSNAT_2,HGSNAT_3,HGSNAT_4,HGSNAT_5,HGSNAT_6,HGSNAT_7,HGSNAT_8,HGSNAT_9,HGSNAT_10,HGSNAT_11,HGSNAT_12,HGSNAT_13,HGSNAT_14,HGSNAT_15,HGSNAT_16,HGSNAT_17,

									HGSNAT_18
	71772	71779	deletion	chr7:5998625-6018327	49.4	685	287	0.419	RSPH10B_6,RSPH10B_5,RSPH10B_4,RSPH10B_3,RSPH10B_2,PMS2_15,PMS2_14,PMS2_13
	95350	95357	deletion	chr10:17865074-17898362	39.4	708	303	0.428	MRC1L1_2,MRC1L1_3,MRC1L1_4,MRC1L1_5,MRC1L1_6,MRC1L1_7,MRC1L1_8,MRC1L1_9
	95384	95391	duplication	chr10:18112045-18145281	34.9	150	476	3.17	MRC1_2,MRC1_3,MRC1_4,MRC1_5,MRC1_6,MRC1_7,MRC1_8,MRC1_9
	71941	71944	deletion	chr7:6838856-6844686	28.1	136	1	0.00735	CCZ1B_15,CCZ1B_14,CCZ1B_13,CCZ1B_12
	60875	60877	deletion	chr5:175386973-175388437	22.2	181	51	0.282	THOC3_6,THOC3_5,THOC3_4
	71029	71032	deletion	chr6:167789455-167794824	16.6	287	157	0.547	TCP10_6,TCP10_5,TCP10_4,TCP10_3
	96933	96937	duplication	chr10:51970584-51978390	15	196	345	1.76	ASAH2_11,ASAH2_10,ASAH2_9,ASAH2_8,ASAH2_7
	71932	71934	deletion	chr7:6819242-6821107	12.3	201	97	0.483	RSPH10B2_12,RSPH10B2_13,RSPH10B2_14
	151556	151561	deletion	chr17:16664740-16676947	11	1336	955	0.715	CCDC144A_13,CCDC144A_13,CCDC144A_14,CCDC144A_14,CCDC144A_15,CCDC144A_15,CCDC144A_16,CCDC144A_16,CCDC144A_17,CCDC144A_18,CCDC144A_17
	95948	95948	deletion	chr10:29580773-29580935	10.7	85	18	0.212	LYZL1_2
	39298	39299	deletion	chr3:57620286-57620496	10.4	100	34	0.34	DENND6A_14,DENND6A_13
	157450	157454	deletion	chr17:61972383-61973871	10.3	201	107	0.532	CSH1_5,CSH1_4,CSH1_3,CSH1_2,CSH1_1
JT384	119943	119977	deletion	chr12:88374109-88473006	119	2809	1582	0.563	C12orf50_13,C12orf50_12,C12orf50_11,C12orf50_10,C12orf50_9,C12orf50_8,C12orf50_7,C12orf50_6,C12orf50_5,C12orf50_4,C12orf50_3,C12orf50_2,C12orf29_1,C12orf29_2,C12orf29_3,C12orf29_4,C12orf29_5,C12orf29_6,C12orf29_7,CEP290_54,CEP290_53,CEP290_52,CEP290_51,CEP290_50,CEP290_49,CEP290_48,CEP290_47,CEP290_46,CEP290_45,CEP290_44,CEP290_43,CEP290_42,CEP290_41,CEP290_40,CEP290_39
	182072	182091	deletion	chr22:25115440-25158466	100	2266	1222	0.539	PIWIL3_21,PIWIL3_20,PIWIL3_19,PIWIL3_18,PIWIL3_17,PIWIL3_16,PIWIL3_15,PIWIL3_14,PIWIL3_13,PIWIL3_12,PIWIL3_11,PIWIL3_10,PIWIL3_9,PIWIL3_8,PIWIL3_7,PIWIL3_6,PIWIL3_5,PIWIL3_4,PIWIL3_3,PIWIL3_2
	119983	120020	duplication	chr12:88480169-88560359	95.6	3807	5308	1.39	CEP290_33,CEP290_32,CEP290_31,CEP290_30,CEP290_29,CEP290_28,CEP290_27,CEP290_26,CEP290_25,CEP290_24,CEP290_23,CEP290_22,CEP290_21,CEP290_20,CEP290_19,CEP290_18,CEP290_17,CEP290_16,CEP290_15,CEP290_14,CEP290_13,CEP290_12,CEP290_11,CEP290_10,CEP290_9,CEP290_8,CEP290_7,CEP290_6,CEP290_5,CEP290_4,CEP290_3,CEP290_2,TMTC3_2,TMTC3_3,TMTC3_4,TMTC3_5,TMTC3_6,TMTC3_7

	71922	71944	deletion	chr7:6797310-6844686	42.5	603	286	0.474	RSPH10B2_2,RSPH10B2_3,RSPH10B2_4,RSPH10B2_5,RSPH10B2_6,RSPH10B2_7,RSPH10B2_8,RSPH10B2_9,RSPH10B2_10,RSPH10B2_11,RSPH10B2_12,RSPH10B2_13,RSPH10B2_14,RSPH10B2_15,RSPH10B2_16,RSPH10B2_17,RSPH10B2_18,RSPH10B2_19,RSPH10B2_20,CCZ1B_15,CCZ1B_14,CCZ1B_13,CCZ1B_12
	6070	6087	duplication	chr1:46083136-46096278	40.9	2370	3151	1.33	NASP_14,NASP_15,CCDC17_13,CCDC17_12,CCDC17_11,CCDC17_10,CCDC17_9,CCDC17_8,CCDC17_7,CCDC17_6,CCDC17_5,CCDC17_4,CCDC17_3,CCDC17_2,CCDC17_1,GPBP1L1_13,GPBP1L1_12,GPBP1L1_11
	95384	95394	deletion	chr10:18112045-18155690	29.5	136	2	0.0147	MRC1_2,MRC1_3,MRC1_4,MRC1_5,MRC1_6,MRC1_7,MRC1_8,MRC1_9,MRC1_10,MRC1_11,MRC1_12
	71773	71780	duplication	chr7:6000323-6022622	27.4	286	547	1.91	RSPH10B_5,RSPH10B_4,RSPH10B_3,RSPH10B_2,PMS2_15,PMS2_14,PMS2_13,PMS2_12
	110109	110109	deletion	chr11:89531299-89531797	20.3	190	43	0.226	TRIM49_8
	96932	96937	deletion	chr10:51965743-51978390	18.4	160	58	0.362	ASAH2_12,ASAH2_11,ASAH2_10,ASAH2_9,ASAH2_8,ASAH2_7
	60876	60880	duplication	chr5:175388055-175395211	11.5	159	268	1.69	THOC3_5,THOC3_4,THOC3_3,THOC3_2,THOC3_1
JT443	71940	71944	duplication	chr7:6837995-6844686	35.4	120	409	3.41	RSPH10B2_20,CCZ1B_15,CCZ1B_14,CCZ1B_13,CCZ1B_12
	95384	95389	deletion	chr10:18112045-18138693	32.3	214	44	0.206	MRC1_2,MRC1_3,MRC1_4,MRC1_5,MRC1_6,MRC1_7
	93184	93194	duplication	chr9:135937382-135947151	25.7	1300	1789	1.38	CEL_1,CEL_2,CEL_3,CEL_4,CEL_5,CEL_6,CEL_7,CEL_8,CEL_9,CEL_10,CEL_11
	71932	71934	duplication	chr7:6819242-6821107	21	125	321	2.57	RSPH10B2_12,RSPH10B2_13,RSPH10B2_14
	96932	96937	deletion	chr10:51965743-51978390	19.5	179	68	0.38	ASAH2_12,ASAH2_11,ASAH2_10,ASAH2_9,ASAH2_8,ASAH2_7
	71027	71032	duplication	chr6:167786658-167794824	19.1	474	690	1.46	TCP10_8,TCP10_7,TCP10_6,TCP10_5,TCP10_4,TCP10_3
	151556	151560	duplication	chr17:16664740-16676023	18.5	763	1104	1.45	CCDC144A_13,CCDC144A_13,CCDC144A_14,CCDC144A_14,CCDC144A_15,CCDC144A_15,CCDC144A_16,CCDC144A_16,CCDC144A_17
	71754	71758	deletion	chr7:5959481-5966179	15.1	627	392	0.625	CCZ1_12,CCZ1_13,CCZ1_14,CCZ1_15,RSPH10B_20
	95350	95355	duplication	chr10:17865074-17891768	13.5	347	527	1.52	MRC1L1_2,MRC1L1_3,MRC1L1_4,MRC1L1_5,MRC1L1_6,MRC1L1_7
	71777	71780	duplication	chr7:6013031-6022622	12.5	83	183	2.2	PMS2_15,PMS2_14,PMS2_13,PMS2_12
	110121	110126	deletion	chr11:89768381-89774718	12.5	204	100	0.49	TRIM49C_3,TRIM49C_4,TRIM49C_5,TRIM49C_6,TRIM49C_7,TRIM49C_8
	61269	61271	deletion	chr5:177166602-177168333	12.1	65	8	0.123	FAM153A_8,FAM153A_7,FAM153A_6

