Identifying Novel Genes that cause Congenital Heart Disease

Anna Wilsdon; B Med Sci, BM BS

School of Life Sciences

University of Nottingham

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy April 2021

Main supervisor – Prof. J David Brook

Declaration

I hereby declare that this thesis is the result of my own work, which has been undertaken during my period of registration for the PhD degree at The University of Nottingham, and that it has not been presented to this or any other university for the award for any degree or diploma other than that for which I am a candidate.

Acknowledgements

Contents

Index of Figures

Index of Tables

Abstract

Complex genetic networks underlie the development of the heart. Variants in some of these genes can lead to congenital heart disease (CHD). Despite the hundreds of genes that have been identified as causing CHD in humans, they only account for a small proportion of individuals with CHD. We carried out whole exome sequencing in the largest cohort of individuals with CHD reported at the time of publication, and identified three novel genome wide significant syndromic CHD genes; *CDK13*, *PRKD1* and *CHD4*.

Individuals with mutations in *CHD4* show neurodevelopmental disability, genital abnormalities and share some phenotypic overlap with other chromatinopathies. Mutations in *PRKD1* also cause syndromic CHD, but identification of further affected individuals is required to determine if there is a consistent phenotype. *Prkd1* mouse models do not have a high incidence of CHD, but this gene may be important in future work as it plays a role in cardiac hypertrophy.

Individuals with mutations in *CDK13* show a recognisable phenotype and the mouse model shows embryonic lethality and atrioventricular canal defects. The precise mechanism by which heterozygous mutations cause disease in humans remains unclear. The phenotypic and genotypic spectrum has been expanded by subsequent reports of individuals with mutations in *CKD13*.

We also identify a role for inherited variants with reduced penetrance in individuals with non-syndromic CHD. This is significant, as the vast majority of individuals with CHD have non-syndromic CHD. It is an important step in understanding the potential oligogenic pathogenesis of the majority of CHD.

Ultimately we aim to increase our knowledge of the genes and networks that underlie CHD, to improve diagnostic yield in individuals affected with CHD. I was able to feedback pathogenic mutations in CHD genes to participants in this study locally. Following this unbiased approach of non-targeted testing in a cohort with multiple types of CHD, will improve our knowledge of genotype phenotype correlations. We also hope that understanding more about the genes involved in cardiogenesis and CHD might have relevance for the failing heart, and development of treatments in the future too.

List of Abbreviations

a.a.	Amino acid
AD	Autosomal dominant
ADHD	Attention Deficit Hyperactivity Disorder
AF	Atrial fibrillation
Ao	Aorta
Array-CGH	Array Comparative Genomic Hybridisation
AS	Aortic Valve Stenosis
ASD	Atrial Septal Defect
AV	Aortic Valve or atrioventricular
AVCD	Atrioventricular canal defect
AVN	Atrioventricular Node
AVSD	Atrioventricular Septal Defect
BAV	Bicuspid Aortic Valve
BSU	Biosupport Unit
bp	base pair(s)
CCA	Common Carotid Artery
CDK	Cyclin dependant kinase
CHD	Congenital heart defect/disease or Chromodomain-Helicase-DNA binding gene/protein
CHDFIDD	CHD, dysmorphic facial features, and intellectual developmental disorder
CNV	Copy number variant/variation
CoA	Coarctation of the aorta
CRD	Cysteine rich domain
CRL	Crown Rump Length
CT	Computed tomography
CTD	C terminal domain
DA	Ductus Ateriosus

DAG	Diacylglycerol
DCM	Dilated Cardiomyopathy
DDD	Deciphering Developmental Disorders Project
Deln/Dels	Deletion
DEXA	Dual-energy X-ray Absorptiometry
DGV	Database of Genomic Variation
DMDD	Deciphering Mechanisms of Developmental Disorders Study
DNM	De novo mutation
DORV	Double Outlet Right Ventricle
Dpf	Days post fertilisation
EA	Ebstein anomaly
ECG	Electrocardiography/electrocardiogram
Echo	Echocardiography/echocardiogram
EFIC	Episcopic fluorescence image capture
EMT	Epithelial (endothelial) to mesenchymal transition
ESCC	European Society of Cardiology Codes
FDR	False discovery rate
FHF	First heart field
FISH	Fluorescent In Situ Hybridisation
GWAS	Genome Wide Association Study
HCM	Hypertrophic cardiomyopathy
HDAC	Histone deacetylase
Het	Heterozygous/heterozygote
HGMD	Human Genome Mutation Database
HLHS	Hypoplastic Left Heart Syndrome
Hom	Homozygous/homozygote
Hpf	Hours post fertilisation
HPO	Human Phenotype Ontology
HREM	High resolution episcopic microscopy
IAA	Interrupted aortic arch

IAVCD	Intermediate atrioventricular canal defect
IBD	Identity by Descent
ID	Intellectual Disability
INTERVAL	Interval study, to determine safe frequency of blood donation
IPCC	International Society for Nomenclature of Paediatric and Congenital Heart Disease
IVC	Inferior Vena Cava
IVS	Intact/inter ventricular septum or intact ventricular septum in the context of TGA
IUGR	Intrauterine growth restriction/retardation
LA	Left atrium
LoF	Loss of Function
LV	Left ventricle
LVNC	Left ventricular non-compaction
LVOT	Left Ventricular Outflow Tract
LVOTO	Left Ventricular Outflow Tract Obstruction
MA	Mitral Valve Atresia
MAF	Mean allele frequency
MiRNA	Micro RNA
MR	Mitral Regurgitation
MRI	Magnetic resonance imaging
MS	Mitral Stenosis
MV	Mitral valve
MVP	Mitral valve prolapse
NDD	Neurodevelopmental disability
NS-CHD	Non syndromic CHD
NTD	N terminal domain
NTP	Nucleotide triphosphates
OCT	Optical computed tomography
OFC	Occipitofrontal circumference

OFT	Outflow tract
OPT	Optical projection tomography
OR	Odds ratio
PA	Pulmonary valvar atresia
PAPVD	Partial anomalous pulmonary venous drainage
PBS	Phosphate Buffered Solution
PC	Post coitum
PCA	Principle component analysis
PDA	Patent ductus arteriosus
PFA	Paraformaldehyde
PFO	Patent Foramen Ovale
PH	Pleckstrin homology domain
PHT	Primary Heart Tube
Prkd1Em1	*Prkd1*em1(IMPC)Wtsi mouse
Prkd1Em2	*Prkd1*em2(IMPC)Wtsi mouse
PS	Pulmonary Valve Stenosis
PTV	Protein truncating variant
PV	Pulmonary Valve
RA	Right atrium
RR	Relative risk
RV	Right ventricle
RVOT	Right ventricular outflow tract
RVOTO	Right ventricular outflow tract obstruction
S-CHD	Syndromic CHD
SAN	Sinoatrial Node
SCL	Subclavian
SDW	Sterile Distilled Water
Ser	Serine
Sg/sgRNA	Single guide (site)
SHF	Second Heart Field

SIT	Situs inversus totalis
SNHL	Sensorineural hearing loss
SnoRNA	Small nucleolar RNAs
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SVC	Superior Vena Cava
TA	Truncus arteriosus
TADA	Transmission and De novo Association
TAPVD/R	Total Anomalous Pulmonary Venous Drainage/return
TE	Time to echo (MRI)
TGA	Transposition of the Great Arteries
TOE	Transoesophageal Echocardiography
TOF	Tetralogy of Fallot
TR	Tricuspid Regurgitation
TR^{MRI}	Repetition time in MRI
TTE	Transthoracic Echocardiography
TV	Tricuspid Valve
VEP	Variant effect predictor (Ensembl resource)
VSD	Ventricular septal defect
VUS	Variant of uncertain Significance
WES	Whole exome sequencing
WGS	Whole Genome Sequencing
WPW	Wolff Parkinson White
WT	Wild Type
WTSI	Wellcome Trust Sanger Institute
XLD	X-linked dominant

Chapter 1 Introduction

1.1 Overview of Congenital Heart Disease

Congenital Heart Disease (or Congenital Heart Disorders/defects, CHD) are abnormalities in the structure of the heart or great vessels that are present at birth. These terms encompass a wide range of defects that may be mild and require no intervention, or may be life-limiting and require complex surgery soon after birth.

CHD is important for two reasons. It is the most common congenital abnormality, and has a higher mortality that many other birth defects [1, 2]. It affects 7-9 babies per 1000 births [1, 3]. If you include milder lesions such as bicuspid aortic valve (BAV), the incidence rises to 2-3% [4]. Prevalence has been estimated at 13 per 1000 children, and 6 per 1000 adults [5]. Almost 60% of individuals with CHD have three common subtypes; atrial septal defect (ASD), ventricular septal defect (VSD) or patent ductus arteriosus (PDA) [6]. Tetralogy of Fallot (TOF) is the most common cyanotic form of CHD, and affects approximately 1 in every 2500-3000 live births [1, 7].

Given that CHD is a common congenital abnormality, it is an important issue for healthcare providers, and the potentially complex needs of this growing group of individuals must be considered. Individuals with CHD are living longer due to better medical and surgical care, with 95% reaching adulthood [8]. The prevalence of both milder and more complex lesions has risen by around 10% every 5 years since 1970 [6]. As a result, individuals with CHD now represent a considerable cohort of individuals, some of whom require lifelong specialist medical care. Their needs range from complicated neonatal and paediatric surgery, to lifelong surveillance to monitor for late complications, as some types of CHD are associated with high morbidity and mortality in adulthood [9]. In addition, given that individuals are now more likely to survive to reproductive

age, they may ask about the risks of their children also being affected with CHD. Provision for genetic counselling and testing should be provided [10].

There is a significant genetic component and strong heritability to CHD. The classical pedigree displaying Mendelian inheritance is uncommon, and in the majority of cases there appears to be a complex pattern of inheritance. It has become apparent that the mechanisms underlying CHD are complicated, and that perturbation of gene expression contributes towards abnormality in the developing heart. Transcription factors such as *NKX2.5* and the genes they regulate, form an important part of these networks. Aberrations in copy number, and mutations in many of these genes have been identified as causing CHD [11-18]. However a single mutation in a specific gene can produce different CHD phenotypes and there is now mounting evidence of reduced penetrance with seemingly deleterious mutations in CHD causing genes [12, 19]. This and the fact that CHD is usually only seen in a single individual within a family, suggests that monogenic causes are rare and there are likely to be additional genetic and non-genetic factors which contribute.

Whilst a number of genes have been identified previously by investigation of families with a Mendelian pattern of CHD, sequencing of large cohorts of affected individuals is now at the forefront of research to discover novel CHD genes [12, 14, 15, 19]. This method has been successful in identifying new CHD genes. Close collaboration between clinicians and research groups will be required to obtain large enough cohorts to fully understand the genetic basis of this complex group of conditions.

Research in the Brook laboratory has focussed on identifying genes essential in heart embryogenesis that cause CHD. This thesis aims to identify further genes that cause CHD, provide an animal model for support, and consider the mechanism by which variants are deleterious. Identification of a genetic cause of an individual's CHD allows accurate genetic counselling, may guide prognosis,

medical management and antenatal care. Therefore it is important that we try to understand the genetic aberrations that cause CHD.

The heart has no real regenerative power, so damage is cumulative and can have long lasting consequences despite remodelling. Understanding the pathways involved in successful cardiac development will help us understand mechanisms behind CHD, and may also provide opportunities to develop treatments for the damaged heart.

1.2 Heart development

CHD results from perturbations in the normal development of the heart. A specific structure may fail to develop, or form abnormally due to altered haemodynamics within the heart and great vessels. Abnormalities include defects of laterality, septation and formation of specific structures such as the valves. The consequences depend on the structure affected, and may differ between the fetal and postnatal circulation states. Given that the heart is the first fully functioning organ in embryogenesis, severe cardiac abnormalities can result in embryonic lethality.

This thesis focusses on identifying genetic causes of CHD in humans, but also involves the use of both zebrafish and mice. The early stages of development of the heart follows a similar path in all three species, with formation of a heart tube and subsequent looping. Both the mouse and zebrafish have been used as model organisms in cardiac research extensively in the past. Understanding the normal development of the human heart can help us understand how and why CHD develops. Cardiogenesis in the human is discussed in more detail below. This is followed by a discussion of the links between genes known to be important in heart development, and those that have subsequently been found to play a role in CHD.

1.2.1 Human Heart Development

Embryogenesis begins with formation of the zygote, after fusion of the ova and sperm. Active circulation of fetal blood is required relatively early on in gestation, so that oxygen can be delivered to the growing embryo. Diffusion is adequate initially, but cannot maintain fetal growth and development. The heart by necessity, is there for the first organ to develop.

The heart develops from a number of different cell populations or heart fields. They contribute to formation of a single tube initially, which then loops to form the chambers of the heart. Looping of the heart tube is the first evidence of the left-right axis in the body. Formation of the heart tube and its subsequent development into the four chambered heart and outflow tract is detailed in section 1.2.1.2. The mature heart is formed by day 50 of gestation [20].

1.2.1.1 The heart fields

The heart is derived from four main sources; the primary and secondary heart fields (both from cardiac mesenchymal progenitor cell populations), the cardiac neural crest, and the proepicardium (Figure 1.1).
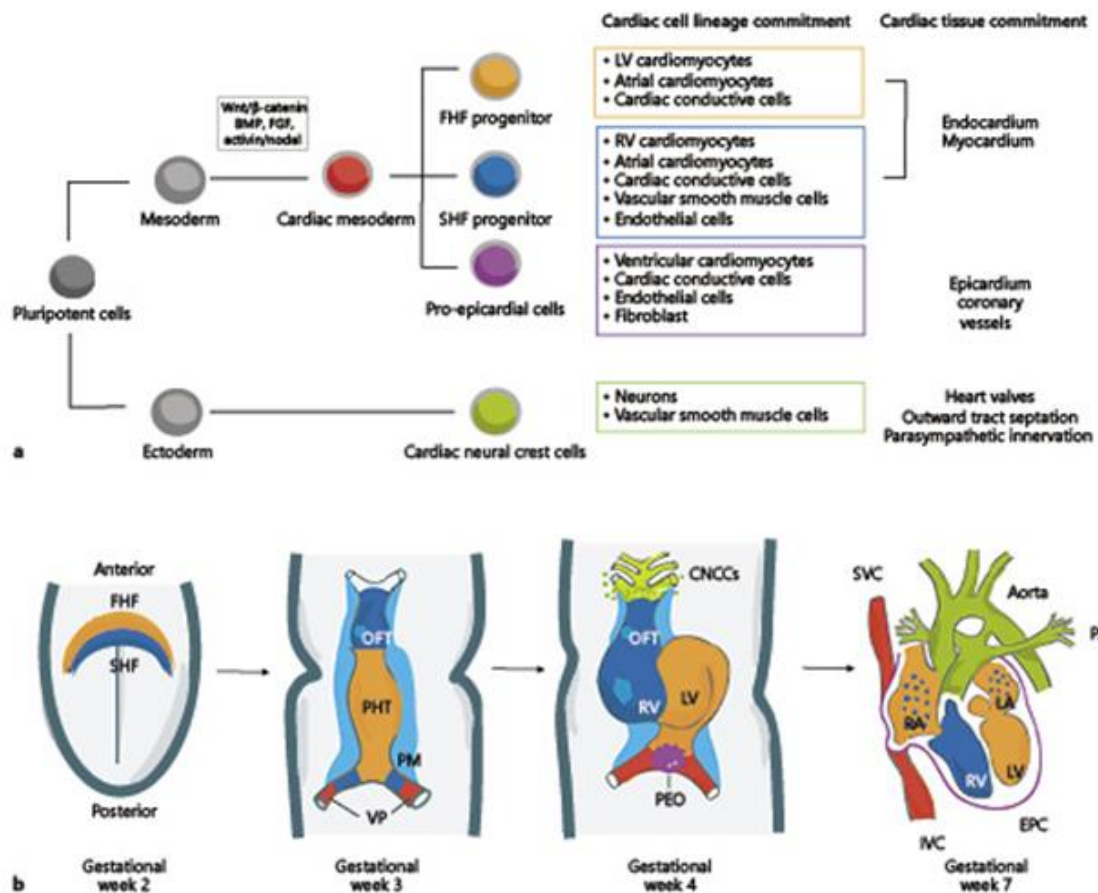
Figure 1.1 Early cardiac development.