JT448	60875	60880	duplication	chr5:175386973-175395211	26.2	158	360	2.28	THOC3_6,THOC3_5,THOC3_4,THOC3_3,THOC3_2,THOC3_1
	95377	95386	deletion	chr10:17951323-18122792	21.5	97	0	0	MRC1L1_29,MRC1L1_30,TMEM236_1,TMEM236_2,TMEM236_3,TMEM236_4,MRC1_1,MRC1_2,MRC1_3,MRC1_4
	142997	143006	duplication	chr16:18532088-18550286	19.4	408	636	1.56	NOMO2_19,NOMO2_18,NOMO2_17,NOMO2_16,NOMO2_15,NOMO2_14,NOMO2_13,NOMO2_12,NOMO2_11,NOMO2_10
	74165	74168	deletion	chr7:64838912-64864788	19.1	727	445	0.612	ZNF92_2,ZNF92_3,ZNF92_4
	71932	71941	deletion	chr7:6819242-6838910	16.6	227	114	0.502	RSPH10B2_12,RSPH10B2_13,RSPH10B2_14,RSPH10B2_15,RSPH10B2_16,RSPH10B2_17,RSPH10B2_18,RSPH10B2_19,RSPH10B2_20,CCZ1B_15
	71773	71779	duplication	chr7:6000323-6018327	15.7	168	306	1.82	RSPH10B_5,RSPH10B_4,RSPH10B_3,RSPH10B_2,PMS2_15,PMS2_14,PMS2_13
	88092	88099	duplication	chr9:35562378-35606282	15.4	610	838	1.37	FAM166B_5,FAM166B_4,FAM166B_3,FAM166B_2,FAM166B_1,TESK1_1,TESK1_2,TESK1_3
	96932	96937	deletion	chr10:51965743-51978390	15.2	142	55	0.387	ASAH2_12,ASAH2_11,ASAH2_10,ASAH2_9,ASAH2_8,ASAH2_7
	25115	25116	deletion	chr2:96652500-96652954	12.9	100	27	0.27	ANKRD36C_3,ANKRD36C_2
	75855	75857	deletion	chr7:98452864-98457961	12.9	273	152	0.557	TMEM130_5,TMEM130_4,TMEM130_3
	34150	34151	deletion	chr2:241631332-241633966	12.3	55	0	0	AQP12A_1,AQP12A_2,AQP12A_3
	95350	95355	duplication	chr10:17865074-17891768	12.2	328	485	1.48	MRC1L1_2,MRC1L1_3,MRC1L1_4,MRC1L1_5,MRC1L1_6,MRC1L1_7
	95389	95393	deletion	chr10:18138509-18152611	12.2	58	2	0.0345	MRC1_7,MRC1_8,MRC1_9,MRC1_10,MRC1_11
	71922	71925	deletion	chr7:6797310-6803732	12.1	248	130	0.524	RSPH10B2_2,RSPH10B2_3,RSPH10B2_4,RSPH10B2_5
	151556	151558	duplication	chr17:16664740-16667489	10.9	592	808	1.36	CCDC144A_13,CCDC144A_13,CCDC144A_14,CCDC144A_14,CCDC144A_15,CCDC144A_15
JT538	21847	21905	deletion	chr2:32701282-33037682	49.9	5347	4068	0.761	BIRC6_34,BIRC6_35,BIRC6_36,BIRC6_37,BIRC6_38,BIRC6_39,BIRC6_40,BIRC6_41,BIRC6_42,BIRC6_43,BIRC6_44,BIRC6_45,BIRC6_46,BIRC6_47,BIRC6_48,BIRC6_49,BIRC6_50,BIRC6_51,BIRC6_52,BIRC6_53,BIRC6_54,BIRC6_55,BIRC6_56,BIRC6_57,BIRC6_58,BIRC6_59,BIRC6_60,BIRC6_61,BIRC6_62,BIRC6_63,BIRC6_64,BIRC6_65,BIRC6_66,BIRC6_67,BIRC6_68,BIRC6_69,BIRC6_70,BIRC6_71,BIRC6_72,BIRC6_73,BIRC6_74,TTC27_1,TTC27_2,TTC27_3,TTC27_4,TTC27_5,TTC27_6,TTC27_7,TTC27_8,TTC27_9,TTC27_10,TTC27_11,TTC27_12,TTC27_13,TTC27_14,TTC27_15,TTC27_16,TTC27_17,TTC27_18
	151874	151885	deletion	chr17:18291534-18397681	38.2	241	56	0.232	EVPLL_10,LGALS9C_1,LGALS9C_2,LGALS9C_3,LGALS9C_4,LGALS9C_5,LGALS9C_6,LGALS9C_7,LGALS9C_8,LGALS9C_9,LGALS9C_10,LGALS9C_11

	94293	94303	deletion	chr9:140387466-140416168	37.4	868	520	0.599	PNPLA7_20,PNPLA7_19,PNPLA7_18,PNPLA7_17,PNPLA7_16,PNPLA7_15,PNPLA7_14,PNPLA7_13,PNPLA7_12,PNPLA7_11,PNPLA7_10
	71932	71944	deletion	chr7:6819242-6844686	33.6	197	37	0.188	RSPH10B2_12,RSPH10B2_13,RSPH10B2_14,RSPH10B2_15,RSPH10B2_16,RSPH10B2_17,RSPH10B2_18,RSPH10B2_19,RSPH10B2_20,CCZ1B_15,CCZ1B_14,CCZ1B_13,CCZ1B_12
	95384	95410	deletion	chr10:18112045-18196627	31.4	137	1	0.0073	MRC1_2,MRC1_3,MRC1_4,MRC1_5,MRC1_6,MRC1_7,MRC1_8,MRC1_9,MRC1_10,MRC1_11,MRC1_12,MRC1_13,MRC1_14,MRC1_15,MRC1_16,MRC1_17,MRC1_18,MRC1_19,MRC1_20,MRC1_21,MRC1_22,MRC1_23,MRC1_24,MRC1_25,MRC1_26,MRC1_27,MRC1_28
	163282	163292	deletion	chr19:863093-881714	23.8	552	347	0.629	CFD_5,MED16_16,MED16_15,MED16_14,MED16_13,MED16_12,MED16_11,MED16_10,MED16_9,MED16_8,MED16_7
	113222	113231	deletion	chr12:1949906-1984503	18	693	478	0.69	CACNA2D4_26,CACNA2D4_25,CACNA2D4_24,CACNA2D4_23,CACNA2D4_22,CACNA2D4_21,CACNA2D4_20,CACNA2D4_19,CACNA2D4_18,CACNA2D4_17
	96932	96937	deletion	chr10:51965743-51978390	15.2	130	50	0.385	ASAH2_12,ASAH2_11,ASAH2_10,ASAH2_9,ASAH2_8,ASAH2_7
	71922	71924	deletion	chr7:6797310-6799925	15.1	100	18	0.18	RSPH10B2_2,RSPH10B2_3,RSPH10B2_4
	159003	159008	deletion	chr17:74921049-74936930	15.1	518	352	0.68	MGAT5B_8,MGAT5B_9,MGAT5B_10,MGAT5B_11,MGAT5B_12,MGAT5B_13
	21827	21839	deletion	chr2:32660565-32690232	14.9	1321	993	0.752	BIRC6_14,BIRC6_15,BIRC6_16,BIRC6_17,BIRC6_18,BIRC6_19,BIRC6_20,BIRC6_21,BIRC6_22,BIRC6_23,BIRC6_24,BIRC6_25,BIRC6_26
	71775	71779	duplication	chr7:6005269-6018327	13.9	134	249	1.86	RSPH10B_3,RSPH10B_2,PMS2_15,PMS2_14,PMS2_13
	184997	185004	duplication	chr22:50943232-50945095	11.4	930	1154	1.24	LMF2_10,LMF2_9,LMF2_8,LMF2_7,LMF2_6,LMF2_5,LMF2_4,LMF2_3
	176320	176328	duplication	chr20:33583140-33587208	11.2	923	1135	1.23	MYH7B_26,MYH7B_27,MYH7B_28,MYH7B_29,MYH7B_30,MYH7B_31,MYH7B_32,MYH7B_33,MYH7B_34
	71754	71759	duplication	chr7:5959481-5968025	10.1	362	482	1.33	CCZ1_12,CCZ1_13,CCZ1_14,CCZ1_15,RSPH10B_20,RSPH10B_19
	154364	154369	duplication	chr17:39525656-39537464	10	663	835	1.26	KRT33B_1,KRT34_7,KRT34_6,KRT34_5,KRT34_4,KRT34_3
JT542	95384	95410	duplication	chr10:18112045-18196627	26.6	120	332	2.77	MRC1_2,MRC1_3,MRC1_4,MRC1_5,MRC1_6,MRC1_7,MRC1_8,MRC1_9,MRC1_10,MRC1_11,MRC1_12,MRC1_13,MRC1_14,MRC1_15,MRC1_16,MRC1_17,MRC1_18,MRC1_19,MRC1_20,MRC1_21,MRC1_22,MRC1_23,MRC1_24,MRC1_25,MRC1_26,MRC1_27,MRC1_28
	21849	21892	duplication	chr2:32703704-32875300	24.7	5435	6608	1.22	BIRC6_36,BIRC6_37,BIRC6_38,BIRC6_39,BIRC6_40,BIRC6_41,BIRC6_42,BIRC6_43,BIRC6_44,BIRC6_45,BIRC6_46,BIRC6_47,BIRC6_48,BIRC6_49,BIRC6_50,BIRC6_51,BIRC6_52,BIRC6_53,BIRC6_54,BIRC6_55,BIRC6_56,BIRC6_57,BIRC6_58,BIRC6_59,BIRC6_60,BIRC6_61,BIRC6_62,BIRC6_63,BIRC6_64,BIRC6_65

									BIRC6_66,BIRC6_67,BIRC6_68,BIRC6_69,BIRC6_70,BIRC6_71, BIRC6_72,BIRC6_73,BIRC6_74,TTC27_1,TTC27_2,TTC27_3,TT C27_4,TTC27_5
	96505	96531	duplication	chr10:47658500- 47758904	23.4	2609	3420	1.31	ANTXRL_1,ANTXRL_2,ANTXRL_3,ANTXRL_4,ANTXRL_5,ANTX RL_6,ANTXRL_7,ANTXRL_8,ANTXRL_9,ANTXRL_10,ANTXRL_1 1,ANTXRL_12,ANTXRL_13,ANTXRL_14,ANTXRL_15,ANTXRL_1 6,ANTXRL_17,ANXA8L2_1,ANXA8L2_2,ANXA8L2_3,ANXA8L2_4 ,ANXA8L2_5,ANXA8L2_6,ANXA8L2_7,ANXA8L2_8,ANXA8L2_9, ANXA8L2_10
	33392	33408	duplication	chr2:233243514- 233273111	18.8	2494	3138	1.26	ALPP_1,ALPP_2,ALPP_3,ALPP_4,ALPP_5,ALPP_6,ALPP_7,ALP P_8,ALPP_9,ALPP_10,ALPP_11,ALPPL2_1,ALPPL2_2,ALPPL2_ 3,ALPPL2_4,ALPPL2_5,ALPPL2_6
	95360	95377	deletion	chr10:17908572- 17951363	17.7	139	37	0.266	MRC1L1_12,MRC1L1_13,MRC1L1_14,MRC1L1_15,MRC1L1_16, MRC1L1_17,MRC1L1_18,MRC1L1_19,MRC1L1_20,MRC1L1_21, MRC1L1_22,MRC1L1_23,MRC1L1_24,MRC1L1_25,MRC1L1_26, MRC1L1_27,MRC1L1_28,MRC1L1_29
	71922	71943	duplication	chr7:6797310- 6841154	15.6	584	831	1.42	RSPH10B2_2,RSPH10B2_3,RSPH10B2_4,RSPH10B2_5,RSPH1 0B2_6,RSPH10B2_7,RSPH10B2_8,RSPH10B2_9,RSPH10B2_10 ,RSPH10B2_11,RSPH10B2_12,RSPH10B2_13,RSPH10B2_14,R SPH10B2_15,RSPH10B2_16,RSPH10B2_17,RSPH10B2_18,RSP H10B2_19,RSPH10B2_20,CCZ1B_15,CCZ1B_14,CCZ1B_13
	21827	21839	duplication	chr2:32660565- 32690232	12.5	1738	2200	1.27	BIRC6_14,BIRC6_15,BIRC6_16,BIRC6_17,BIRC6_18,BIRC6_19, BIRC6_20,BIRC6_21,BIRC6_22,BIRC6_23,BIRC6_24,BIRC6_25, BIRC6_26
	71776	71776	deletion	chr7:6006495- 6006747	10.9	62	0	0	RSPH10B_2
	94292	94299	duplication	chr9:140379040- 140400212	10.7	853	1163	1.36	PNPLA7_21,PNPLA7_20,PNPLA7_19,PNPLA7_18,PNPLA7_17,P NPLA7_16,PNPLA7_15,PNPLA7_14
	60875	60879	duplication	chr5:175386973- 175394270	10.3	266	407	1.53	THOC3_6,THOC3_5,THOC3_4,THOC3_3,THOC3_2
	183132	183146	duplication	chr22:36682765- 36697711	10	1748	2136	1.22	MYH9_35,MYH9_34,MYH9_33,MYH9_32,MYH9_31,MYH9_30,M YH9_29,MYH9_28,MYH9_27,MYH9_26,MYH9_25,MYH9_24,MY H9_23,MYH9_22,MYH9_21
JT587	127755	127824	duplication	chr14:22554730- 22970647	35.4	9099	11812	1.3	TRAV23DV6_1,TRAV23DV6_2,TRDV1_1,TRDV1_2,TRAV24_1,T RAV24_2,TRAV25_1,TRAV25_2,TRAV26-1_2,TRAV8- 7_1,TRAV8- 7_2,TRAV27_1,TRAV27_2,TRAV29DV5_1,TRAV29DV5_2,TRAV 30_1,TRAV30_2,TRAV26- 2_2,TRAV34_1,TRAV34_2,TRAV35_1,TRAV35_2,TRAV36DV7_1 ,TRAV36DV7_2,TRAV38-1_1,TRAV38-1_2,TRAV38- 2DV8_1,TRAV38- 2DV8_2,TRAV39_1,TRAV39_2,TRAV40_1,TRAV40_2,TRAV41_1 ,TRAV41_2,TRDV2_1,TRDV2_2,TRDJ1_1,TRDJ4_1,TRDJ2_1,TR DJ3_1,TRDC_1,TRDC_2,TRDC_3,TRDV3_2,TRAJ61_1,TRAJ59_ 1,TRAJ58_1,TRAJ57_1,TRAJ56_1,TRAJ54_1,TRAJ53_1,TRAJ52