Reproduced from Tan et al. [21], with permission from Karger Basel. Early cardiac development. **a** Cardiac cell lineage and specification during development demonstrating the commitment of pluripotent cells toward mature cardiac cell types within the heart development. **b** Schematic of cardiac morphogenesis in humans. At the second week of gestation, the cardiogenic mesodermal cells migrate toward the anterior side of the embryo to form the first heart field (FHF) or cardiac crescent and second heart field (SHF) that are specified to form specific segments of the primary heart tube (PHT), which is patterned along the anteroposterior axis to form the various regions and chambers of the looped and mature heart during weeks 3 and 4. The FHF gives rise to the beating PHT and will eventually give rise to the left ventricle (LV) and parts of the right and left atria (RA and LA, respectively). The SHF, located behind the PHT and within the pharyngeal mesoderm by gestational week 3, will contribute to the formation of the RV, parts of the atria and outflow tract, and later to the base of the aorta and pulmonary artery. At gestational week 3, the

cells at the venous pole contribute to the formation of the superior and inferior vena cavas (SVC and IVC, respectively). By gestational week 4, the cardiac neural crest cells migrate in from the dorsal neural tube, forming smooth muscle cells within the aortic and pulmonary arteries. In addition, the proepicardial organ formed by the proepicardial progenitor cell clusters later contributes to the formation of the epicardium. The 4 chambers form by the end of week 7. Wnt, Wingless integrated; FGF, fibroblast growth factor; BMP, bone morphogenetic proteins; OFT, outflow tract; PM, pharyngeal mesoderm; VP, venous pole; CNCCs, cardiac neural crest cells; LV, left ventricle; RV, right ventricle; PEO, proepicardial organ; EPC, epicardium; and PA, pulmonary artery.

Initial development centres on formation of the heart tube. Subsequent migration of different cell populations to the arterial (cranial) pole, and the venous (caudal) pole, cause elongation of the tube and formation of the structures within the heart.

- The primary (or first) heart field produces the initial heart tube, and later primarily the left ventricle (LV) [22]. It also contributes to the atria, part of the right ventricle (RV), tricuspid (TV) and mitral valves (MV).

- The second heart field (SHF) produces the outflow tract, and is split into anterior and posterior components. It contributes to all chambers with the exception of the LV [23].
  - The anterior SHF moves to the arterial pole of the heart. It contributes to the right ventricle (RV) and outflow tract (OFT, conus cordis and truncus arteriosus). The mesoderm extends to form the myocardium of the RV, and the right side of the interventricular septum. The SHF also forms part of the aortic and pulmonary (semilunar) valves (AV, PV) and the walls of the great arteries [24, 25]. Cells in the aorta are added from the right, and cells in the pulmonary trunk are added from the left which may contribute to the crossing over of the aorta and pulmonary trunk [26].
  - The posterior SHF reaches the venous pole of the heart. It contributes to the atrioventricular canal, the atria, the dorsal

mesenchymal protrusion, the central chamber of the sinus venosus and the outflow tract [22, 27]. The myocardium derived from these cells covers the sinus venosus and forms part of the dorsal wall of the right and left atrium [24, 28, 29].

- The cardiac neural crest is derived from ectodermal lineage, from the neural tube cells near the first three somites. This is in contrast to the rest of the heart that is derived from mesoderm [30]. It contributes structurally to the outflow tract septum, the endocardial cushions/valves and the parasympathetic nervous system [31, 32]. These cells migrate from the dorsal neural tube, through the posterior pharyngeal arches to the anterior part of the heart tube [31, 32]. The cells then migrate to the outflow tract between weeks three and four, and are involved in separation of the aorta and pulmonary trunk. They produce myocardium for the outflow tract and a small outflow tract septum with the endocardial cushions.

- The proepicardium is a transient tissue which derives from the proepicardial organ. It is a collection of epithelial cells, with a core of mesenchymal cells, within an extracellular matrix in the dorsal mesoderm of the septum transversum in mammals [22, 33, 34]. It contributes to the epicardium, fibroblasts in the myocardium and the coronary arteries. Proepicardial cells are produced from a specific population of cells in the splanchnic mesoderm. The cells migrate to the primary heart tube just as it begins looping. The cells attach to the AV junction and migrate over the myocardium to form the epicardium [33]. A few of the cells undergo EMT (epithelial to mesenchymal transition) and with the other mesenchymal cells of the proepicardium, form the smooth muscle cells, perivascular fibroblasts and endothelial cells of the coronary arteries [35-37]. There is also some contribution to the muscular ventricular septum and the atria by a small number of proepicardial cell precursors [38, 39].

1.2.1.2 Stages of Formation of the heart

Formation of the heart centres on the creation of a single linear heart tube, which then loops and forms the chambers of the heart. An overview of the elements of cardiogenesis are shown in figure 1.1 and 1.2, and the steps are discussed in more detail below.

1
- Gastrulation
  - Formation of the three germ layers

2
- First and Second heart fields

3
- Heart tube forms

4
- Cardiac looping

5
- Septae form

6
- Outflow tract forms

7
- Cardiac valves form

8
- Vasculature forms

9
- Conduction system forms
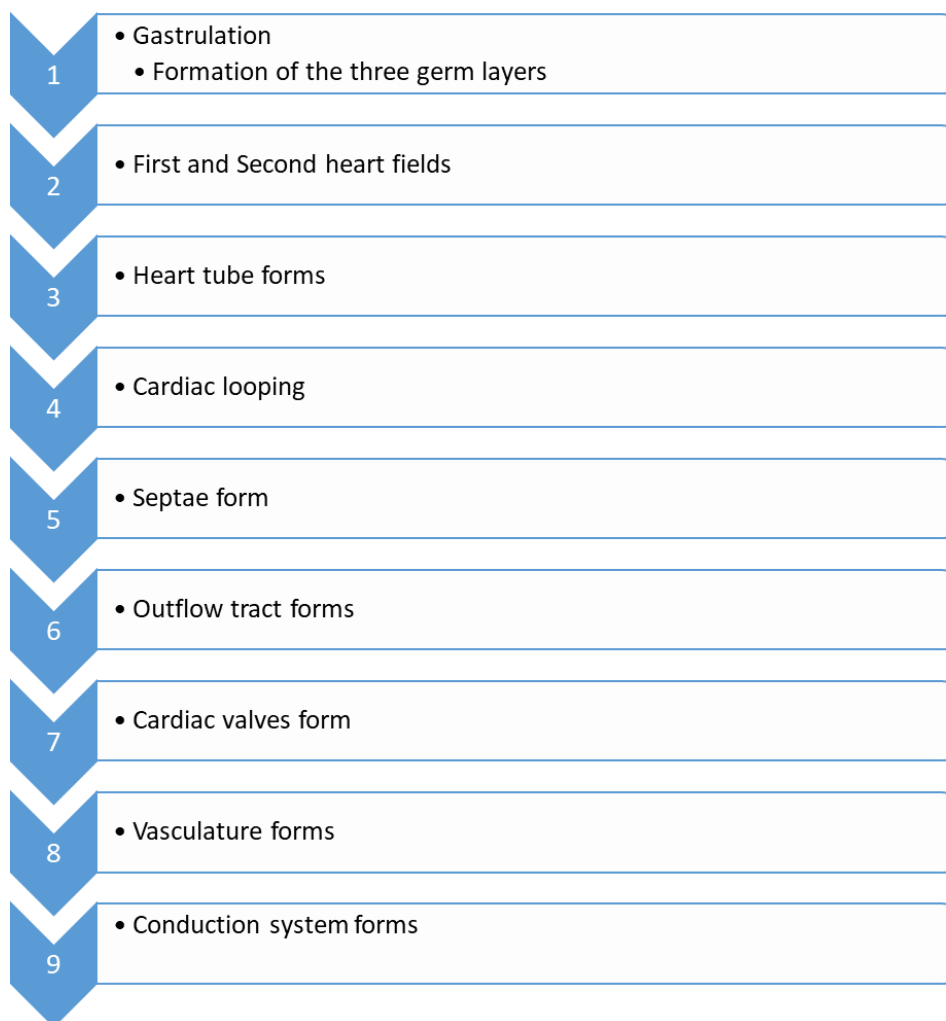
Figure 1.2. Steps in cardiogenesis.

An overview of the elements of formation of the heart in normal embryogenesis. The steps are described in more detail in the following sections.

1.2.1.2.1 Gastrulation and formation of the Cardiogenic Mesoderm

Following fertilisation (day 0), the blastocyst develops and implants into the uterine lining. Two layers of cells form not long after this; the epiblast and hypoblast. This is the bilaminar embryonic disc or germ disc. The primitive streak forms within the epiblast of the germ disc [40]. The groove of the primitive streak marks out the axis of the embryo.

By day 15, some of the epiblast cells that are located around the primitive streak, migrate between the epiblast and hypoblast layers (Figure 1.4). These cells become the endoderm and mesoderm. The cells that do not migrate become the ectoderm [39]. This produces a trilaminar structure, comprising of dorsal ectoderm, mesoderm and ventral endoderm. This is gastrulation. The heart requires contributions from all three of these cell types [41].
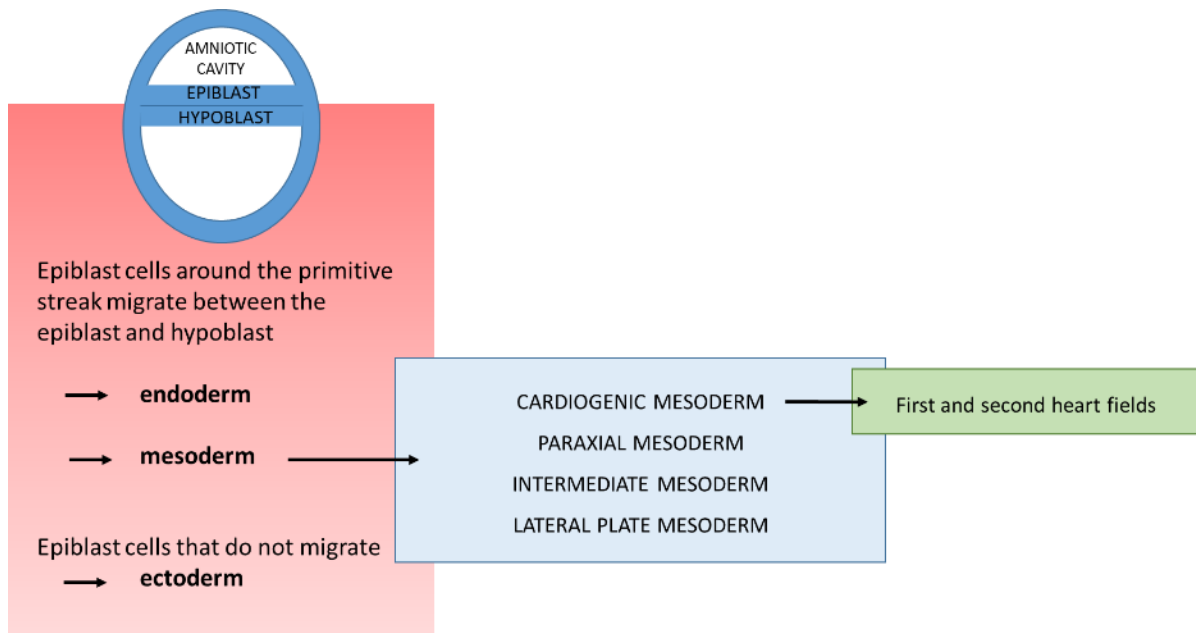
Figure 1.3 Derivation of the heart fields.

Representation of formation of the trilaminar structure in gastrulation and derivation of the first and second heart fields. The endoderm contributes to the epithelium of the respiratory, urogenital and gastrointestinal tract. The ectoderm is required for the epidermis and parts of the central nervous system. The mesoderm contributes to most other structures.

The cells that produce mesoderm delineate to form four specific mesodermal cell populations; the cardiogenic mesoderm, the paraxial mesoderm, the intermediate and lateral plate mesoderm. The first and second heart field derive from the cardiogenic mesoderm [39].

These early processes in embryogenesis are tightly controlled by a number of important gene networks (Figure 1.4). During the process of the epiblast cells invaginating through the primitive streak, *BMP*, *NODAL*, *Wnt/β-catenin* and *FGF* control the differentiation of epiblasts to mesodermal cells which express *Brachyury*, a T-box transcription factor [42]. As described already, the cells migrate to the lateral splanchnic mesoderm and differentiate into cardiogenic mesoderm and then the first and second heart fields. These cardiac precursor

cells express *Eomesodermin*, a T-box transcription factor, which plays a key role in formation of the definitive endoderm [43, 44], but which also activates *Mesp1* which is seen in cardiac progenitors as well as other mesodermal lineage cells [45-48]. This process is controlled by Wnt/ β-catenin signalling, which commands the expression of *VEGFR2 (Flk-1)* and esomesodermin [49].
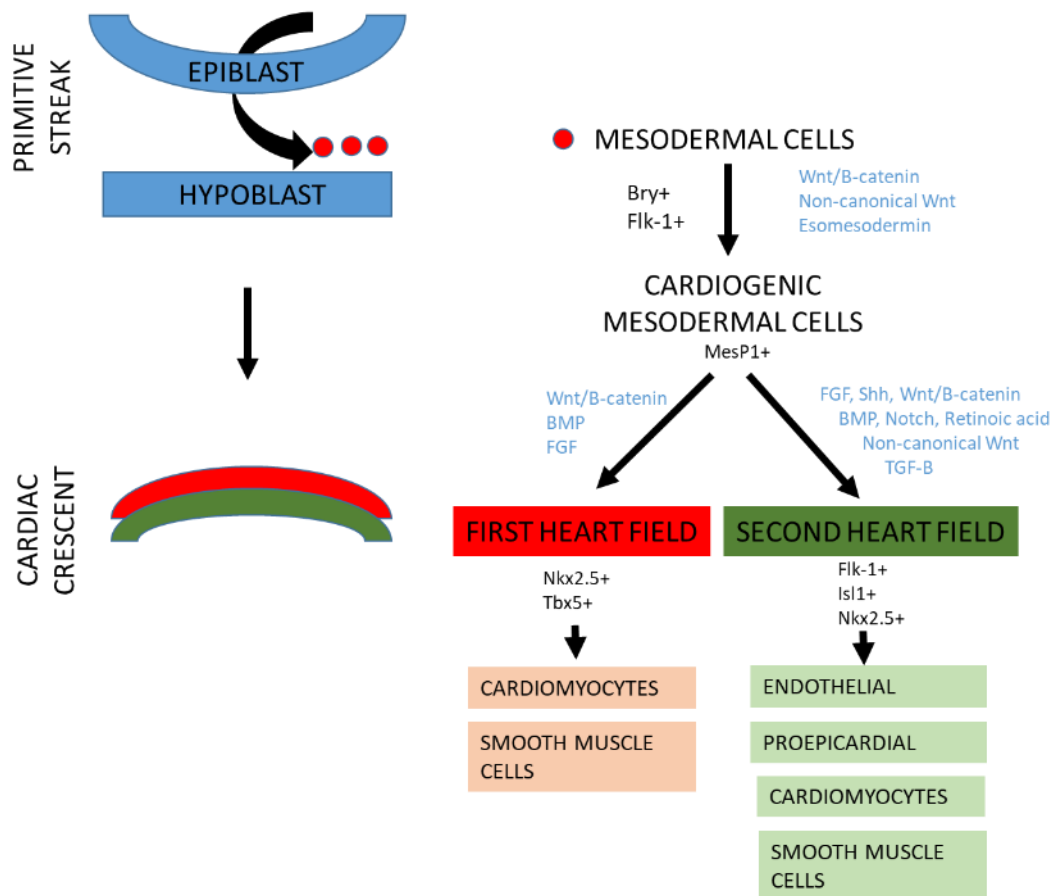
Figure 1.4. Genetic controls of cardiogenesis

Representation of gastrulation and the genetic controls guiding mesodermal differentiation into the heart fields, which are located within the cardiac crescents. Additional information on gene pathways involved from references: [50-55]. *Flk-1* = vascular endothelial growth factor receptor 2, *VEGFR2*.

1.2.1.2.2 Formation of the First/Primary and Second heart fields

Once the cardiac mesodermal cells have differentiated, they migrate in a cranio-lateral direction to form the primary heart field. This migration pattern results in changes to the orientation of the cells from a rostro-caudal pattern in the primitive streak, to a medio-lateral formation in the cardiogenic fields (Figure 1.5). The primary heart field will go on to produce the linear heart tube.