									_1,TRAJ50_1,TRAJ49_1,TRAJ48_1,TRAJ47_1,TRAJ46_1,TRAJ45_1,TRAJ44_1,TRAJ43_1,TRAJ42_1,TRAJ41_1,TRAJ40_1,TRAJ39_1
	95380	95411	deletion	chr10:18060293-18198276	29.9	162	1	0.00617	TMEM236_2, TMEM236_3, TMEM236_4, MRC1_1, MRC1_2, MRC1_3, MRC1_4, MRC1_5, MRC1_6, MRC1_7, MRC1_8, MRC1_9, MRC1_10, MRC1_11, MRC1_12, MRC1_13, MRC1_14, MRC1_15, MRC1_16, MRC1_17, MRC1_18, MRC1_19, MRC1_20, MRC1_21, MRC1_22, MRC1_23, MRC1_24, MRC1_25, MRC1_26, MRC1_27, MRC1_28, MRC1_29
	50372	50396	deletion	chr4:104037735-104081936	19.7	1711	1110	0.649	CENPE_45, CENPE_44, CENPE_43, CENPE_42, CENPE_41, CENPE_40, CENPE_39, CENPE_38, CENPE_37, CENPE_36, CENPE_35, CENPE_34, CENPE_33, CENPE_32, CENPE_31, CENPE_30, CENPE_29, CENPE_28, CENPE_27, CENPE_26, CENPE_25, CENPE_24, CENPE_23, CENPE_22, CENPE_21
	76853	76862	deletion	chr7:102957278-102968201	13.9	732	421	0.575	DNAJC2_13, DNAJC2_12, DNAJC2_11, DNAJC2_10, DNAJC2_9, DNAJC2_8, DNAJC2_7, DNAJC2_6, DNAJC2_5, DNAJC2_4
	129092	129100	deletion	chr14:39760224-39777791	11.1	402	274	0.682	MIA2_9, CTAGE5_5, CTAGE5_6, MIA2_10, CTAGE5_7, MIA2_11, CTAGE5_8, MIA2_12, CTAGE5_9, MIA2_13, CTAGE5_10, MIA2_14, CTAGE5_11, MIA2_15, CTAGE5_12, MIA2_16, CTAGE5_13, MIA2_17
JT624	80563	80580	duplication	chr8:11968942-12287960	65.5	448	1485	3.31	ZNF705D_5, ZNF705D_6, USP17L2_1, FAM86B1_8, FAM86B1_7, FAM86B1_6, FAM86B1_5, FAM86B1_4, FAM86B1_3, FAM86B1_2, FAM86B1_1, DEFB130_2, DEFB130_1, FAM86B2_8, FAM86B2_7, FAM86B2_6, FAM86B2_5, FAM86B2_4
	171925	171936	duplication	chr19:49055511-49114104	42.3	1424	2144	1.51	SULT2B1_1, SULT2B1_2, SULT2B1_3, SULT2B1_4, SULT2B1_5, SULT2B1_6, SULT2B1_7, FAM83E_5, FAM83E_4, SPACA4_1, FAM83E_3, FAM83E_2
	133509	133528	duplication	chr15:20739498-21071610	29.5	268	528	1.97	GOLGA6L6_8, GOLGA6L6_7, GOLGA6L6_6, GOLGA6L6_5, GOLGA6L6_4, GOLGA6L6_3, GOLGA6L6_2, GOLGA6L6_1, AC012414.1_1, POTE_B_11, POTE_B_10, POTE_B_9, POTE_B_8, POTE_B_7, POTE_B_6, POTE_B_5, POTE_B_4, POTE_B_3, POTE_B_2, POTE_B_1
	51616	51626	deletion	chr4:144258343-144390342	29.3	529	271	0.512	GAB1_1, GAB1_2, GAB1_3, GAB1_4, GAB1_5, GAB1_6, GAB1_7, GAB1_8, GAB1_9, GAB1_10, GAB1_11
	134035	134039	deletion	chr15:30844282-30848215	25.1	149	22	0.148	GOLGA8Q_1, GOLGA8Q_2, GOLGA8Q_3, GOLGA8Q_4, GOLGA8Q_5
	102268	102270	duplication	chr10:135139332-135141519	22.4	330	604	1.83	CALY_5, CALY_4, CALY_3
	86336	86338	deletion	chr8:145317757-145318513	22.3	107	2	0.0187	KM-PA-2_11, KM-PA-2_10, KM-PA-2_9
	1868	1870	deletion	chr1:13716802-13743098	20.8	97	1	0.0103	PRAMEF17_2, PRAMEF17_3, PRAMEF20_2
	139030	139030	deletion	chr15:79065426-79065545	19.9	118	11	0.0932	ADAMTS7_14
	11168	11184	deletion	chr1:146399556-146420191	15.2	277	149	0.538	NBPF12_8, NBPF12_9, NBPF12_10, NBPF12_11, NBPF12_12, NBPF12_13, NBPF12_14, NBPF12_15, NBPF12_16, NBPF12_17, NBPF

									12_18,NBPF12_19,NBPF12_20,NBPF12_21,NBPF12_22,NBPF12_23,NBPF12_24
	86358	86360	duplication	chr8:145486843-145487599	14.5	180	336	1.87	BOP1_13,BOP1_12,BOP1_11
	1853	1855	duplication	chr1:13495968-13522259	14.2	79	191	2.42	PRAMEF16_2,PRAMEF16_3,PRAMEF21_2
	11149	11164	duplication	chr1:146232495-146253110	14.1	140	256	1.83	NBPF12_16,NBPF12_15,NBPF12_14,NBPF12_13,NBPF12_12,NBPF12_11,NBPF12_10,NBPF12_9,NBPF12_8,NBPF12_7,NBPF12_6,NBPF12_5,NBPF12_4,NBPF12_3,NBPF12_2,NBPF12_1
	46668	46672	deletion	chr4:9175517-9218723	12.8	69	7	0.101	FAM90A26P_5,FAM90A26P_6,FAM90A26P_7,USP17L13_1,USP17L10_1,USP17L11_1
	96454	96455	duplication	chr10:46312644-46322838	10.8	107	211	1.97	FAM25E_1,AGAP4_7
	169777	169778	deletion	chr19:39261667-39262670	10.7	49	0	0	LGALS7_4,LGALS7_3
	154326	154327	deletion	chr17:39215917-39222097	10.4	176	85	0.483	KRTAP2-3_1,KRTAP2-4_1
JT668	46667	46675	deletion	chr4:9174900-9237700	55.7	346	31	0.0896	FAM90A26P_4,FAM90A26P_5,FAM90A26P_6,FAM90A26P_7,USP17L13_1,USP17L10_1,USP17L11_1,USP17L12_1,USP17L13_2,USP17L15_1
	86325	86346	deletion	chr8:145314144-145322083	55.2	269	0	0	MROH1_36,MROH1_37,MROH1_38,MROH1_39,MROH1_40,MROH1_41,MROH1_42,MROH1_43,KM-PA-2_14,KM-PA-2_13,KM-PA-2_12,KM-PA-2_11,KM-PA-2_10,KM-PA-2_9,KM-PA-2_8,KM-PA-2_7,KM-PA-2_6,KM-PA-2_5,KM-PA-2_4,KM-PA-2_3,KM-PA-2_2,SCXB_1
	96508	96534	duplication	chr10:47667256-47769665	47.2	1987	2801	1.41	ANTXRL_4,ANTXRL_5,ANTXRL_6,ANTXRL_7,ANTXRL_8,ANTXRL_9,ANTXRL_10,ANTXRL_11,ANTXRL_12,ANTXRL_13,ANTXRL_14,ANTXRL_15,ANTXRL_16,ANTXRL_17,ANXA8L2_1,ANXA8L2_2,ANXA8L2_3,ANXA8L2_4,ANXA8L2_5,ANXA8L2_6,ANXA8L2_7,ANXA8L2_8,ANXA8L2_9,ANXA8L2_10,ANXA8L2_11,ANXA8L2_12,AL603965.1_2
	95384	95393	deletion	chr10:18112045-18152611	41	219	22	0.1	MRC1_2,MRC1_3,MRC1_4,MRC1_5,MRC1_6,MRC1_7,MRC1_8,MRC1_9,MRC1_10,MRC1_11
	71926	71951	duplication	chr7:6805338-6862991	34	746	1141	1.53	RSPH10B2_6,RSPH10B2_7,RSPH10B2_8,RSPH10B2_9,RSPH10B2_10,RSPH10B2_11,RSPH10B2_12,RSPH10B2_13,RSPH10B2_14,RSPH10B2_15,RSPH10B2_16,RSPH10B2_17,RSPH10B2_18,RSPH10B2_19,RSPH10B2_20,CCZ1B_15,CCZ1B_14,CCZ1B_13,CCZ1B_12,CCZ1B_11,CCZ1B_10,CCZ1B_9,CCZ1B_8,CCZ1B_7,CCZ1B_6,CCZ1B_5
	96932	96937	deletion	chr10:51965743-51978390	27.1	235	76	0.323	ASAH2_12,ASAH2_11,ASAH2_10,ASAH2_9,ASAH2_8,ASAH2_7
	95350	95355	duplication	chr10:17865074-17891768	24	272	521	1.92	MRC1L1_2,MRC1L1_3,MRC1L1_4,MRC1L1_5,MRC1L1_6,MRC1L1_7
	144262	144265	duplication	chr16:29057332-29062028	21.7	183	401	2.19	CTB-134H23.2_3,CTB-134H23.2_4,CTB-134H23.2_5,CTB-134H23.2_6

	80621	80621	duplication	chr8:13424502-13425460	19.3	347	883	2.54	C8orf48_1
	96454	96458	duplication	chr10:46312644-46329564	18.2	241	507	2.1	FAM25E_1,AGAP4_7,AGAP4_6,AGAP4_5,AGAP4_4
	88584	88586	duplication	chr9:43624656-43629320	16.5	1408	2201	1.56	SPATA31A6_4,SPATA31A6_3,SPATA31A6_2
	86358	86360	duplication	chr8:145486843-145487599	15	253	453	1.79	BOP1_13,BOP1_12,BOP1_11
	74581	74583	deletion	chr7:74191614-74193768	14.3	536	310	0.578	NCF1_2,NCF1_3,NCF1_4
	134202	134205	deletion	chr15:32889653-32890378	12.8	575	359	0.624	GOLGA8N_5,GOLGA8N_6,GOLGA8N_7,GOLGA8N_8
	48971	48973	deletion	chr4:75480840-75484724	12.1	72	11	0.153	AREGB_1,AREGB_2,AREGB_3
	25134	25135	deletion	chr2:96696155-96696596	12	86	20	0.233	GPAT2_7,GPAT2_6
	147618	147620	deletion	chr16:75239232-75239890	11.9	337	190	0.564	CTRB2_5,CTRB2_4,CTRB2_3
	133985	133986	deletion	chr15:30384912-30385191	11.4	195	91	0.467	GOLGA8J_16,GOLGA8J_17
	1868	1870	deletion	chr1:13716802-13743098	11.2	214	116	0.542	PRAMEF17_2,PRAMEF17_3,PRAMEF20_2
	182121	182123	deletion	chr22:25428577-25445331	11.2	54	2	0.037	KIAA1671_2,KIAA1671_3,KIAA1671_4
	1853	1855	duplication	chr1:13495968-13522259	10.6	118	217	1.84	PRAMEF16_2,PRAMEF16_3,PRAMEF21_2
	11148	11157	deletion	chr1:146231724-146244717	10.1	71	19	0.268	NBPF12_17,NBPF12_16,NBPF12_15,NBPF12_14,NBPF12_13,NBPF12_12,NBPF12_11,NBPF12_10,NBPF12_9,NBPF12_8
JT714	80564	80584	deletion	chr8:11970084-12338444	163	1300	270	0.208	ZNF705D_6,USP17L2_1,FAM86B1_8,FAM86B1_7,FAM86B1_6,FAM86B1_5,FAM86B1_4,FAM86B1_3,FAM86B1_2,FAM86B1_1,DEFB130_2,DEFB130_1,FAM86B2_8,FAM86B2_7,FAM86B2_6,FAM86B2_5,FAM86B2_4,FAM86B2_3,FAM86B2_2,FAM86B2_1,AC130352.1_5
	80366	80412	deletion	chr8:7220458-7809913	106	1804	934	0.518	DEFB4B_2,DEFB103B_2,SPAG11B_3,SPAG11B_2,SPAG11B_1,DEFB104B_2,DEFB106B_2,DEFB105B_1,DEFB105B_3,DEFB107B_1,DEFB107B_2,RP11-1118M6.1_1,RP11-1118M6.1_2,RP11-1118M6.1_3,RP11-1118M6.1_4,AC084121.16_4,AC084121.16_3,AC084121.16_2,AC084121.16_1,DEFB107A_2,DEFB107A_1,DEFB105A_3,DEFB105A_1,DEFB106A_2,DEFB104A_2,SPAG11A_1,SPAG11A_2,SPAG11A_3,DEFB103A_2,DEFB4A_2,ZNF705B_3,ZNF705B_4,ZNF705B_5,ZNF705B_6
	46667	46675	deletion	chr4:9174900-9237700	54	298	25	0.0839	FAM90A26P_4,FAM90A26P_5,FAM90A26P_6,FAM90A26P_7,USP17L13_1,USP17L10_1,USP17L11_1,USP17L12_1,USP17L13_2,USP17L15_1