`
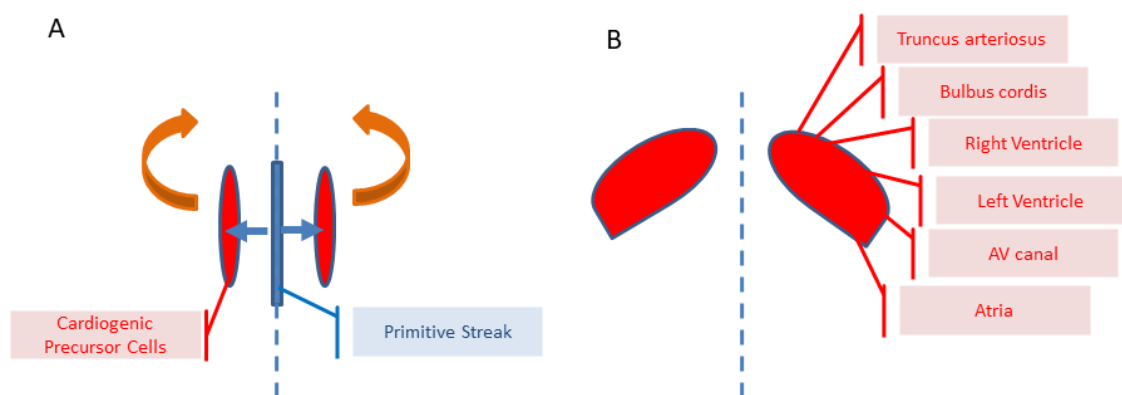
Figure 1.5 Cell migration in cardiogenesis.

A: The cardiogenic precursors migrate through the primitive streak to form the intraembryonic mesoderm. These cells then migrate craniolaterally to form the cardiogenic fields. B: The migration pattern results in a change to the orientation of the cells from a rostro-caudal pattern in the primitive streak, to a medio-lateral formation in the primary heart field.

Medial and caudal to the first heart field, the second heart field forms in the pharyngeal mesoderm. The two lie adjacent at day 15, in the cardiac crescent [39] (Figures 1.1, 1.4). The second heart field will contribute cells to the arterial and venous poles of the heart tube when it has formed [24, 56]. At the stage of the cardiac crescent, at day 15 *NKX2.5, GATA4, MEF2C, SRF* and *FOXP1* are expressed and the *WNT*, *BMP* and *FGF* pathways are active (Figure 1.4) [51-53,

55]. Canonical Wnt/β-catenin signalling inhibits cardiogenesis, whereas non-canonical Wnt/Calcium and Wnt/polarity signals support cardiogenesis [39].

1.2.1.2.3 Formation of the Heart Tube and Aorta

The heart tube is derived from cells from the first heart field. However the majority of the heart is produced from cells derived from the second heart field which are added to and elongate the heart tube [24, 57, 58].

The lateral plate mesoderm (Figure 1.3) of the embryo splits to form the intraembryonic space/pericardial coelom on day 18-19 of gestation. This results in the formation of the somatic (dorsal or somatopleuric) mesoderm and splanchnic (visceral, ventral or splanchnopleuric) mesoderm on either side of the pericardial coelom. Cells from the splanchnic mesoderm merge centrally at the caudal end of the embryo where the heart will develop. The myocardial and endocardial cardiogenic precursors are located here. The angioblastic cords develop within the heart field, and these embryonic precursor cells go on to form the endocardial heart tubes by day 18. These two heart tubes come together in the midline and fuse to become the primordial heart tube. Fusion continues in a caudal direction and occurs as the embryo folds to close the ventral foregut.

The differentiation of splanchnic mesoderm, to become cardiogenic mesoderm as described above, is influenced by a number of different genes, including *BMP2, FGF8, SHH* [59, 60]. Medial migration of the heart tubes is influenced by the *MIL* gene. *Mesp1* and *Gata4* are required for the central fusion [61-63]. Impaired migration of the bilateral sections of mesoderm towards each other means that the midline heart tube fails to form, and cardia bifida results instead. The enzyme Furin, is also required for ventral closure. Mouse embryos without Furin are embryonic lethal and cardia bifida is often observed [64, 65].

The endocardial tube consists of myocardium and endocardium separated by the cardiac jelly, which is an extracellular matrix. The epicardium forms later on from the proepicardium [56]. When active blood flow is needed to sustain the embryo, precardiomyocytes differentiate and produce a myocardial sleeve around the heart tube which begins to contracting by day 21 [66], and possibly earlier [67].

The single heart tube is modified further to produce the truncus arteriosus, bulbus cordis, primordial ventricle, primordial atrium and sinus venosus. (Figure 1.6) The sinus venosus receives the blood, which then passes through the tube to the bulbus cordis, the aortic sac and then the first aortic arch. The sinus venosus includes a left and right horn, which will become the right atrium (RA), coronary sinus and vena cava.
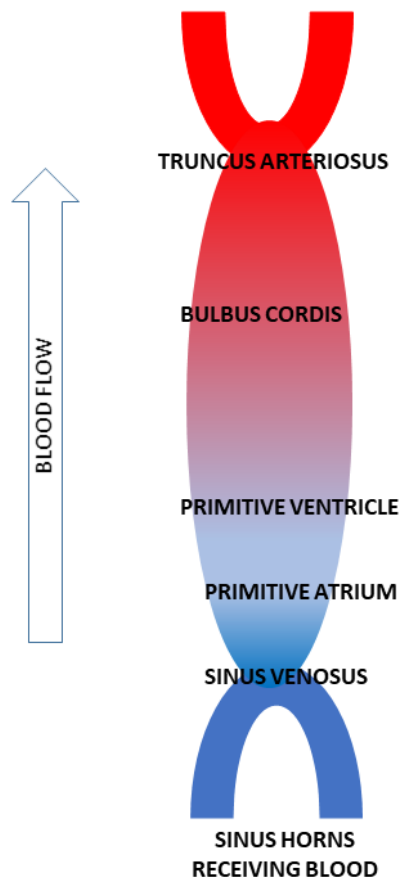
Figure 1.6 The linear heart tube.

The linear heart tube and subsections that will go on to form the chambers and other important sections of the heart. The ventricles will form from the outer curvature of the loop and the venous pole will form the atrial appendages [68].

At the stage of the linear heart tube, at day 20, *GATA4/5/6, MESP1/2, TBX5* and *miles-apart* are all expressed, and the *BMP* and *FGF* pathways are active and control terminal differentiation to myocardiocytes [51-53, 55]. Blood enters the heart tube from the cardinal, umbilical and vitelline veins and then passes through to the aortic arches and dorsal aorta [66].

The dorsal aortae form separately in the somatic mesoderm, and are separated from the heart tube by the primitive foregut [66]. They attach to the heart tubes prior to their fusion in the midline and are pulled forwards to make the first

aortic arch arteries as the embryo folds. Additional arches form later which become the arteries supplying the upper body (e.g. internal and external carotids). Below the first arch, the aortae fuse to produce a single vessel to supply the rest of the body.

1.2.1.2.4 Heart Tube Looping

Heart tube looping occurs during days 22-28 [66]. Looping of the heart tube is the first anatomical evidence of the left-right axis in the body. Its purpose is to bring the sections of the heart tube into alignment, to produce the normal cardiac anatomy. Growth of the bulbus cordis and primitive ventricle, increases the length of the heart tube and causes it to curve towards the left side of the embryo. This is the formation of the C shaped loop. Additional pressure from the dorsal mesocardium and omphalomesenteric veins help create this C shape. A group of myocardial progenitor cells which express *Isl1* add myocardium onto the loop [69]. The S-shaped loop then forms as a result of caudal migration of the ventricle section and positions the atria above the ventricles (Figures 1.1 and 1.7).
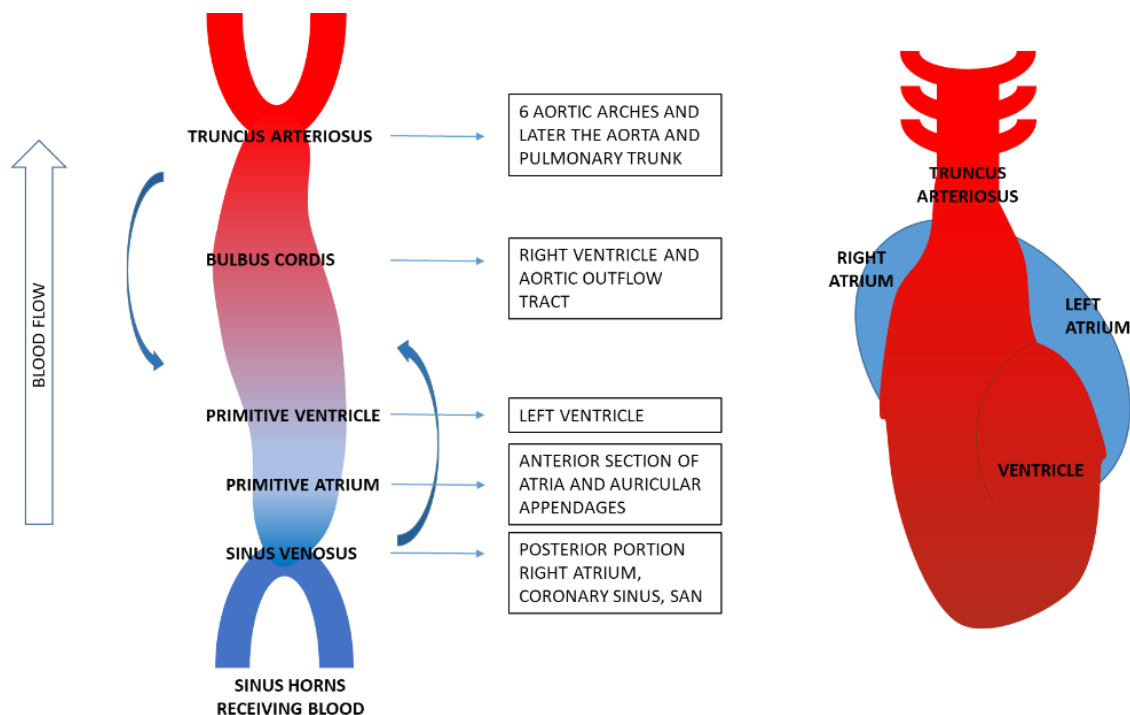
Figure 1.7 Heart tube looping.

Representation of the linear heart tube components and their final position once looping has finished.

The left and right horns of the sinus venosus go on to form the coronary sinus and the sinoatrial orifice. The sinoatrial orifice has two valves which separate the right ventricle and sinus venosus. It is eventually sited in the right atrium and will allow blood from the superior and inferior vena cavae into the heart. The coronary sinus also forms in the right atrium and drains blood used to oxygenate the heart into the right atrium here. The four pulmonary veins become fused with the posterior wall of the left atrium to receive oxygenated blood from the lungs.

As mentioned previously, looping of the heart tube is the first evidence of laterality in the embryo. This occurs at around three weeks when the cells of the primitive streak have established the right/left, medial/lateral and cranio-caudal axes. The primitive node containing cilia is part of the primitive streak and is found at its cranial end. Within the node are the motile cilia centrally and

the non-mobile mechanosensory cilia peripherally [70]. Clockwise rotation of the angulated cilia results in unidirectional flow towards the left [71]. Additional factors such as gradients of signalling molecules, movement of vesicles containing signalling modules, and activation of mechano or chemosensory cilia, may also contribute to forming a gradient across the embryo [70-72].

The leftward flow leads to induction of *NODAL*, one of the *TGFβ* subfamily, by calcium influx [73]. As the expression levels of *NODAL* increase on the left hand side, the levels of transcription factor *PITX2* and *LEFTY* gene expression also increase [74]. The LEFTY protein competes for some of the same receptors as NODAL, and so acts antagonistically to block NODAL signalling on the right hand side [75]. *CITED2* is also important in determining the axis and heart development. *Cited2* is expressed in the lateral mesoderm and acts on *Nodal, Lefty2/1* and *Pitx2* [76]*. Cited2* knockout mouse embryos are lethal with severe defects including cardiac laterality abnormalities [77].

In addition to *NODAL*, other members of the TGFβ subfamily including *BMP (BMP4, BMP2)*, *GDF* and *WNT* also contribute to laterality and associated disorders [78]. They interact with cell surface receptors which use SMAD and other non-SMAD pathways [78-80]. *Fgf* expression in the heart fields bring the inflow and outflow tracts together in close proximity [24]. *FGF8, cSNR* (inhibitor of *Pitx2*) and *NKX2.5* act on the right side, and *Shh, CFC, Nodal, Lefty, BMP2, NKX2.5, NKX2.3* and *Pitx2* act on the left to establish correct laterality [77, 81-85]. At day 30, when the heart tube has finished the looping process, *NKX2.5, SNAI1, PITX2, HAND1/2, XIRP1, LEFTY1/2, FOXH1* and *FOXP1* are all expressed, and *NODAL* and *Hedgehog* pathways are active having helped set the axis [55, 58, 86].

The determining factors that lead to the differentiation of atrial and ventricular cells are not yet fully clear. Possible contributors include *COUP-TFII* [87], localised synthesis of retinoic acid [88], and *Irx4* [89]. The left and right atria are determined by left sided expression of *Pitx2,* which supresses the

development of pacemaker tissue in the left atrium [90, 91]. Transcription factors *Tbx5* and *Hand1* determine the left and right ventricular myocytes [92, 93].

Migration of cells from the proepicardial organ onto the heart tube begins just as the heart begins to undergo looping to produce the epicardium. This will produce the coronary vasculature. In addition, whilst looping and ballooning is taking place, the cells in the secondary heart field undergo EMT (epicardial to myocardial transformation) and contribute myocardial cells to the outflow tract.

1.2.1.2.5 Formation of the Atrial Septum

Septation of the heart occurs after cardiac looping has been completed. In the fourth and fifth weeks of gestation, the atrial septum begins to form [20]. Four components come together to form the atrial septum [94-97]:

- The superior atrioventricular cushion from endocardium that has undergone epithelial to mesenchymal transition (EMT).
- The inferior atrioventricular cushion from endocardium that has undergone EMT.
- The primary atrial septum and mesenchymal cap, formed from the endocardium that has undergone EMT.
- The dorsal mesenchymal protrusion.

The primary atrial septum is formed from the second heart field. It forms at the top of the atria, and fuses inferiorly in the atrioventricular canal region with the atrioventricular cushions (Figure 1.8) [20]. This divides the cavity into the left and right atria.

Cells from the second heart field allow growth of the atrial septum primum towards the atrioventricular endocardial cushions. There is a mesenchymal cap on the lower end of the myocardial atrial septum primum. The atrioventricular

endocardial cushions sit on a bed of myocardial cells which have a unique pattern of gene expression that is different to the standard myocardium present in the heart chambers. This encourages EMT and suppress the chamber myocardium specific gene expression. *TBX2, BMP2* and *NFATC2/3/4* are all required. EMT of these endothelial cells produces the superior and inferior atrioventricular cushions and the mesenchymal cap on the primary septum. The endocardial cells in the atrioventricular cushions express *ALK2/3/5, VEGFR* and *NOTCH1* [94] and Wnt/β-catenin signalling is required for formation of the cushions, and their subsequent expansion and remodelling [98]. In addition, a non-myocardial second heart field structure called the dorsal mesenchymal protrusion, extends towards the cushions with the primary septum myocardial to divide the atria [20].
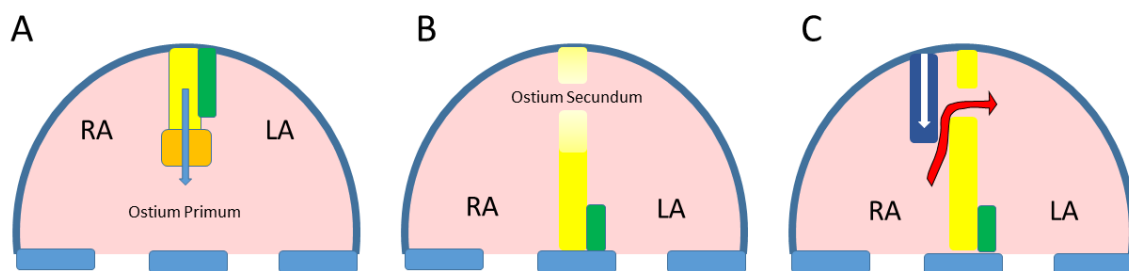
Figure 1.8 Formation of the atrial septum.

A: The atrial septum primum (primary atrial septum) grows downwards from the top of the atrial cavity (blue arrow) towards the atrioventricular endocardial cushions (blue). Attached is the mesenchymal cap (orange) and the dorsal mesenchymal protrusion (green) extends downwards with it. This process closes the ostium primum. B. The ostium secundum forms caudally as a result of breakdown of the atrial septum primum. This produces the ostium secundum. C: An additional septum, the atrial septum secundum (secondary atrial septum) forms by inward folding of the roof of the atria and extends downwards to form a flap that covers the ostium secundum (white arrow). This is the foramen ovale which will close at birth, but allows unidirectional flow of blood from right to left atrium in the foetus (red arrow). LA: Left atrium, RA: Right atrium.

The two atria remain connected by an opening that allows the fetal circulation to bypass the lungs. The ostium (foramen) primum is the gap between the septum primum and atrioventricular endocardial cushions. This closes when the mesenchymal cap fuses anteriorly with the cushions, and posteriorly with the dorsal mesenchymal protrusion [20]. The more caudal ostium (foramen) secundum results from perforations in the septum primum. These are produced by apoptosis as the ostium primum closes. A more muscular septum secundum forms around day 33 and extends alongside the septum primum. The septum secundum overlaps the ostium secundum during weeks 5-6 of gestation, but leaves the ostium secundum open. This is the foramen ovale, which acts as a valve that allows flow from right to left across the atrial septum. It does not fuse until after birth when pulmonary pressures rise [20]. There is no need for circulation through the pulmonary vasculature in the foetus, as the placenta contributes all the required diffusion.

Tbx5, *Gata4* and *Nkx2.5* are required to define the precursors for, and the actual process of atrial septum formation [99-101]. *Tbx5* is thought to be crucial in starting the process of atrial septation and may regulate the effects of Hedgehog signalling on the posterior second heart field [102]. *Tbx5+/-* mice develop atrial septal defects (ASD) [103]. Mice with conditional *Tbx5* haploinsufficiency in the second heart field develop ASDs, due to a lack of atrial septal progenitors [102]. *Tbx5* regulates Hedgehog binding protein *Gas1* and hedgehog transcription factor target *Osr1,* which are expressed in the second heart field, and ASD formation could be rescued in these mice by activation of hedgehog signalling [104].

In addition to *TBX5*, mutations in other transcription factors *TBX20*, *GATA4* and *NKX2.5* have been established as a cause of ASDs in humans [105]. Structural proteins are also important, including the sarcomeric genes *ACTC1*, *MYH6* and *MYH7* which are also associated with ASDs [106-108].

Defective formation of the dorsal mesenchymal protrusion, produced from the second heart field can contribute to atrioventricular canal defects (AVCD) [109] and ostium primum ASDs [29]. Conditional deletion of BMP receptor *Alk3* from the cardiac venous pole, can lead to impaired formation of the dorsal mesenchymal protrusion and an ostium primum defect in mice [110]. Mechanisms include reduced cell proliferation, abnormal apoptosis and impaired migration [29].

1.2.1.2.6 Division of the Atrioventricular Canal

Division of the atrioventricular canal occurs between weeks four to five. The superior and inferior endocardial cushions fuse together and form the atrioventricular septum, which splits the atrioventricular canal into the atria and ventricles. Two further endocardial cushions will later become the mitral and tricuspid valves [20]. Endothelial cells in the endocardial cushions undergo EMT, as a result of regulation by *TGFβ* and *NOTCH* signalling [68]. The increasing numbers of mesenchymal cells this produces allows growth of the endocardial cushions.

1.2.1.2.7 Ventricular Septation and the Atrioventricular Canal

A muscular interventricular ridge develops as the medial walls of the growing ventricles come together in the midline. This is the muscular ventricular septum. It stretches from the apex of the heart, to the atrioventricular endocardial cushions. Increasing ventricular trabeculation may also contribute to its formation [111] (Figure 1.9).

The interventricular foramen is a small gap that lies above the muscular septum. The four endocardial cushions of the atrioventricular canal (superior, inferior and two lateral cushions) grow and come together to form a septum which divides atrioventricular canal into right and left sides. This process helps close the

ostium primum in the atrial septum, as well as the interventricular foramen inferiorly by the end of 7 weeks gestation [111, 112].
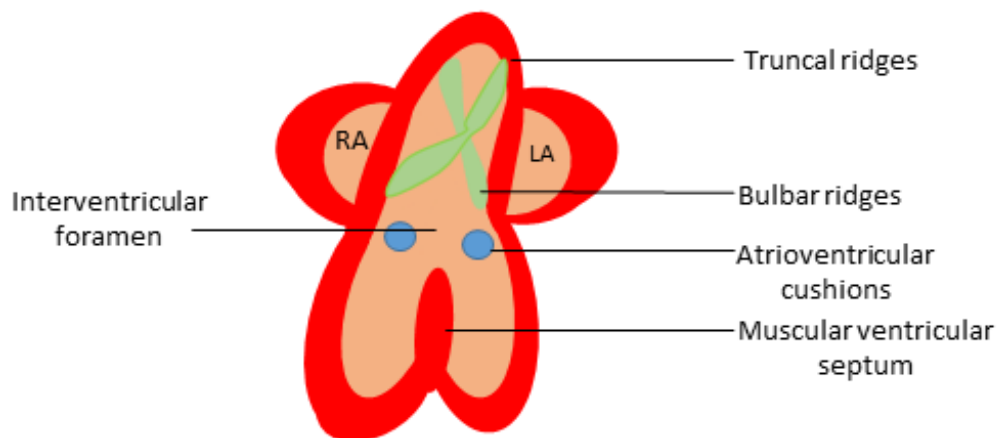
Figure 1.9 Development of the ventricular and outflow tract septum.

The muscular septum grows caudally from the floor of the ventricles. The atrioventricular cushions and components of the forming outflow tract septum (green) contribute to the membranous ventricular septum.

Superiorly, the aorticopulmonary septum rotates and then fuses with the muscular ventricular septum. This produces the membranous interventricular septum. There is an additional contribution to the membranous septum from the atrioventricular endocardial cushions [112-114].

Interestingly, there is evidence from transgenic mouse lines that the left and right sides of the septum do not overlap as the septum forms, and there is a right and left myocardial identity [115]. *Tbx5*, *Gata4*, *Nkx2*.5, *Sall4* and the *Hand* transcription factors contribute to septum formation [116, 117]. However, loss of these genes does not result in a completely absent interventricular septum, but instead results in a variety of septal defects and wider effects on the heart as a whole [50].

1.2.1.2.8 Ventricular Trabeculation

Formation of the trabeculae in the ventricles is required for a number of reasons. The trabeculae allow increased oxygen diffusion in the heart at a stage when coronary circulation is not present, allowing a concurrent increase in myocardial mass. They are also needed for formation of the Purkinje fibres and act as an anchor for the atrioventricular valves [118]. The trabeculae are highly conserved across species [50].

Animal studies suggest that myocardial projections form from the myocardium along the outer curve of the looping heart [119, 120]. These "sheets" are the beginnings of the trabeculae. They increase the surface area across which diffusion can occur, to facilitate gas exchange in the developing heart [118, 121, 122]. Perfusion is taken over later by the coronary plexus, which forms between five to eight weeks of gestation. This will supply blood to the myocardium [118]. As the coronary plexus forms, the trabeculae merge with each other at their bases. This increases the width of the compact myocardium. The recesses between the trabeculae become capillaries which are continuous with the coronary arterial system [121].

Neuregulin/ErbB, Notch, Ephrin/Eph, Hand2, YAP, DAAM1, Numb Family Proteins and *Bmp10* are all involved in the process of trabeculation [123-130]. Failure of normal compaction results in Left Ventricular Non Compaction (LVNC) [121].

1.2.1.2.9 Septation of the Outflow Tract

The outflow tract forms from two adjacent sections of the heart tube; the truncus arteriosus and the bulbus cordis sections. Septation of the outflow tract is required to form the separate pulmonary artery and aorta. This relies on the truncal and bulbar ridges and occurs during the seventh and eighth week of gestation [20] (Figure 1.9, Figure 1.10).

Septation of the cardiac outflow tract.
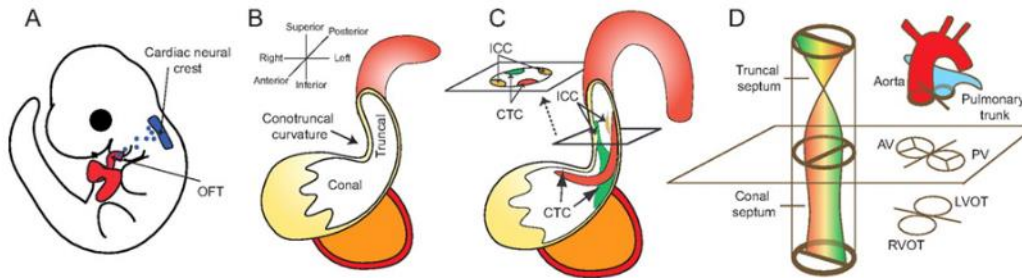
Figure 1.10 Septation of the cardiac outflow tract.

Reproduced with permission from Chien-Jung Lin et al. (A) Left lateral view of an E10 mouse embryo. The neural crest at rhombomere 6-8 gives rise to cells (blue) that migrate to and colonize the distal cardiac outflow tract (OFT). (B) The cardiac OFT contains conal (proximal) and truncal (distal) cushions. The boundary between the conal and truncal cushions is marked by an outer curvature of the OFT (the conotruncal curvature). (C) The conotruncal cushions (CTCs) and intercalated cushions (ICCs) develop within the OFT. These cushions occupy four quadrants of the OFT (shown in cross-section). The conotruncal cushions fuse to septate the OFT, as shown in D. (D) Fusion of the conotruncal cushions forms a spiral septum, the truncal part of which divides the OFT into aorta and pulmonary trunk, whereas the conal part septates the OFT into left and right ventricular outlets (LVOT, RVOT). The aortic valves (AV) and pulmonic valves (PV) develop at the conotruncal junction.

The bulbar ridges are formed in the bulbus cordis during the fifth week gestation from proliferation of mesenchymal cells [20]. The bulbar ridges grow upwards to reach into the truncus arteriosus and join the truncal ridges, above the bulbus cordis. Cardiac neural crest cells begin their migration from between the first and third somites of the neural tube and move through the pharyngeal arches into the ridges to help produce the conotruncal septum. The ridges form a helix shaped aorticopulmonary septum. The two arteries are completely separated during weeks 6-8, after fusion of the ridges in a zipper like fashion from a distal to proximal position. The helix shape means that the pulmonary trunk wraps around the aorta.

The bulbar ridges also grow inferiorly to fuse with the atrioventricular endocardial cushions and the interventricular septum. The conal endocardial cushions are the proximal outflow tract cushions which grow and fuse to form the outlet septum, as part of ventricular septation and are a necessary part of conotruncal septation [20]. The bulbar ridges also go on to form part of the aortic and pulmonary valves with the truncal endocardial cushions. The bulbus cordis produces the infundibulum of the right ventricle and aortic vestibule in the left ventricle.

Abnormalities of the outflow tract and conotruncal region can lead to double outlet right ventricle (DORV), transposition of the great arteries (TGA), TOF, VSD, common arterial trunk and interrupted aortic arch (IAA). *BMP2, SHH, FOXC1/C2, NKX2.5, TBX1, PITX2, FGF8* and *FGF10* are all required in septation of the outflow tracts [117].

1.2.1.2.10 Valve Formation

EMT in the endocardial cushions is central to the process of valve formation. The endocardial cushions are found in the atrioventricular canal (section 1.2.1.2.6) and the outflow tract (section 1.2.1.2.9). EMT causes expansion of the

endocardial cushions in the areas that will become the valves. EMT is initiated by localised expression of genes that support EMT such as *Bmp2*, which controls migration of these transformed cells in the cardiac jelly [131]. The cushions and mesenchymal cells then remodel to produce a more elongated valve shape. Apoptosis of cells at the base, and proliferation at the tip extend the cushion and produce the recognisable shape of the valve leaflet [132].

The mitral and tricuspid valves form between 5-8 weeks gestation [94]. The mesenchymal cells generated from the atrioventricular endocardial cushions by EMT, divide the canal into left and right atrioventricular canals and form the mitral and tricuspid valves. Later migration of the epicardial cells onto the valve leaflets and cushions completes the process of valve formation [20]. Both the mitral and tricuspid valves become attached to the heart wall by the chordae tendinae and papillary muscles.

The semilunar valves form from two conotruncal cushions and two intercalated cushions [133-135]. Whilst the atrioventricular valves rely on EMT alone for their formation, the outflow tract valves rely on both EMT in the endocardial cushions, and an additional contribution from the cardiac neural crest cells [20]. The outflow tract forms from the second heart field and neural crest cells join to provide the cells for the cushions of the *distal* outflow tract. The *proximal* outflow tract cushions are derived from the underlying endothelium and EMT [136]. The neural crest cells are also required for the formation of the outflow tract septum and the ventricular septum (sections 1.2.1.2.7 and 1.2.1.2.9) [20, 137].

Control of growth and formation of the atrioventricular canal and valves requires *Bmp2,* which induces EMT [131]. *GATA4* expression in the endocardium also activates other factors required for EMT such as *ErbB3* [138]. *Notch* and *Bmp2* signalling are required to control whether EMT in invasive or non-invasive [139]. BMP2 supports invasive EMT [131]. In addition, TGFβ is required to induce the expression of Snail family of transcription factors (*Snai1*, *Snai2*), which causes

the cells to become invasive after undergoing EMT [139]. BMP2 inhibits endothelial Gsk3β which stabilises Snail1 and also supports invasiveness. In contrast to *Bmp2*, *Notch1* seems to support non-invasiveness. Mouse studies show that active *Notch1* signalling in the endocardium results in expression of *Hey1* and *Heyl*, an activated mesenchyme gene pattern in the ventricles and non-invasive EMT [140]. Inhibition of Snail1, TGFβ2 or Notch1 reduced the BMP2 controlled transformation and invasion.

The signalling pathways supporting the formation of the outflow tract valves share many similarities with those that contribute to the atrioventricular valves, but there are also a number of differences (Figure 1.11) [50, 94]. Mouse models suggest that VEGF signalling through VEGFR1 receptors is required for EMT in the outflow tract, but not the atrioventricular canal [136]. *VEGFR1* is down regulated after EMT has occurred. VEGF then acts via VEGFR2 receptors in the atrioventricular cushions to facilitate remodelling into their final valve leaflet shape. *VEGF* expression in myocardium where the valves form, is inhibited by the actions of calcineurin/NFAT signalling in mice [141].

Genes and pathways essential for cardiac septation and valve development.
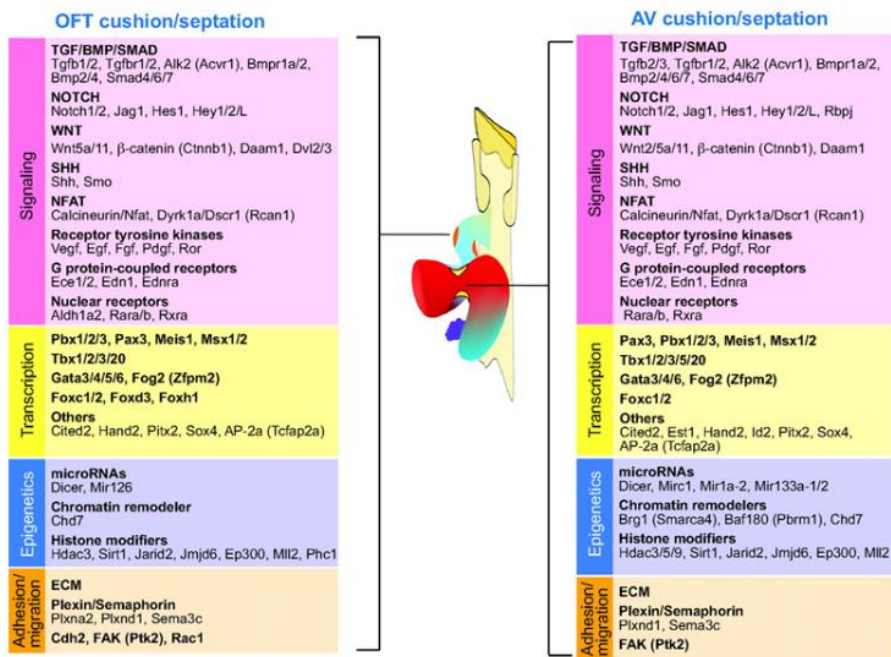
Chien-Jung Lin et al. Development 2012;139:3277-3299

© 2012.

Figure 1.11 Genetic controls of septation.

Reproduced with permission from Lin et al [94]. Many of the same genetic pathways play a role in both the septation, and formation of valves, in the atrioventricular canal and the outflow tract. However, the process is also presided over by multiple other genes and networks.

The processes that regulate elongation and maturation of the valves are less well understood, in comparison to their formation. *Calcineurin* and *Nfat* signalling are required for these later stages in the mouse and zebrafish [141]. *Sox9* knockout mouse models are embryonic lethal with hypoplastic endocardial cushions and abnormal extracellular matrix suggesting it is required for normal valve formation [142]. Conditional deletion of *Sox9* causes thickened heart valve leaflets and calcium deposits, which are common features of valve disease. This suggests that *Sox9* is required for early and later processes of valve

development. *Scx* is also expressed form E15.5 onwards, and postnatally in remodelling heart valves. Knockout mice develop thickened heart valves [143].