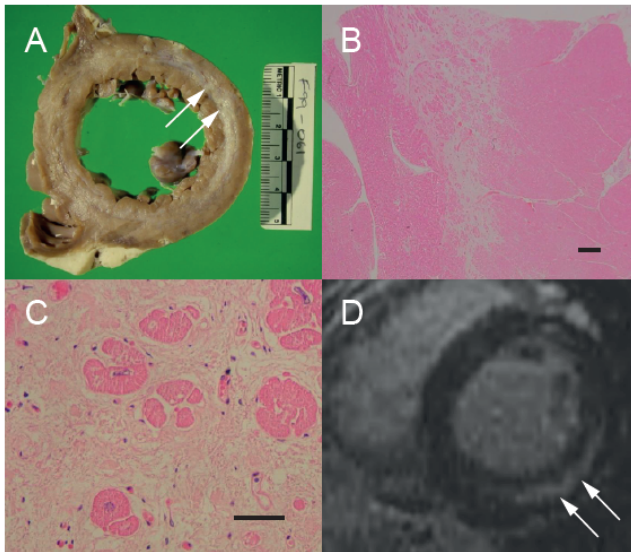
	144038	144052	deletion	chr16:28403074-28478223	45.9	463	161	0.348	EIF3CL_10,EIF3CL_9,EIF3CL_8,EIF3CL_7,EIF3CL_6,EIF3CL_5,EIF3CL_4,EIF3CL_3,EIF3CL_2,NPIPL1_7,NPIPL1_6,NPIPL1_5,NPIPL1_4,NPIPL1_3,NPIPL1_2
	101374	101385	duplication	chr10:121189888-121275164	44.2	1477	2130	1.44	GRK5_7,GRK5_8,GRK5_9,GRK5_10,GRK5_11,GRK5_12,GRK5_13,GRK5_14,GRK5_15,GRK5_16,RGS10_5,RGS10_4
	71941	71954	deletion	chr7:6838856-6864368	41.7	593	287	0.484	CCZ1B_15,CCZ1B_14,CCZ1B_13,CCZ1B_12,CCZ1B_11,CCZ1B_10,CCZ1B_9,CCZ1B_8,CCZ1B_7,CCZ1B_6,CCZ1B_5,CCZ1B_4,CCZ1B_3,CCZ1B_2
	79056	79061	deletion	chr7:142045364-142104553	37	1001	568	0.567	TRBV4-2_1,TRBV4-2_2,TRBV7-8_2,TRBV7-8_1,TRBV6-9_2,TRBV6-9_1
	134035	134040	duplication	chr15:30844282-30848525	26.4	265	503	1.9	GOLGA8Q_1,GOLGA8Q_2,GOLGA8Q_3,GOLGA8Q_4,GOLGA8Q_5,GOLGA8Q_6
	95386	95402	deletion	chr10:18122629-18169599	25.3	154	29	0.188	MRC1_4,MRC1_5,MRC1_6,MRC1_7,MRC1_8,MRC1_9,MRC1_10,MRC1_11,MRC1_12,MRC1_13,MRC1_14,MRC1_15,MRC1_16,MRC1_17,MRC1_18,MRC1_19,MRC1_20
	61269	61273	deletion	chr5:177166602-177171968	24.9	118	3	0.0254	FAM153A_8,FAM153A_7,FAM153A_6,FAM153A_5,FAM153A_4
	61260	61267	deletion	chr5:177156025-177164113	23.8	363	182	0.501	FAM153A_17,FAM153A_16,FAM153A_15,FAM153A_14,FAM153A_13,FAM153A_12,FAM153A_11,FAM153A_10
	96454	96458	deletion	chr10:46312644-46329564	21.8	309	137	0.443	FAM25E_1,AGAP4_7,AGAP4_6,AGAP4_5,AGAP4_4
	71743	71756	duplication	chr7:5938432-5963593	19.8	767	1061	1.38	CCZ1_1,CCZ1_2,CCZ1_3,CCZ1_4,CCZ1_5,CCZ1_6,CCZ1_7,CCZ1_8,CCZ1_9,CCZ1_10,CCZ1_11,CCZ1_12,CCZ1_13,CCZ1_14
	127793	127800	deletion	chr14:22907540-22932202	18.4	586	359	0.613	TRDD1_1,TRDD2_1,TRDD3_1,TRDJ1_1,TRDJ4_1,TRDJ2_1,TRDJ3_1,TRDC_1
	75449	75454	duplication	chr7:92019286-92028263	17.5	639	923	1.44	ANKIB1_15,ANKIB1_16,ANKIB1_17,ANKIB1_18,ANKIB1_19,ANKIB1_20
	5149	5149	deletion	chr1:39987049-39987242	16.4	240	99	0.412	BMP8A_4
	134027	134030	deletion	chr15:30701745-30702470	15.1	131	40	0.305	GOLGA8R_8,GOLGA8R_7,GOLGA8R_6,GOLGA8R_5
	124177	124178	duplication	chr13:24465429-24468340	14.6	890	1255	1.41	C1QTNF9B_3,C1QTNF9B_2
	173903	173903	deletion	chr19:56662315-56662350	14.3	139	45	0.324	AC024580.1_2
	27341	27342	deletion	chr2:132288153-132289375	13	609	395	0.649	CCDC74A_3,CCDC74A_4
	144130	144131	deletion	chr16:28763890-28769404	12.9	81	15	0.185	A-575C2.4_1,A-575C2.4_2
	149208	149209	deletion	chr17:3445817-3446922	12.8	232	117	0.504	TRPV3_6,TRPV3_5
	96933	96937	duplication	chr10:51970584-51978390	12.6	182	301	1.65	ASAH2_11,ASAH2_10,ASAH2_9,ASAH2_8,ASAH2_7
	48967	48971	deletion	chr4:75312252-	12.4	121	46	0.38	AREG_2,AREG_3,AREG_4,AREG_5,AREGB_1

				75480899					
	134158	134161	duplication	chr15:32684984-32685771	11.8	243	374	1.54	GOLGA8K_19,GOLGA8K_18,GOLGA8K_17,GOLGA8K_16
	103289	103291	deletion	chr11:5270599-5274635	11.5	150	69	0.46	HBG1_2,HBG1_1,HBG2_8
	27180	27181	duplication	chr2:130898732-130899954	11.2	681	950	1.4	CCDC74B_4,CCDC74B_3
	110120	110120	deletion	chr11:89608779-89609185	11.1	178	82	0.461	TRIM64B_1
	86336	86337	duplication	chr8:145317757-145318304	10.7	141	245	1.74	KM-PA-2_11,KM-PA-2_10
	11149	11164	deletion	chr1:146232495-146253110	10.2	246	150	0.61	NBPF12_16,NBPF12_15,NBPF12_14,NBPF12_13,NBPF12_12,NBPF12_11,NBPF12_10,NBPF12_9,NBPF12_8,NBPF12_7,NBPF12_6,NBPF12_5,NBPF12_4,NBPF12_3,NBPF12_2,NBPF12_1
	70084	70084	duplication	chr6:150343117-150343379	10.1	103	202	1.96	RAET1L_2
	73250	73255	deletion	chr7:38304951-38339764	10.1	477	313	0.656	TRGC1_1,TRGJ1_1,TRGJP_1,TRGJP1_1,TRGV11_1,TRGV10_1
	88544	88558	deletion	chr9:42368416-42410483	10.1	116	51	0.44	ANKRD20A2_1,ANKRD20A2_2,ANKRD20A2_3,ANKRD20A2_4,ANKRD20A2_5,ANKRD20A2_6,ANKRD20A2_7,ANKRD20A2_8,ANKRD20A2_9,ANKRD20A2_10,ANKRD20A2_11,ANKRD20A2_12,ANKRD20A2_13,ANKRD20A2_14,ANKRD20A2_15,RP11-146D12.2_1
JT788	151876	151886	deletion	chr17:18387190-18430117	111	855	71	0.083	LGALS9C_2,LGALS9C_3,LGALS9C_4,LGALS9C_5,LGALS9C_6,LGALS9C_7,LGALS9C_8,LGALS9C_9,LGALS9C_10,LGALS9C_11,FAM106A_1
	71921	71944	deletion	chr7:6744791-6844686	49.4	594	247	0.416	ZNF12_2,RSPH10B2_2,RSPH10B2_3,RSPH10B2_4,RSPH10B2_5,RSPH10B2_6,RSPH10B2_7,RSPH10B2_8,RSPH10B2_9,RSPH10B2_10,RSPH10B2_11,RSPH10B2_12,RSPH10B2_13,RSPH10B2_14,RSPH10B2_15,RSPH10B2_16,RSPH10B2_17,RSPH10B2_18,RSPH10B2_19,RSPH10B2_20,CCZ1B_15,CCZ1B_14,CCZ1B_13,CCZ1B_12
	46667	46675	deletion	chr4:9174900-9237700	47.7	432	45	0.104	FAM90A26P_4,FAM90A26P_5,FAM90A26P_6,FAM90A26P_7,USP17L13_1,USP17L10_1,USP17L11_1,USP17L12_1,USP17L13_2,USP17L15_1
	86325	86346	deletion	chr8:145314144-145322083	44.9	259	0	0	MROH1_36,MROH1_37,MROH1_38,MROH1_39,MROH1_40,MROH1_41,MROH1_42,MROH1_43,KM-PA-2_14,KM-PA-2_13,KM-PA-2_12,KM-PA-2_11,KM-PA-2_10,KM-PA-2_9,KM-PA-2_8,KM-PA-2_7,KM-PA-2_6,KM-PA-2_5,KM-PA-2_4,KM-PA-2_3,KM-PA-2_2,SCXB_1
	88658	88690	deletion	chr9:69200360-69424176	39.3	631	274	0.434	FOXD4L6_1,CBWD6_15,CBWD6_14,CBWD6_13,CBWD6_12,CBWD6_11,CBWD6_10,CBWD6_9,CBWD6_8,CBWD6_7,CBWD6_6,CBWD6_5,BX255923.1_1,CBWD6_4,CBWD6_3,CBWD6_2,CBWD6_1,CR769776.1_1,ANKRD20A4_1,ANKRD20A4_2,ANKRD20A