1.2.1.2.11 The Aorta and Pulmonary Artery

Cells from the primary heart field form cardiac myoblasts which produce the endocardial tubes[144]. Initially bilateral ventral and dorsal aortae form, which are connected through the first aortic arch[145] (Figure 1.12). The two ventral aortae coalesce to form the aortic sac. The aortic sac is superior to the truncus arteriosus. It splits to form two horns, the right becomes the brachiocephalic artery and the left forms the part of the aorta proximal to the brachiocephalic trunk with rest of the aortic sac. The aortic sac also gives rise to the aortic arches [146]. Six pairs are produced in total, but they form and regress at different times. The two dorsal aortae remain separate at this point and receive the aortic arches on each side, from the aortic sac in the midline.
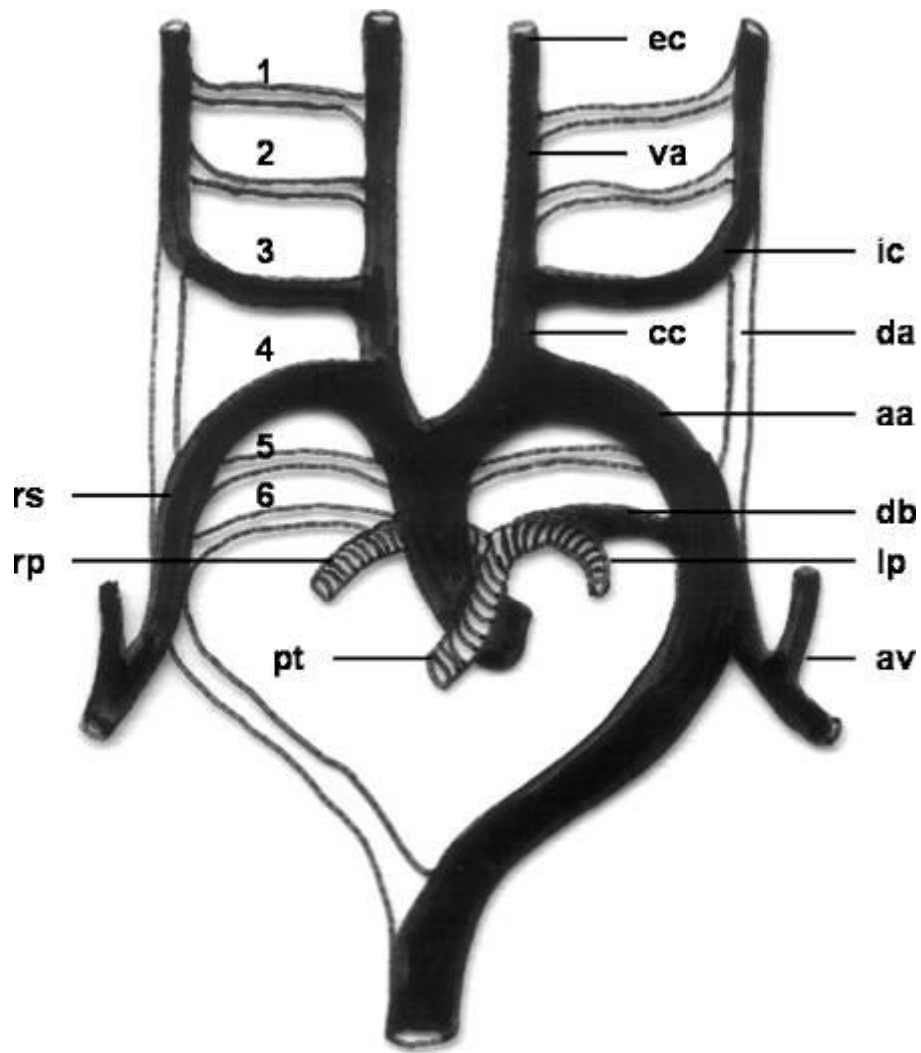
Figure 1.12 Schematic drawing of the development of the aortic arch and its branches.

Reproduced with permission from Kau et al (Thieme) [145]. 1, first aortic arch; 2, second aortic arch; 3, third aortic arch; 4, fourth aortic arch; 5, fifth aortic arch; 6, sixth aortic arch; aa, aortic arch; va, ventral aorta; da, dorsal aorta; cc, common carotid artery; ic. Internal carotid artery; ec, external carotid artery; rs, right subclavian artery; av, vertebral artery; pt, pulmonary trunk; rp, right pulmonary artery; lp, left pulmonary artery; db, ductus arteriosus botalli.

The pharyngeal arches start to form from day 28. The aortic arches are the vasculature of the pharyngeal arches, and are paired with a cranial nerve. There are five paired arches; I, II, III, IV and VI (there is no fifth arch) [145]. These

aortic arches form segments of the aortic arch, its branches and the pulmonary arteries.

The aortic arch itself forms from the aortic sac, the left forth aortic arch and the dorsal aorta [146]. The most proximal branch of the aortic arch, is the brachiocephalic trunk on the right hand side. As described already, this is formed from the right horn of the aortic sac. It has two branches; the right subclavian artery (SCL) which is formed from the 4th right aortic arch and dorsal aorta [147], and the right common carotid artery (CCA) formed from the proximal segment of the 3rd aortic arch [145]. The second branch of the aortic arch is the left CCA, which is also formed from the 3rd aortic arch [145]. The final branch of the aortic arch is the left SCL artery, which is formed from the 7th intersegmental artery [147]. The descending aorta forms from the dorsal aorta [148].

The pulmonary trunk and right and left pulmonary arteries are derived from the 6th aortic arch [147]. The also produces the ductus arteriosus (DA).

1.2.1.2.12 The Coronary Arterial System

At 42 days gestation, the truncus arteriosus is divided into the aorta and pulmonary artery. Surrounding the aorta is a capillary ring. *CXCL12* and *CXCR4* expression encourages expansion of this ring to form a vascular plexus [149]. As the workload of the heart increases, relative hypoxia in the myocardium induces *HIF-1α* and *HIF-1β* [150] and subsequent expression of angiogenic genes including *VEGF*, *FGF2* and *PDGF* [151-153]. *VEGF-A*, *VEGF-B*, and *angiopoietin-2* are also produced in the myocardium in response to *Shh* signals [154]. These gene pathways facilitate formation of the vascular plexus.

Following on from formation of the vascular plexus, apoptosis in the wall of the aorta allows endothelial cells from this plexus to create a connection between the capillaries and the aorta [155-157] (Figure 1.13). These are the beginnings

of the coronary ostia, sited above the right and left aortic cusps. This process requires cardiomyocytes and neural crest cells [149]. Blood can now flow into the vascular plexus from the aorta. The flow and associated sheer stress, induces differentiation and migration of vascular smooth muscles cells and their attachment at these sites [158].
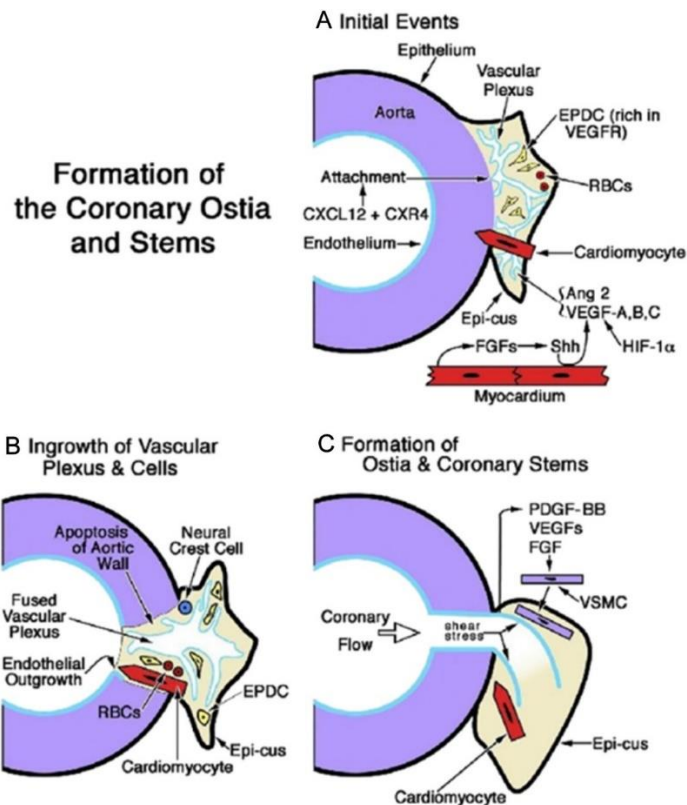
Figure 1.13 Formation of the coronary ostia.

Reproduced with permission from Tomanek et al [149]. Fig. 1. A. Formation of the coronary ostia and stems is initiated when the capillary ring that encircles the aortic root expands and attaches as a vascular plexus (in response to CXCL12 + CXCR4). The sites of ostial formation are adjacent to an epicardial cusp (Epi-cus), a thickened portion of the subepicardium, which contains epicardial derived cells (EPDC) that are rich in VEGF receptors and erythroblasts (RBCs). Cardiomyocytes guide the attachment of the vascular plexus. FGFs from the myocardium promote Sonic hedgehog (Shh) signalling, which together with hypoxia inducible factor-1 (Hif-1) stimulates VEGFs and angiopoietin (Ang) 2, thus facilitating angiogenesis of the vascular plexus. B. Myocardial-derived cardiomyocytes and neural crest cells facilitate the entry of the vascular plexus into an opening in the aortic wall, created by apoptosis. An endothelial ingrowth demarcates the pathway of the forming ostium. C. The onset of coronary flow and shear stress are key to the remodelling of the vascular plexus. Differentiation, migration, and attachment of VSMCs are influenced by 1) PDGF-BB activation in endothelial cells and the

ligand's interaction with PDGFR-β in VSMC progenitors, and 2) the influence of VEGFs and FGFs.

Vascular smooth muscle cell recruitment requires *PDGF-BB*, *PDGFR-β*, *VEGFs* and *FGFs* [149, 159, 160] (Figure 1.10). The vascular smooth muscle cell precursors migrate with the proepicardium before differentiating [35]. Neural crest cells also facilitate recruitment of vascular smooth muscle cells [161], as well as being required to guide the entry of endocardial cells through the aortic wall described above [162]. The *Notch* pathway is also required for normal coronary artery formation [163-168].

1.2.1.2.13 The Conduction System

A functioning heart is required early on in human embryogenesis to provide an adequate supply of oxygen and nutrients to the developing embryo. This requires not only correct anatomical development, but also a working and coordinated conduction system. Myocardial cells specialise to become the cells of the conduction system. The myocardial cells of the primitive heart tube spontaneously depolarise allowing conduction of an electrical impulse [169], but the ability of the cells to contract at this stage is limited due to their immaturity. The impulse is generated in the venous part of the heart tube initially, and remains at this site throughout elongation of the tube [170].

As the heart tube loops and forms chambers, a number of gene programmes required for conduction are expressed to create myocardium that has high contractility and fast conduction properties [171]. In other sections of myocardium, this "working pattern" is supressed, so the cells maintain lower rates of proliferation, slow conduction and automaticity [169]. This nodal phenotype is retained in the sinus venosus, atrioventricular canal and the inner curvature of the tube. The slower conduction in the atrioventricular canal means that the myocytes are slow to relax, effectively fulfilling the task of the valves and preventing retrograde flow at a point before they are formed.

Myocardium from the second heart field is incorporated into the posterior wall of the atria [169] and the sinoatrial node (SAN) myocardium forms at the base of the cardinal veins, where the right atrium and right sinus horn meet [172]. SAN precursor cells form at the same time as the sinus venosus, and have a distinct genetic lineage from those cells that form the atria [173]. As development of the heart progresses, the SAN takes over the pacemaker activity from the sinus venosus pacemaker cells, which become atrialised [169, 174].

Tbx3 facilitates the formation of conductive tissue by promoting the development of ion channels needed for pacemaker activity and spontaneous depolarisation [173, 175, 176]. It also represses the normal "working myocardium" cell patterning. Its expression contributes to formation of the SAN, and the atrioventricular node (AVN) which is described next.

Cells from the left and right posterior second heart field produce the anterior and posterior AVN [27]. The atrioventricular canal contains most of the cells required to form the AVN [177, 178]. Markers of the canal include *Bmp2*, *Tbx2* and *Tbx3* [175, 179]. In addition to suppression of the "working myocardium" gene expression patterns which are associated with rapid conducting cells, the slower conduction in the atrioventricular canal is maintained by expansion of the cardiac jelly during valve formation, which acts as a physical barrier to conduction [180]. The annulus fibrosis then develops from the epicardial mesenchyme that reaches the atrioventricular cushions and passes between the atrioventricular canals and the ventricles [181-183]. This means that there is no myocardial connection between the atria and ventricles, with the exception of the atrioventricular bundle attached to the AVN [169]. This ensures that the AVN is the singular gateway for impulses to pass through and prevent lethal and re-entry arrhythmias [177, 184]. *Tbx2*, *Tbx3* and *Bmp2*, *Gata4*, *Gata6* and *Notch* signalling are essential in formation of the AVN [131, 179, 184-189].

Electrical impulses are conducted through the ventricles by the Bundle of His and the Purkinje fibres, which are both derived from myocardium [190, 191]. Cells expressing *G1N2* and *Tbx3* within the atrioventricular canal forge a path into the ventricular septum and form the interventricular ring, which in turn forms the Bundle of His [175, 192]. Subendocardial myocytes in the septal trabeculae form the bundle branches and Purkinje fibres [193-195]. The expression of another transcription factor *Id2* is controlled by upstream *Tbx5* and *Nkx2.5*, and is required for formation of the His-Purkinje system [196]. *Id2* deficient mice show both abnormalities in the structure and function of the conduction system, and compound haploinsufficiency with *Tbx5* or *Nkx2.5* resulted in lack of specification of the ventricular conduction system in the embryo.

The ventricular conduction system is designed to enable fast conduction and expresses gap junction forming genes *Cx40* and *Scn5a* to facilitate this. This gene expression sets the cells apart from both the normal working myocardium, and the nodal tissue which has slower conduction properties [169]. Irx3 represses Connexin 43 (*Gja1*) and activates Connexin 40 transcription (*Gja5*) [195]. The final step required is repolarisation which is dependent on *Irx5* [197]. *Irx5* maintains a repolarisation gradient by supressing the expression of *Kv4.2* potassium channel genes, by recruiting the cardiac transcription repressor *mBop* to repress *Kv4.2*.

1.2.1.2.14 Fetal Circulation

In utero, the fetal heart is required to deliver oxygen and nutrients from 4 weeks gestation, when passive diffusion becomes insufficient [198, 199]. By week 10, the placenta has become the dominant organ for gas exchange, as opposed to the passive gas exchange that has occurred through the yolk sac and placenta prior to this [199]. The workload of the heart gradually increases throughout fetal life, and there is a sudden step up in requirements after birth [21]. This is due to the change from the fetal, to neonatal and then postnatal circulation (Figure 1.14).

The fetal cardiovascular system is designed to funnel oxygenated blood from the placenta directly to the brain and organs, bypassing the lungs. There are four shunts in place to do this; the ductus venosus, the ductus arteriosus, the foramen ovale and the placenta. In addition to the anatomical adaptations, there are additional biochemical factors which support this, such as prostaglandin and endothelin-1 which increase pulmonary vascular resistance in the fetal circulation [21, 200-202]. The reduction in pulmonary vascular pressure after birth and release of bradykinin, means that the blood begins to flow through the pulmonary circulation, and the ductus begins to close immediately [202-204]. Fusion is normally complete by three weeks of age.
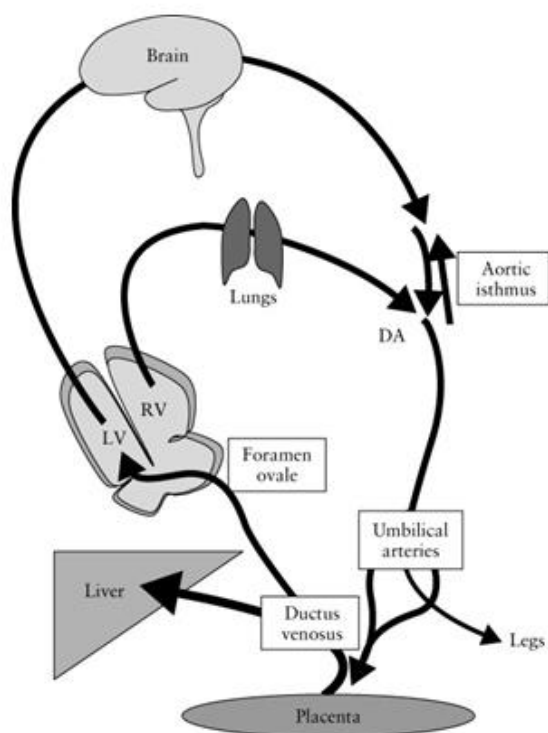
Figure 1.14 The fetal circulation.

Reproduced with permission Wiley 2006 [205]. Diagram illustrating the serial partitioning of nutrient- and oxygen-rich blood reaching the fetus via the umbilical vein. The first partition at the level of the ductus venosus distributes the majority of umbilical venous blood to the liver. Umbilical venous blood that continues towards the heart is partitioned towards the left ventricle (LV) at the foramen ovale. This blood supplies the brain and upper part of the body via the brachiocephalic circulation and the myocardium via the coronary circulation. A minor proportion of blood from the right ventricle (RV) supplies the lungs, while the remainder continues through the ductus arteriosus (DA) towards the aorta. At the aortic isthmus, bloodstreams directed towards the descending aorta are partitioned based on the relationship of blood-flow resistance in the brachiocephalic and subdiaphragmatic circulations. While net forward flow is maintained under physiological conditions, diastolic flow reversal occurs when brachiocephalic resistance falls and/or subdiaphragmatic (placental) resistance rises. Finally, the major proportion of descending aortic blood is partitioned at the umbilical arteries, to return to the placenta for respiratory and nutrient exchange.

1.3 Pathophysiology of CHD

Understanding the process of formation of the heart, allows us to determine the mechanisms by which different types of CHD develop. Defects might arise as a result of a structure failing to form, or abnormal apoptosis. This is the case in secundum ASDs, where the defect lies within the fossa ovalis. The flap may not be sufficiently sized to cover the fossa, or there may be perforations resulting from either failure of formation of the floor of the fossa from the primary atrial septum, or failure of normal apoptosis [206].

An ostium primum ASD is a form of atrioventricular defect, rather than atrial septal defect. There is an AVSD, but blood is only shunted between the atria as the atrioventricular valve leaflets fuse with the crest of the muscular ventricular septum, to divide the atria from the ventricles. This and other atrioventricular canal defects are a result of failure of normal growth of the vestibular spine (mesenchymal cap) [206, 207]. The endocardial cushions may fuse normally. Depending on whether the bridging leaflets are fused to the atrial or ventricular septum, or neither (floating) determines where shunting can occur.

A key part of normal ventricular septation is transfer of the aorta from over the RV to over the LV. This reduces the size of the interventricular foramen by bringing the fused outflow tract cushions and the top of the muscular ventricular septum into closer proximity. The foramen will then be closed by the atrioventricular cushions. Failure of this process will result in a VSD [206]. A muscular VSD results from a different mechanism, which is failure of trabeculations to come together. This leaves a pathway in the septum between the trabeculae, through which blood can flow.

TOF and DORV can be considered part of a spectrum. Abnormalities which develop whilst both arterial trunks remain over the RV, when the intermediate cushions have already created a separate aorta and pulmonary artery, will result

in DORV [206]. DORV can also result from abnormal development of the atrioventricular region of the heart. Abnormal alignment of the tricuspid orifice so that it is aligned with the left part of the outflow tract septum has also been demonstrated [208]. The proximal cushions should go on to form the subpulmonary infundibulum. If they obstruct the pulmonary trunk outlet, TOF results. Hypertrophy of the RV then occurs as a consequence of the outflow tract obstruction.

Defects can result from a variety of mechanisms which interact. For example, an initial event may be precipitated by an environmental change or genetic variant, and then be compounded by altered haemodynamics within the heart. Hypoplastic left heart can be a consequence of reduced blood flow secondary to aortic and mitral valve atresia or stenosis. Hypoplasia of the RV can also occur in the setting of an atretic pulmonary valve. The size of the ventricle will vary, depending on the point at which blood flow fell too low to allow normal development of the ventricle.

Normal blood flow and the ability of cells to respond to changes in flow, are required for normal cardiovascular development [209]. Transcription factors such as *Klf2a* show variable expression levels in different flow conditions, and control endocardial cell shape and size, and interactions with the myocardial cells, to ensure correct chamber formation in the heart [210, 211]. Cilia are also important for facilitating response to shear stress and are found on endocardial cells. Stimulation of cilia is required to initiate ventricular trabeculation for example [212]. Cilia are also important for establishing the left right axis of the heart and are implicated in heterotaxy defects [71, 213].

1.4 The Mouse as a model for CHD

I planned to use mouse models to evaluate the impact of knockout of novel CHD genes. The mouse is an excellent cardiovascular model, as its heart closely

resembles the human heart in its embryology and genetics. A vast multitude of genetically engineered lines have been produced, along with mouse atlases describing their anatomy and standard phenotyping procedures (www.mousephenotype.org) [214]. The mouse heart follows a similar course to that already described above, with initial formation of a single heart tube which loops and expands to form the chambers of the heart. This is described in more detail in appendix A. The mouse heart may begin to beat as early as E7.5 [215].

A large proportion of knockout mouse lines are early embryonic lethal, and this often results from defective cardiovascular development [216]. This is not surprising given that many of the other organ systems are not required prior to birth when the placenta is functioning, but the heart is required to sustain the embryo. Failure to form the linear heart tube, establish circulation or adequate cardiac function can cause death pre E11.0 [217]. Abnormal looping and failure of formation of the correct anatomy of the chambers is less likely to cause lethality, and can instead result in a different spectrum of cardiac abnormalities in the surviving embryo. Death from E11.0 onwards results from a different spectrum of abnormalities including persistent truncus arteriosus (PTA), failure to separate the atria and ventricles, or failure of the conduction system [216]. Thus, lethality itself and the timing of lethality are both important outcomes to consider in assessing a mouse model.

1.4.1 Comparison of embryogenesis of the mouse and human heart

The embryology and anatomical structure of both the mouse and human foetus follow a similar course [218]. Atrial and ventricular septation and the structures present at the heart tube stage are similar in both species. The human heart is more pyramidal in shape as it sits on the diaphragm as opposed to the more ellipse shaped mouse heart. The atrial appendages are also more prominent in the mouse, and are relatively small in the human. Other important differences include a single superior vena cava (SVC) in humans and bilateral SVCs in mice [218]. In the mouse, the left SVC remains and drains into the atrium via the

coronary sinus, whereas humans have the brachiocephalic vein, which drains into the right sided SVC. There are 2-4 pulmonary vein orifices in the human, compared with one in the mouse.

There is no delamination of the tricuspid valve (TV) in the mouse in utero, and no moderator band [218]. The septal leaflet of the TV may remain attached to the septum in the newborn mouse. In both the mouse and human, the tricuspid valve is positioned inferior to the mitral valve, and both have chordae in the right ventricle between the tricuspid valve and the ventricular septum. Both mice and humans have two papillary muscles in the left ventricle.

The atrioventricular septum is thinner in humans than the mouse, and fibrous as opposed to the muscular mouse septum [218]. Different ventricular trabeculations allow identification of the left and right ventricles in the human. The left ventricle has smooth trabeculations compared with the coarser trabeculations in the right ventricle. In the mouse the difference is less pronounced, but some difference can be seen on the septal surfaces of the ventricles, where it is smoother on the right at E14.5 [219]. More subtle differences are present, but not apparent without specific detailed quantitative analysis, and it is difficult to tell the ventricles apart based on trabeculation patterns alone in mice.

Another significant difference between the human and mouse, is the fraction of time in utero during which cardiogenesis occurs. The gestational period in the mouse is 20 days and cardiogenesis is mainly complete at E15.5, which is drastically different to the human foetus which has an extended period between the heart being completely formed and birth. This may produce a different spectrum of CHD in the human compared to the mouse as there is far more potential for morphological changes due to haemodynamic stress in the human [218]. Specific examples of CHD that are potentially time dependant include HLHS, PS and Ebstein anomaly, which were not seen in 10 000 ENU mutagenic mice [220]. HLHS is a prime example of a time dependant form of CHD, which is

not present throughout embryogenesis [221]. It can occur as a result of LV outflow obstruction, or obstruction of inflow into the LV. Most commonly it results from aortic valve defects, which leads to higher afterload in the LV and subsequent dilatation. As a result the LV might actually appear normal in size initially. Later, reduced blood flow through the LV results in failure of the chamber to enlarge and further negative feedback on growth. A similar process occurs with MV abnormalities causing reduced preload.

Despite the differences, the mouse is an ideal model to study CHD given the anatomical and genetic similarities, possibility for manipulation with CRISPR and the extensive phenotype data that has already been collected in knockout mice [214].

1.5 Imaging the Mouse Heart

Traditional methods used to phenotype the mouse heart include paraffin embedding and sectioning. Newer methods include ultrasound, computed tomography (CT), optical projection tomography (OPT), magnetic resonance imaging (MRI), episcopic fluorescence image capture (EFIC) and high resolution episcopic microscopy (HREM). I had planned to use HREM to determine the cardiac anatomy of animal models used in this thesis.

HREM provides exceptionally high resolution imaging of tissue and has the added benefit of allowing 3D image reconstruction to study anatomy [222, 223]. Tissue specimens are fixed in 4% PBS-buffered formaldehyde, then dehydrated and embedded in a JB4 resin. The blocks are then mounted on a microtome for sectioning and an image is taken after each slice is removed. The sections can be 0.5 to 10μm thick. The digital images captured can then be used to create a 3D model of the structure. This method also allows study of cell morphology and gene expression, in addition to anatomy.

HREM has been used extensively to study cardiovascular development in a number of animal models including the mouse. The development of the wild type (WT) C57BL6 mouse has been characterised using HREM, which provides an important control aspect for any other mouse models studied[224]. HREM analysis forms an important part of the pipeline in the DMDD Study (Deciphering the Mechanisms of Developmental Disorders https://dmdd.org.uk/) and is used to image knock out mice.

1.6 Aetiology of CHD

CHD usually occurs as a "one off" in a family, and pedigrees exhibiting Mendelian inheritance are rare. Initial research focussed on these rare families to identify genes that cause CHD [225]. Despite the apparent scarcity of families with monogenic CHD, first degree family members have an increased risk of CHD, and this along with concordance in monozygotic twins supports a genetic component in the aetiology of CHD [226]. The overall sibling recurrence risk is 3.1% [227], but more severe types of CHD may have higher rates of recurrence (including heterotaxy, right and left ventricular outflow tract obstruction (RVOTO/LVOTO)) [3, 228-230]. Ethnic origin does appear to contribute to CHD risk also supporting a genetic contribution, although there is often a complex relationship with socioeconomic factors which can make this relationship difficult to interpret [231, 232].

Recurrence risk in children is twice as high when CHD is present in the mother, compared to the father [233]. In addition, around 50% of affected siblings will be affected with a different type of CHD confirming variable expressivity [234]. Even if the siblings are not affected with CHD, they have a higher chance of other congenital abnormalities [227]. It is therefore likely that CHD is oligo or polygenic [12, 235]. How exactly each type of CHD occurs and why penetrance and expressivity are variable is not known but there are likely to be additional maternal and environmental factors involved [236-238].

Despite advances is our knowledge of genes that cause CHD, it can be difficult to elucidate the cause of an individual's CHD. De novo and inherited mutations in single genes [12, 14, 16], more complex structural variation [17, 18, 239] and epigenetic factors [240] all contribute. At the moment we can identify a genetic cause in up to 40% of people with CHD [241-244]; this is much more likely in those who have CHD in combination with other medical issues including neurodevelopmental disability (NDD) and specific types of [242, 243].

Studies have already begun to look at the potential gains of genetic evaluation in certain clinical populations including neurodevelopmental clinics and cardiology units [245, 246]. Knowledge of the aetiology of CHD can help guide treatment and investigations, provide information about prognosis and allow appropriate genetic counselling and advice about recurrence risk. For this reason it is beneficial to expand our knowledge of the genetic causes of CHD, which is the focus of this thesis.

1.6.1 Genetic Causes of CHD

Genetic causes of CHD include chromosome aneuploidy, large chromosome deletions and duplications, smaller copy number variations and single nucleotide variants. Other mechanisms include epigenetic factors and somatic mutations. Examples are shown below in table 1.1. The majority of these identified genetic causes result in syndromic CHD (S-CHD), that is CHD in combination with other congenital abnormalities or neurodevelopmental disability (NDD). This is reflected in the higher chance of identifying a genetic cause in individuals with S-CHD, compared to non-syndromic CHD (NS-CHD, CHD in isolation).

This thesis centres on single gene causes of CHD. Transcription factors and genes involved in chromatin modification are important examples of monogenic causes of CHD. Many of these genes that cause CHD, are required for normal cardiac development in the embryo. Other pathways that have proved important

in development of the cardiovascular system, require more evidence of their role in causing CHD in humans, such as the *SHH* pathway [247-249] (Table 1.1).

A genetic variant that causes CHD can be detected in around 15-25% of individuals with the use of karyotype, FISH (fluorescent in situ hybridisation) and array-CGH (array-comparative genomic hybridisation) [246, 250-253]. Nowadays, gene panels and whole exome sequencing (WES) are used in addition, and WES may identify a diagnosis in up to 28% of individuals referred for clinical testing with CHD and a positive family history [254].

Chromosomal and Structural causes of CHD	
Abnormality	Phenotype
Trisomy 21	40-50% have CHD ASD, VSD, AVSD, PDA [255-257]
Trisomy 18	90% have CHD ASD, VSD, PDA, AVSD, valvular dysplasia, TOF, DORV [258-261]
Trisomy 13	55 - 80% have CHD ASD, VSD, PDA, HLHS, heterotaxy [259, 261-264]
Turner Syndrome 45, X	35% have CHD, 22% develop aortic dilatation [265], BAV, CoA, PAPVD, HLHS, [266, 267]
Copy Number Variation	
22q11.2 deletion syndrome (Di George Syndrome)	60-80% have CHD, accounts for 18% of ToF. ToF, IAA (type B), VSD, TA, ASD, BAV, CoA, right sided aortic arch. The phenotype is variable and include NS-CHD [268-271]
Williams-Beuren Syndrome (del 7q11.23)	50-85% have CHD Supravalvular AS, supravalvular PS, peripheral pulmonary stenosis due to haploinsufficiency of *ELN* [272, 273]
Smith Magenis Syndrome (del 17p11.2)	<45% have CHD ASD, VSD, MVP, PS, PA [274]
Wolf Hirschhorn Syndrome	50-65% have CHD ASD, VSD, PDA, PS [275]

(del 4p16.3)	
1q21.1 Deletions and duplication	30% of each CNV have CHD. PDA, ASD, VSD, TOF, TA, PS. TGA 26066539 Candidate genes include deletion of *PLXNA2* and duplication of *GJA5* [276, 277]
1p36 deletion	70% have CHD ASD, VSD, PDA, BAV, TOF, CoA [278]
1q41q42 deletion	50% have CHD BAV, ASD, TGA, VSD [279]
8p23.1 deletion	50-75% have CHD AVSD, ASD, VSD, PS, TOF [280] CHD may result from haploinsufficiency of *GATA4*
Single Gene Disorders: Transcription Factors	
NKX2.5	Mutations are associated with AD ASD, HLHS, EA, TOF, VSD and conduction disorders leading to AV block [281-283] Identification of *NKX2.5* mutations is important as individuals have a higher risk of conduction disease and sudden cardiac death [284]. NKX2.5 interacts with TBX5 [99, 282] and mouse models suggest CHD results from reduced NKX2.5 nuclear import, downregulated BMP and Notch signaling and dysregulation of genes required for cardiogenesis [285].
TBX5 Holt Oram Syndrome	75% have CHD AD VSD, ASD, AVSD, hypoplastic LV, conduction defects leading to AV block [99, 286-290] Mainly truncating variants, but gain of function mutations have also been reported [291-295]
GATA4	AD ASD, AVSD, PS, TOF VSD [296, 297] Non-coding variants in *GATA4* have been associated with BAV [298] Gata4 signalling is required by Hh-responsive progenitor cells in the second heart field, which is required for development of the outflow tract [299].
GATA5	AD ASD, BAV, DORV, TOF, VSD [300, 301]
GATA6 CHD and pancreatic agenesis	AD NS-CHD TA, TOF and ASD and S-CHD with pancreatic agenesis [302-304]
TBX1	AD TOF, PA, ASD, aorticopulmonary collateral artery, IAA, isolated left pulmonary artery, VSD [305, 306]
TBX20	ASD, VSD, MS, DCM [307, 308]
HAND1	AD Single Ventricle, VSD [309, 310]