									4_3,ANKRD20A4_4,ANKRD20A4_5,ANKRD20A4_6,ANKRD20A4_7,ANKRD20A4_8,ANKRD20A4_9,ANKRD20A4_10,ANKRD20A4_11,ANKRD20A4_12,ANKRD20A4_13,ANKRD20A4_14,ANKRD20A4_15
	11127	11164	duplication	chr1:146215076-146253110	33.7	416	780	1.88	NBPF12_38,NBPF12_37,NBPF12_36,NBPF12_35,NBPF12_34,NBPF12_33,NBPF12_32,NBPF12_31,NBPF12_30,NBPF12_29,NBPF12_28,NBPF12_27,NBPF12_26,NBPF12_25,NBPF12_24,NBPF12_23,NBPF12_22,NBPF12_21,NBPF12_20,NBPF12_19,NBPF12_18,NBPF12_17,NBPF12_16,NBPF12_15,NBPF12_14,NBPF12_13,NBPF12_12,NBPF12_11,NBPF12_10,NBPF12_9,NBPF12_8,NBPF12_7,NBPF12_6,NBPF12_5,NBPF12_4,NBPF12_3,NBPF12_2,NBPF12_1
	61028	61050	duplication	chr5:176310772-176385155	32.9	3323	4555	1.37	HK3_15,HK3_14,HK3_13,HK3_12,HK3_11,HK3_10,HK3_9,HK3_8,HK3_7,HK3_6,HK3_5,HK3_4,HK3_3,HK3_2,UIMC1_15,UIMC1_14,UIMC1_13,UIMC1_12,UIMC1_11,UIMC1_10,UIMC1_9,UIMC1_8,UIMC1_7
	11169	11184	deletion	chr1:146400092-146420191	31.7	443	181	0.409	NBPF12_9,NBPF12_10,NBPF12_11,NBPF12_12,NBPF12_13,NBPF12_14,NBPF12_15,NBPF12_16,NBPF12_17,NBPF12_18,NBPF12_19,NBPF12_20,NBPF12_21,NBPF12_22,NBPF12_23,NBPF12_24
	96932	96937	deletion	chr10:51965743-51978390	26.9	282	105	0.372	ASAH2_12,ASAH2_11,ASAH2_10,ASAH2_9,ASAH2_8,ASAH2_7
	144262	144265	deletion	chr16:29057332-29062028	24.6	194	29	0.149	CTB-134H23.2_3,CTB-134H23.2_4,CTB-134H23.2_5,CTB-134H23.2_6
	95348	95352	deletion	chr10:17837921-17875838	23.4	158	10	0.0633	TMEM236_4,MRC1L1_1,MRC1L1_2,MRC1L1_3,MRC1L1_4
	142916	142928	deletion	chr16:16345821-16367932	22	1114	645	0.579	NOMO3_8,NOMO3_9,NOMO3_10,NOMO3_11,NOMO3_12,NOMO3_13,NOMO3_14,NOMO3_15,NOMO3_16,NOMO3_17,NOMO3_18,NOMO3_19,NOMO3_20
	139030	139030	deletion	chr15:79065426-79065545	20.5	204	12	0.0588	ADAMTS7_14
	71762	71779	duplication	chr7:5977390-6018327	15.4	725	1044	1.44	RSPH10B_16,RSPH10B_15,RSPH10B_14,RSPH10B_13,RSPH10B_12,RSPH10B_11,RSPH10B_10,RSPH10B_9,RSPH10B_8,RSPH10B_7,RSPH10B_6,RSPH10B_5,RSPH10B_4,RSPH10B_3,RSPH10B_2,PMS2_15,PMS2_14,PMS2_13
	25134	25135	deletion	chr2:96696155-96696596	15.1	89	5	0.0562	GPAT2_7,GPAT2_6
	147620	147628	deletion	chr16:75239812-75258086	14.9	1269	808	0.637	CTRB2_3,CTRB2_2,CTRB2_1,CTRB1_1,CTRB1_2,CTRB1_3,CTRB1_4,CTRB1_5,CTRB1_6
	25393	25407	duplication	chr2:97877284-97911808	14.8	333	540	1.62	ANKRD36_57,ANKRD36_58,ANKRD36_59,ANKRD36_60,ANKRD36_61,ANKRD36_62,ANKRD36_63,ANKRD36_64,ANKRD36_65,ANKRD36_66,ANKRD36_67,ANKRD36_68,ANKRD36_69,ANKRD36_70,ANKRD36_71
	103288	103291	duplication	chr11:5269590-5274635	14.5	189	448	2.37	HBG1_3,HBG1_2,HBG1_1,HBG2_8

	169777	169778	deletion	chr19:39261667-39262670	14.1	74	0	0	LGALS7_4, LGALS7_3
	71014	71015	deletion	chr6:167594122-167595404	12.3	107	20	0.187	TCP10L2_7, TCP10L2_8
	96454	96461	deletion	chr10:46312644-46342795	12	245	99	0.404	FAM25E_1, AGAP4_7, AGAP4_6, AGAP4_5, AGAP4_4, AGAP4_3, AGAP4_2, AGAP4_1
	95382	95394	deletion	chr10:18084908-18155690	11.4	269	314	1.17	TMEM236_4, MRC1_1, MRC1_2, MRC1_3, MRC1_4, MRC1_5, MRC1_6, MRC1_7, MRC1_8, MRC1_9, MRC1_10, MRC1_11, MRC1_12
	95382	95386	duplication	chr10:18084908-18122792	11.1	127	297	2.34	TMEM236_4, MRC1_1, MRC1_2, MRC1_3, MRC1_4
	166916	166917	deletion	chr19:14508471-14508747	11	110	28	0.255	CD97_7, CD97_8
	61010	61021	duplication	chr5:176289626-176305640	10.8	2329	3029	1.3	UNC5A_2, UNC5A_3, UNC5A_4, UNC5A_5, UNC5A_6, UNC5A_7, UNC5A_8, UNC5A_9, UNC5A_10, UNC5A_11, UNC5A_12, UNC5A_13
	27177	27178	deletion	chr2:130897465-130897669	10.6	595	279	0.469	CCDC74B_7, CCDC74B_6

Appendix K



A: Post-mortem section of the left ventricle from Patient P3, family 2b, showing circumferential scarring in the mid-myocardium. Arrows mark areas of fibrosis. B: Low-power microscopy of posterior wall of the left ventricle showing loose mid-myocardial fibrosis in patient P3, staining is eosin and haematoxylin. C: High-power microscopy of posterior wall of the left ventricle showing loose mid-myocardial fibrosis in patient P3, staining is eosin and haematoxylin. D: Cardiac MRI showing mid-myocardial fibrosis (arrows) in patient P4 at the age of 25 years. Adapted from Kennedy *et al.*, 2016.