HAND2	AD PS, TOF, VSD [311-313]
CITED2	AS, ASD, PS, SIT, dextocardia, TGA, TOF, RVOTO, TAPVR, VSD [52, 94, 314]

Single Gene Disorders: Structural Proteins

MYH6	AD mutations cause ASD, DCM [108, 315]. AR missense mutations cause HLHS [316]
MYH7	AD HCM and DCM including LVNC, EA [317]
ACTC1	AD ASD, HCM, DCM, LVNC [106]
MYBPC3	AD ASD, MR, PDA, VSD [318, 319]

Single Gene Disorders: Chromatin Modifiers

KMT2D (*MLL2*)	AD Kabuki Syndrome [320] VSD, ASD, CoA [321] H3K4 methyltransferase.
KDM6A	XLD Kabuki Syndrome [320] CoA, ASD (AVSD, PS, hypoplastic RV are rare) [320]. Removes H3K27 methylation [322].
CHD7	AD CHARGE syndrome, 80% have CHD including ToF, ASD, VSD, Conotruncal defects, DORV, PS [323, 324]. Chromatin remodeller.
NSD1	Sotos Syndrome, ASD, VSD [325, 326]
CASZ1	A zinc finger transcription factor that interacts with histones and is associated with AD VSD [327]
SMYD4	Interacts with HDAC1 and plays a role in histone acetylation, linked with DORV and TOF [328]

Single Gene Disorders: Specific Pathways

RAS MAPK Signalling Pathway Noonan Syndrome and the Rasopathies (including cardiofaciocutaneous Syndrome (CFC) and Costello syndrome)	*PTPN11, KRAS, SOS1, SOS2, RAF1, BRAF, MEK1, CBL, NRAS, NF1, MAP2K1, MAP2K2, RIT1, SHOC2* (mainly gain of function mutations) [329-337] AD Noonan Syndrome 80-90% have CHD, 50-60% have HCM PS, ASD, VSD, AVSD, PDA, HCM [336, 338, 339] CFC – 75% have CHD [339, 340]

NOTCH Signalling Pathway	NOTCH signalling is involved in controlling neural crest cell proliferation/differentiation, AV and left/right patterning [341], cell fate, formation of the heart chambers, and valves through EMT of the cushion progenitor cells [341, 342]. NOTCH1 mutations cause AD CHD; CoA, VSD, HLHS, BAV, TOF and may account for 4.5% of NS-TOF [19, 343, 344]. *NOTCH1* mutations are mainly reported in CHD affecting the left side of the heart and mutations in other Notch pathways genes also cause LVOT defects (*ARHGAP31, MAML1, SMARCA4, JARID2*) [344, 345]. AD Alagille Syndrome. Caused by LoF mutations and deletions of NOTCH ligand *JAG1* (90%) and *NOTCH2 (1-2%)* [346-348] 90% have CHD. TOF, peripheral PS [348-352] Other genes in this pathway associated with CHD in humans include *ADAM17* [353], *HES1* [354-356] and *HEY2* [94, 355, 357]
Wnt/β-Catenin Signalling	This pathway regulates proliferation of SHF cells during cardiogenesis [358]. Non-canonical Wnt signalling is also involved in the planar cell polarity process which regulates chamber remodelling and ciliagenesis [359-361]. Mutations in Wnt pathway genes are enriched in BAV and other types of CHD [12, 14]. Much of the evidence comes from mouse models [362-364]. Other genes in this pathway associated with CHD in humans include *APC* [365] , *BCL9* [16], *DCHS1* [366] and *EDN1* [367].
Bmp Signalling	*Bmp* is required for differentiation of the cardiac mesoderm [358]. It regulates *Nkx2.5* and *Mef2c* [123]. Common variants in *BMP4* are associated with CHD [368]. Other genes known to alter BMP signalling are also associated with CHD (*SMAD6* [369] , *SMAD2* [370] and *GALNT1* [371]).
NODAL Signalling	*NODAL* is a member of the TGF-β Cytokine Superfamily. This includes *BMP-2, TGF-β2, TGF-β3*. Nodal signalling is required to regulate left right patterning in the embryo. Mutations in NODAL are associated with AD heterotaxy and NS-CHD (TGA, DORV, TOF, VSD) [372, 373]. CNVs including *NODAL* have also been identified in individuals with CHD [374].

	Mutations in upstream genes such as *ZIC3,* and downstream targets of *NODAL* including *GDF1, CFC1, TDGF1, FOXH1, SMAD* and *PITX2* have been identified in individuals with CHD [375, 376].
VEGF Signalling	*VEGF* is required for the formation of the atrioventricular valves [377]. Damaging variants in *VEGF* related genes *FLT4, KDR, VEGFA, FGD5* and others have been identified in individuals with TOF [378], and might contribute to CHD in individuals with trisomy 21 [377].

Table 1.1 Overview of genetic causes of CHD.

An overview of some of the important genes and pathways involved in CHD. Structural genes are often also associated with cardiomyopathy. Additional structural genes not included in the table include *DCHS1, TTN, ELN* and *MYH11* [379]. A more comprehensive review of transcription factors and pathways involved in CHD can be found in other dedicated reviews [10, 380]. AD: autosomal dominant AS: aortic stenosis, ASD: atrial septal defect, AV: atrioventricular, AVSD: atrioventricular septal defect BAV: Bicuspid aortic valve, DCM: Dilated Cardiomyopathy, Del: deletion, DORV: Double outlet right ventricle, EA: Ebstein anomaly, HCM: hypertrophic cardiomyopathy, HLHS: Hypoplastic left heart syndrome, IAA: Interrupted aortic arch, LoF: loss of function, LV: Left ventricle, LVNC: Left ventricular non-compaction, LVOT: Left ventricular outflow tract, MS: mitral stenosis, MVP: mitral valve prolapse, PA: Pulmonary atresia, PAPVD: Partial anomalous pulmonary venous drainage, PDA: Patent ductus arteriosus, PS: Pulmonary stenosis, RV: right ventricle, RVOTO: Right ventricular outflow tract obstruction, SIT: Situs inversus totalis, TA: tricuspid atresia, TAPVR: Total anomalous pulmonary venous return, TGA: Transposition of the great arteries, TOF: Tetralogy of Fallot, VSD: ventricular septal defect XLD: X linked dominant.

1.6.1.1 Chromosomal Aneuploidy

Around 8 -13% of CHD is thought to result from chromosomal abnormalities [379, 381, 382]. Trisomy 13, 18 and 21 commonly result in CHD (Table 1.1). Both trisomy 13 and trisomy 18 have a poor prognosis, but can survive to term and women are offered prenatal testing to quantify the risk to their baby. Trisomy 21 is a far more variable condition and does not always include CHD. When it does, they are usually defects of the endocardial cushions such as an AVSD. Turner Syndrome is the only other chromosome aneuploidy with CHD that can survive to term, with the exception of mosaicism for trisomy 8 and 9 [383, 384]. Turner Syndrome is associated with BAV and CoA [265-267].

1.6.1.2 Structural Abnormalities including Copy Number Variation

There are a number of well characterised large deletions associated with CHD such as 22q11.2 deletion syndrome and Williams-Beuren Syndrome (Table 1.1). These large deletions can be detected using fluorescent in situ hybridisation (FISH). More recently, an increasing number of sub microscopic microdeletion and microduplication syndromes have been brought to our attention as a result of increased use of Array-Comparative Genomic Hybridisation (Array-CGH). Some of these copy number variations (CNVs) cause CHD.

CNVs are thought to affect over 10% of the genome [385], but estimates vary depending on the platform used to detect them. CNVs found in normal individuals are likely to represent benign copy number polymorphisms. Other CNVs have been associated with diverse disease processes and cumulative reporting of CNVs has allowed us to recognise some common microdeletion and microduplication syndromes. Many of these syndromes exhibit variable penetrance and expressivity, making it difficult to define the relationship between a CNV and disease. It is likely that other genetic, and possibly

environmental factors, contribute to the phenotype when these CNVs are present. Even monozygotic twins with 22q11.2 deletion, have been reported with discordant phenotypes (although this was an older case report with limited tests available at the time). This variability makes it difficult to assess the exact contribution of CNVs to CHD [18, 239, 276, 386, 387]. It has been suggested that CNVs can be identified in up to 1 in 8 individuals with CHD [388]. However, rare, mainly de novo CNVs are thought to contribute to around 15-20% of S-CHD (mainly 22q11.2 deletions) and 5% of NS-CHD [17, 389].

1.6.1.3 Single Gene Mutations

Around 5%-10% of CHD is thought to result from single gene disorders [244, 390]. Variants have been associated with both NS-CHD and S-CHD. Given the usually simplex nature of CHD and the associated reduced reproductive fitness, we might postulate that these mutations are mainly de novo. More recently however, inherited single gene variants have been shown to play a role in NS-CHD [12]. Additional variability of expression and penetrance associated with a single gene mutation suggests that there are likely to be other contributing genetic and/or environmental factors that modify the phenotype.

Heart development is controlled by complex networks of genes that interact to ensure correct development of the heart tube and its asymmetric looping, to produce a functioning heart, including transcription factors (Figure 1.15). Many additional groups of genes play important roles in this process including sarcomeric, signalling and adhesion genes. Variants in some of these genes can result in CHD (Table 1.1). Homsy et al. suggested that 20% of S-CHD and 2% NS-CHD might result from deleterious mutations in genes known to be in the top quartile of expression in the mouse heart during embryogenesis [14].

Stage	Cardiac Crescent	Linear Heart Tube	Looping Heart	Chamber Septation
Key Regulators	Gata4	Gata4	Gata4	Gata4
	Nkx2.5	Nkx2.5	Nkx2.5	Nkx2.5
	Pitx2	Pitx2	Pitx2	Pitx2
	Mef2c	Mef2c	Mef2c	Mef2c
	Tbx1	Tbx1	Tbx1	Tbx1
	--	Tbx2	Tbx2	Tbx2
	--	Tbx3	Tbx3	Tbx3
	Tbx5	Tbx5	Tbx5	Tbx5
	--	--	Tbx18	Tbx18
	--	--	Tbx20	Tbx20
	--	--	Irx1/6	Irx1/6
	--	--	Hand1/2	Hand1/2
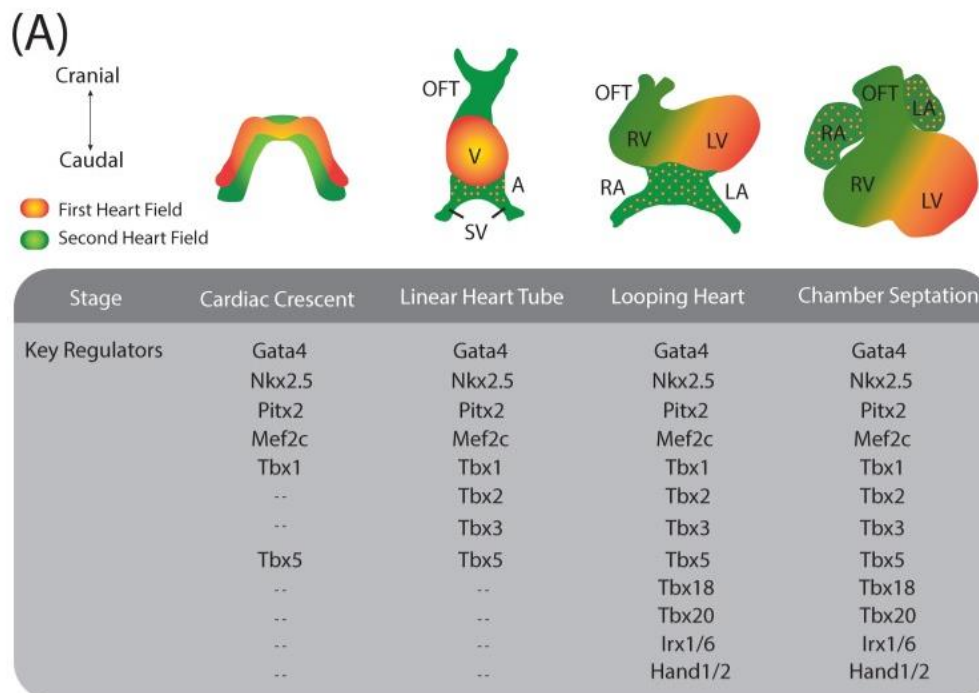
Figure 1.15 Prominent transcription factors in cardiogenesis.

From Franco et al [391]. Reproduced with permission. Figure 1A. Schematic representation of the developing heart delineating distinct developmental stages and the most prominent transcription factors involved in these morphogenetic processes.

Given that some of these pathways and transcription factors function throughout multiple developing organ systems in the embryo, it is not surprising that mutations in these genes can result in additional extra-cardiac phenotypes. Indeed, S-CHD results from single genes in many cases [392]. Some of the more well described single gene causes of CHD are shown in table 1.1.

1.6.1.4 The Epigenetic contribution to CHD

Epigenetics relates to changes in gene expression, rather than actual changes in the genetic code. Epigenetic processes includes DNA methylation, histone and chromatin modification, and regulation of non-coding microRNAs. There is plenty of functional evidence and animal models that demonstrate the importance of genes involved in epigenetic processes in cardiogenesis. It has also become evident that this holds true in human CHD, where de novo mutations in chromatin remodelling enzymes are over-represented [12, 14, 16, 393]. Importantly, some epigenetic mechanisms are heritable [209].

Methylation was the first epigenetic mechanism described, and has an established role in imprinting and X-inactivation [394]. Mouse studies have shown only a small number of differentially methylated sites within a mainly consistent methylome, which are enriched for genes known to contribute to heart development [395]. Methylation patterns of genes involved in cardiac development change throughout development suggesting methylation plays a role in control of transcription during cardiogenesis [395].

Some studies have been carried out comparing methylation patterns in normal hearts and those affected with CHD. Abnormal methylation of *NKX2-5* and *HAND1* in individuals with sporadic TOF, [396] and significant differences in promotor methylation have been identified in CHD candidate genes in individuals with TOF [240]. It is not clear however, whether the altered methylation is an association with, or contributes to, the development of TOF. The current literature does not provide definitive evidence that methylation abnormalities alone cause CHD, but suggests it may contribute in some cases. Larger studies are required.

The second epigenetic mechanism that may contribute to CHD is histone modification and the resultant changes in chromatin conformation. Subunits of

certain chromatin remodelling complexes are expressed during cardiogenesis and interact with key cardiac transcription factors such as *GATA4*, *TBX5* and *NKX2.5* [397, 398]. This include the BAF complex where RNA interference mediated knockdown in mice results in abnormal cardiac development [397, 399].

A number of genes in specific histone modification pathways appear to be significant in the aetiology of CHD [12, 14, 16]. This includes de novo mutations in genes involved in the modification of histone 3 lysine 4 (H3K4). Several genes known to cause S-CHD such as *KMT2D, KDM6A* (Kabuki Syndrome) and *CHD7* (CHARGE Syndrome) are involved in histone and chromatin modification (Table 1.1). Histone acetyl transferases such as p300 and CBP are also important in cardiac development [400]. p300 deficient mice have cardiac defects including reduced ventricular trabeculation and double null p300/CBP mutants were embryonic lethal [401]. The role of histone deactelyation is covered further in chapter 3 in the discussion of novel CHD gene *CHD4*.

A third epigenetic mechanism to consider are microRNAs (miRNAs). MiRNAs are short (19-22bp) non-coding RNAs. Although the function of most is still unknown, it is becoming apparent that they are important in normal heart development [402-408]. Studies suggest that non-coding RNAs may support the pathogenesis of CHD by influencing gene regulation [409].

1.6.1.5 Somatic Mutations

Less is known about the role of somatic variants in the heart, compared to germline variants. 1% of individuals with CHD have mosaic variants identified on blood testing which may contribute to their CHD [410]. Cardiac tissue has a higher yield (5%). Somatic mutations have been identified in hearts affected by sporadic atrial fibrillation and CHD, in *GATA6, Cx40* and *Cx43* [411, 412]. Overall, current evidence suggests that somatic mutations only play a small role

in CHD, or are part of a more complex combination of genetic and environmental factors.

1.6.2 Evidence for Oligogenic inheritance of CHD

The non-Mendelian nature of CHD in the majority, suggests that there is some contribution from either other genetic or environmental causes, or both. It is a minority of families where a monogenic cause is identified and there is now increasing evidence of non-penetrance and a role of inherited variants to support this idea [12]. Supporting evidence is described further in chapter 3, where I describe the contribution of inherited deleterious mutations in genes known to cause CHD in individuals with NS-CHD.

Evidence of an oligogenic patterns are have also been reported in mice and specific human populations. Mutations in *NKX2.5* were initially identified as causing ASD and conduction abnormalities [13, 413], but it has since been shown that the penetrance and expressivity of *NKX2.5* mutations are influenced by the genetic background in mice [414], as well as other mutations in humans [415, 416].