Appendix M

Full coIP blots from Chapter 5, Figure 5-11.

Figure 5-11:A, RUVBL1 Figure 5-11:B, RUVBL2

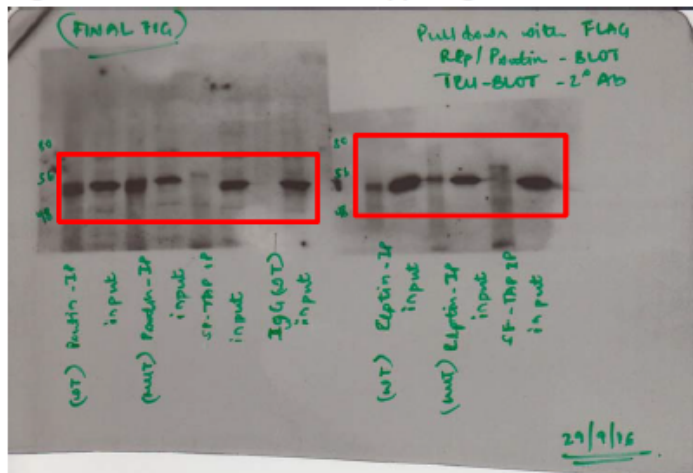


Figure 5-11:C, FLAG

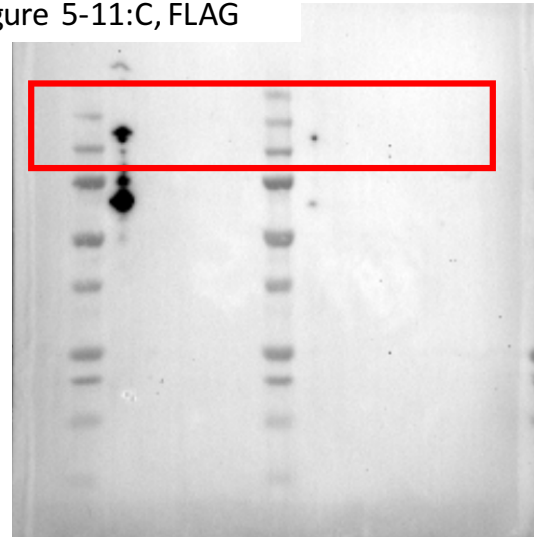
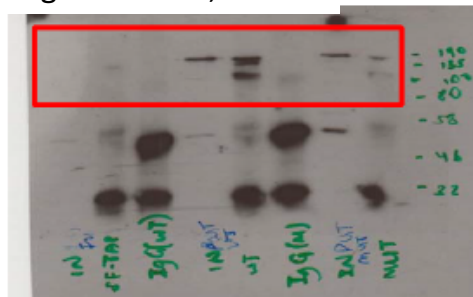


Figure 5-11:D, IFT88



Appendix N

The full list of variants remaining following filtering in family CHD10. PD= Probably Damaging, D= Deleterious, T= Tolerated.

Gene	Variant	Protein change	Polyphen2	SIFT	Condel	CADD	MAF (GnomAD)	Known protein function
<i>PYHIN</i>	1: 158913587TA>T	337fs	-	-	-	-	Not found	Transcription regulation, cell cycle control.
<i>STAB1</i>	3: 52549476TCTC>T	FS1302F	-	-	-	-	Not found	Transmembrane receptor, angiogenesis.
<i>ANKRD20A3</i>	9: 43121654TTAAG>T	259-260fs	-	-	-	-	862/8876	-
<i>CCDC73</i>	11: 32624416T>TGTC	1060-1061fs	-	-	-	-	1/243930	-
<i>MTCH2</i>	11: 47647293A>AAATC	227-228fs	-	-	-	-	4/30688	Mitochondrial inner membrane
<i>CARNS1</i>	11:67191606TCTTCAACGTGGAG>T	796-800fs	-	-	-	-	Not found	Carnasine synthase
<i>CARNS1</i>	11: 67191631C>CG	804-805fs	-	-	-	-	Not found	Carnasine synthase
<i>ATXN3</i>	14: 92537354C>CTG	305-306fs	-	-	-	-	Not found	ATPase binding, associated with Machado-Joseph disease, 109150.
<i>C4ORF36</i>	4: 87809325G>A	Q57*	-	-	-	27.4	Not found	-
<i>PTCH1</i>	9: 98240360 C>T, c.1324 C>T	V442M	PD (0.921)	D (0)	D (0.816)	26.2	Not found	Shh signalling
<i>CRB1</i>	1: 197398621G>A	D907N	PD (0.917)	T (0.1)	D (0.654)	25.1	Not found	Polarity in the eye, associated with RP and LCA.
<i>ITSN1</i>	21: 35206639G>C	W1127S	PD (0.94)	D (0)	D (0.830)	24.3	Not found	Endocytic membrane trafficking with the actin assembly machinery
<i>MTCH2</i>	11: 47660352G>T	H60N	PD (0.642)	D (0.02)	D (0.608)	23.3	3/11598	Mitochondrial inner membrane
<i>ADCK1</i>	14: 78365491T>C	W211R	PD (1)	D (0)	D (0.945)	23.2	Not found	ATP binding
<i>PTGIS</i>	20:48156152G>T	L210I	PD	T	D	22.4	Not found	Cytochrome P450 associated

			(0.726)	(0.06)	(0.587)			with essential Hypertension
<i>PLK1</i>	16: 23691408C>T	L138F	PD (0.995)	D (0.01)	D (0.863)	22.3	Not found	Centromere-located regulator
<i>APLF</i>	2: 68691455C>A	W118C	PD (0.999)	D (0)	D (0.935)	22.2	Not found	Response to DNA strand breaks
<i>ZMYND8</i>	20: 45853209T>G	E960A	PD (0.998)	D (0.01)	D (0.880)	19.74	Not found	RNA polymerase II transcription corepressor
<i>FRG1</i>	4: 190873347A>G	E55G	PD (0.466)	D (0.01)	D (0.558)	19.43	15/27280	Actin binding
<i>PDHX</i>	11: 35016621A>G	T470A	PD (0.955)	D (0.01)	D (0.804)	18.91	Not found	Regulation of acetyl-CoA biosynthetic process from pyruvate, associated with Lacticacidaemia (MIM 245349)
<i>CLUH</i>	17: 2601422G>A	R539C	PD (0.67)	D (0)	D (0.682)	18.84	3/202724	Intracellular distribution of mitochondria
<i>AGL</i>	1: 100379163G>C	D1344H	PD (0.717)	T (0.1)	D (0.564)	17.44	Not found	Glycogen debrancher enzyme. Associated with Glycogen storage disease IIIa, 232400 and Glycogen storage disease IIIb, 232400.
<i>KRT18</i>	12: 53343318C>T	P121S	B (0.438)	D (0.03)	D (0.507)	16.34	1004/12016	Type I intermediate filament chain keratin 18