1.6.2.1 Maternal and Environmental contribution to CHD

Additional non-genetic modifiers and contributors to CHD include haemodynamic changes during embryogenesis, maternal and environmental factors. Epidemiological studies have attempted to pinpoint specific environmental factors associated with CHD [238]. Attributable fractions differ between the CHD lesions, from 13.6% in HLHS to 30% in TGA with intact ventricular septum (IVS). Some suggested associations include exposure to chemicals including solvents and pesticides, as well as paternal marijuana use and anaesthesia. Other paternal risk factors have been suggested including paternal age, but the

evidence is contradictory [417]. Whether these associations prove to be causal or not remains to be seen.

There are a number of well described maternal medical conditions and medications that can increase the risk of CHD in a pregnancy. Cardiogenesis occurs mainly between 2 and 7 weeks of gestation, so maternal factors in this period and the three months leading up to conception are usually considered [418, 419]. Twins have a higher risk of CHD compared to singletons, most likely due to vascular consequences of sharing a single placenta [420-422]. However, dizygotic twins also have a higher incidence of CHD than their siblings, suggesting that there may be an additional maternal environmental component [423].

Older maternal age increases the risk of CHD [424, 425], as does maternal smoking, especially in the first trimester [426, 427]. Alcohol consumption has been suggested to increase the risk of CHD, but this was not supported in a meta-analysis [428]. Therapeutic medications including anticonvulsants, non-steroidal anti-inflammatory drugs, trimethoprim, sulfasalazine, retinoids, ACE inhibitors and thalidomide are associated with CHD [429-431].

Maternal obesity has been associated with a number of congenital abnormalities including CHD, in an apparent dose response relationship [432-434]. Pregestational diabetes (type 1 and 2) is associated with a risk of CHD in the foetus (RR=4, 4-5% compared with 1% [236, 237, 435]. Interestingly, the risk of CHD in a diabetic mother's offspring is not significantly different between those with type 1 or type 2 diabetes [236]. The type of diabetes may correlate with the subtype of CHD seen in the foetus, with AVSD and conotruncal lesions being associated with type 1 diabetes, and heterotaxy and LVOTO being mainly associated with type 2 [435].

The beginning of fortification of food with folic acid was associated with a trend towards a lower incidence of CHD [436, 437]. The exact mechanism by which folic acid could influence cardiogenesis is unclear, but it may play a role in migration of cardiac neural crest cells [438, 439]. The association remains uncertain however.

1.7 Recurrence Risk

Originally CHD was thought to be due to de novo mutations, because of the detrimental effect on reproductive fitness and the low risk of developing CHD, even if you have an affected first degree relative [3]. However recurrence is seen and Ellesoe et al found non-penetrance in apparently obligate carriers in 32% of families with more than one affected individual [440]. Whilst a minority of families have monogenic CHD and a potentially significant recurrence risk despite evidence of non-penetrance [13], in the majority of cases, CHD occurs in a single individual in the family. Only around 2% have an affected first degree relative [3]. Studies have shown however that there is an increased risk to relatives, and estimates of recurrence risk fall between 1-9% [3, 233, 441-448]. Recurrence risk in siblings is thought to be around 3-4%, and 3-10% if a parent is affected [448]. The risk is highest in first degree relatives, and reduces with decreased relatedness [440, 449, 450]. The relative risk to a first degree relative is 3.21, 1.78 to a second degree relative and 1.1 to a third degree relative [3].

The recurrence risk varies by CHD subtype, and some types are more heritable than others [3, 230, 233, 451-454]. The relative risk (RR) for a first degree relative of being affected with the *same* type of CHD varies considerably. For heterotaxy the RR is 79, for RVOTO it is 49, and for AVSD the RR is 25. The RR is much lower for septal defects (ASD RR 7, VSD RR is 3) [234]. Risk appears to be especially high in left sided lesions. Cardiac screening of relatives of individuals with HLHS found that 19 -55% of first degree relatives had a cardiovascular malformation [452]. BAV accounted for the majority of abnormalities, but other more diverse phenotypes included VSD, AVSD and left sided SVC. Family screening has been recommended in those with BAV because up to 20% of relatives are affected [453, 455]. For probands with CoA, another

left sided defect, 9% of screened relatives were affected as were 3% of relatives in families with dTGA [230]. The relatives mainly exhibited left sided lesions when the proband was affected with HLHS or CoA (72% and 67% respectively). The risk to second and third degree relatives was significantly lower (1% or less, and 2% or less respectively).

When recurrence of CHD occurs within a family, the individuals may exhibit the same or different CHD lesions. In a study of 1163 families with over 3000 affected individuals, Ellesoe et al. found that CHD phenotypes in affected family members were more likely to be discordant than concordant [440]. This has been shown in other cohorts [447, 452, 456]. This might suggest that when a particular genetic pathway is affected, it can result in several different defects [382, 457]. However Oyen et al found that it was more likely that siblings would be affected with the same type of CHD [3]. Ellesoe et al. confirmed that higher concordance were seen in more closely related individuals. Higher concordance was also seen in some families with larger numbers of affected relatives, in this case it is more likely that there was a monogenic cause of CHD. Despite this reported discordance, it seems that certain types of CHD are more likely to cluster together in families than others. Examples include TGA and right left patterning abnormalities, which were more likely to occur together. DORV and right left patterning were weakly associated. Whilst there are anatomical similarities between DORV and discordant ventriculo-arterial connections, these results suggest that different genetic and molecular mechanisms might contribute to these two types of CHD. This is supported by the findings of others [450].

Possible explanations for both low recurrence risk and discordant phenotypes within a family include de novo mutations, reduced penetrance, oligogenic mechanisms and influence by additional environmental factors. This makes it particularly difficult to determine a recurrence risk for a specific family unless there is a clear Mendelian pattern of inheritance. What is clear though, is that

there is a risk to close family members of also being affected with CHD with variable expressivity, but that this risk is difficult to quantify and may vary between subtypes.

1.8 Classification of CHD

The spectrum of CHD is vast and includes abnormalities of position, the great arteries, coronary arteries, veins, the valves and connections between the chambers. There are a number of ways in which CHD is classified, and there is no single system that has a clear advantage over the others. Some are based on anatomy, others are based on clinical consequence. Some are useful from a clinical and prognostic point of view, some from a genetic or recurrence point of view. It is important to consider extra-cardiac features too including neurodevelopmental delay. Some of the important classification systems are outlined below as correct phenotypic labelling was required for the WES work I carried out with collaborators described in chapter 3.

1.8.1 Anatomical Based Classification Systems

Two of the main classification systems based on anatomy are the IPCC (International Society for Nomenclature of Paediatric and Congenital Heart Disease) and HPO (Human Phenotype Ontology) systems [458, 459]. Both are hierarchical systems and have been adopted by many clinicians and researchers.

This IPCC system was produced by amalgamating the previous two main coding systems and it focusses solely on CHD, and does not include any acquired heart disease [458]. It has been designed for use in audit and research, by both clinicians and administrators. One of the founding principles is that the code represents the core of the disease, rather than the variable manifestations and consequences of it. It also includes codes for treatments and genetic syndromes

and we used IPCC codes as part of the whole exome sequencing WES project described in chapter 3 [12].

The Human Phenotype Ontology (HPO) system is not just focussed on the heart, but covers the whole body. It also incorporates genetic syndromes. The advantage over the IPCC terms is that it allows coding of extra cardiac abnormalities. However it is not as specific when it comes to CHD, and the IPCC terms do offer a greater breadth of options. HPO terms were also used in the WES work described in chapter 3 to code for extra cardiac features, and in development of a list of genes we were highly confident cause CHD in humans [12].

1.8.2 Clinical Classification Systems

CHD can also be classified by clinical consequence [460]. Commonly it is split into cyanotic and acyanotic types, and the effect of the anatomical changes on pulmonary blood flow. Increased pulmonary blood flow will occur with left to right shunts, and septal defects without pulmonary obstruction. Decreased pulmonary blood flow occurs with right to left shunts and septal defects with pulmonary obstruction. Increased pulmonary blood flow increases the risk of developing pulmonary vascular disease, which can reverse the left to right shunt resulting in cyanosis (Eisenmenger Syndrome). Treatment is indicated early to prevent the complications of pulmonary hypertension. Cyanotic or acyanotic is there for an important distinction in guiding management, as is duct dependency.

1.9 Diagnosis of CHD

CHD may be diagnosed antenatally, after birth, throughout life or in some cases on post mortem. Timing of diagnosis depends on the particular type of CHD. Given that CHD tends to occur in a single individual within a family and monogenic inheritance with full penetrance seems rare, it is difficult to identify

who is at risk of CHD and would benefit from targeted screening. Increasing our knowledge of the genetic aetiology of CHD may help clarify this.

Almost 50% of all individuals who required a procedure to treat CHD in the first year of life were diagnosed antenatally in the UK, although pick up rates vary from region to region (NICOR Report 2012-2015 https://nicor4.nicor.org.uk). USS at 18-22 weeks provides the best view of the fetal heart. Structural, functional and rhythm disturbances can be detected using USS. CHD tends to be picked up incidentally on the 18-20 week fetal anomaly scan, as there is no identified high risk obstetric population for targeted CHD screening (with the exception of maternal diabetes) (https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/398598/Congenital_heart_disease_-_HP_-_FASP76.pdf). Antenatal detection allows parents to decide whether or not to continue with the pregnancy or to come to terms with the diagnosis before the birth. Importantly it can allow the medical teams to determine if any special care will be required. Ensuring delivery in a hospital is carefully planned with neonatal support available to prevent deterioration prior to surgery can improve outcomes [461].

CHD might be detected in the neonatal period after auscultation of a murmur or if the child has symptoms suggestive of CHD. In adulthood, CHD may be detected incidentally or symptoms might arise if there is additional stress on the heart such as pregnancy, or if decompensation and symptoms of heart failure occur. Other sequlae, such as arrhythmias, may become apparent. Some of the more common lesions such as BAV may not present until later life with symptoms of valve dysfunction, or detection of aortic dilatation. ASD and CoA are also sometimes detected in adults for the first time. Later detection can mean that person is at an increased risk of complications even after repair. Repair of an ASD in later life means that there is often severe enlargement of the RA and RV, which may not improve [462]. Therefore, defining a population for screening can improve outcomes and may be facilitated by increasing our

knowledge of genetic factors that increase risk, and identifying the genes which cause CHD.

1.10 Treatment of CHD

Treatment of CHD is beyond the scope of this thesis and will not be discussed in detail. Treatment may be aimed at treating the actual structural lesion, or complications of it such as arrhythmias or cardiac failure. Surgical interventions can correct some anatomical abnormalities, or provide a functioning heart which is not necessarily structurally normal. Some individuals will require multiple surgeries, and in some cases a heart transplant may eventually become necessary. Percutaneous treatments can be used to treat a variety of lesions, including some septal defects. One of the newer areas of development are the fetal interventions which include balloon valvuloplasty. Treatment in many instances is lifelong with medications to optimise heart function and control arrhythmias.

1.11 Prognosis and Complications

Whilst great progress has been made in the treatment of CHD, some individuals do have an associated increase in morbidity and mortality. There is emerging evidence that the pathways controlling heart and brain development are closely linked and genetic variants may potentially affect both organ systems [14]. Understanding which genes cause CHD and how they affect the development of other organs too, may increase our ability to provide prognostic information. Potentially it might help us identify whether non cardiac phenotypes could potentially predict the occurrence of CHD. Additionally, we may be able to tailor treatments by genotype if there is evidence of pharmacogenetic interactions.

1.11.1 Increased mortality

The prognosis for individuals with CHD has improved dramatically over the years, with over 90% now surviving to older age [463]. Unfortunately however, individuals with CHD still find themselves at higher risk of stroke, heart failure and higher mortality overall [8, 464]. Whist this might not be unexpected for individuals with severe CHD, it appears that there is also a higher mortality with the more simple and common defects, such as ASD and VSD [465, 466]. One of the larger studies of adults with CHD has shown those with simple CHD had a mortality rate of 0.8% per annum, those with moderate severity CHD had a mortality rate of 0.7%, and those with severe defects had a mortality rate of 2% (over an 8 year follow up period) [467]. Mortality was lowest in patients who did not require any repair. Additional risk was conveyed if the individual has a genetic syndrome, subpulmonary ventricular dysfunction, cyanosis, pulmonary outflow tract obstruction, pulmonary hypertension, infective endocarditis and single ventricle physiology. Reassuringly though, the majority of adult CHD patients will have a mortality rate only slightly higher than the general population, and many of the complications such as arrhythmias and infective endocarditis are treatable and preventable. Increasing our knowledge of S-CHD genes may help us identify more individuals who may be at higher risk of mortality, and provide an opportunity to optimise their care and organise examination of other organs systems if relevant.

1.11.2 Association with Neurodevelopmental Disability

Around 10% of individuals with CHD have additional neurodevelopmental disability (NDD). This rises to around 50% when you consider those with severe forms of CHD [468]. The aetiology of NDD in CHD is difficult to define given that factors such as time on intensive care and in hospital in general, circulatory arrest during surgery, and reduced cerebral perfusion may feasibly contribute. Clinical factors that identify those at high risk of NDD are those with cyanotic lesions, those requiring open heart surgery in the neonatal or infantile period, CHD with prematurity, ECMO, heart transplants, CPR or prolonged seizures and

abnormal cerebral imaging [392]. However it is becoming apparent that genetic influences are also important and there are shared genes in the pathways required for both brain and heart development [14]. Features such as microcephaly also independently predict NDD in individuals with CHD, which is unlikely to be due to the factors mentioned above and supports an additional genetic contribution [469].

1.11.3 Association with Respiratory Disease

Respiratory dysfunction is an important cause of morbidity and mortality in individuals with CHD. It is also closely tied to the pathogenesis of CHD in cases where ciliopathy genes are altered. Cilia help determine the right left axis of the body and are involved in development of the heart, renal and neurological systems. Primary ciliary dyskinesia results in heterotaxy, CHD and respiratory abnormalities [213]. Extra precautions and preoperative optimisation can make a difference to outcomes in those undergoing surgery and who experience chest infections [470, 471], therefor identifying individuals with ciliopathies is important.

1.11.4 Association with Myocardial Dysfunction

Individuals with CHD (excluding septal defects) have a 25% chance of developing significant cardiac failure by 30 years of age [472]. The risk is related to the type of CHD, their age, the number of operations they have had and the function of the right ventricle. Interestingly cardiac failure in those with CHD may develop by a different mechanism to those who do not have CHD. Norozi et al. suggested that the mechanism may be due to chronotropic incompetence, as indicated by lower maximum heart rates in those with CHD [472]. This may mean that individuals with CHD respond differently to treatments for cardiac failure than those without CHD. Evidence for this is provided by some studies which have shown that enalapril does not help improve exercise capacity in adults who have undergone the Fontan procedure [473], or improve cardiac function and somatic growth in infants with a single ventricle [474].

A few studies have already begun to consider the implications of genetic burden on the chances of individuals with CHD developing heart failure, by considering which pathways are involved and suggesting potential mechanisms [475]. In some cases aberrations in monogenic genes can cause both CHD and heart failure such as *MYH6*. In other instances, cardiac failure might result from a combination of genetic, cardiac, and haemodynamic stress on the heart [107, 476]. Identifying links between the genetic aetiology of an individual's CHD and their risk of cardiac failure may help develop tailored treatment pathways.

1.12 Current Recommendations for Genetic Testing in CHD

Identifying the genetic cause of CHD in an individual is important for a number of reasons. A genetic test result can guide prognosis and identify whether other organ systems should be screened for abnormalities [477]. It can also identify if other family members are at risk, provide accurate recurrence risks, and allow testing in a pregnancy and preimplantation genetic diagnosis.

European and American guidance on genetic testing in CHD have been produced [10, 478]. European guidelines suggest that genetic testing should be reviewed at transition clinics in those with a recognisable syndrome, or CHD in combination with facial dysmorphism, skeletal defects and visceral organ malformations. In addition those with growth delay and NDD including behavioural or psychiatric disorders, should be considered for genetic testing. Individuals with a family history of one or more affected first degree relatives, or multiple miscarriages and/or siblings with birth defects may also benefit from testing [10, 235]. Testing should also be considered in the case of an affected foetus as this may guide management during the pregnancy, birth and neonatal period [235]. Because it may be difficult to ascertain in a neonate has S-CHD or not, testing should be considered in those that have types of CHD which are commonly associated with a known genetic syndrome such as TOF, TA, IAA and left sided obstructive lesions [479].

Generally the guidelines recommend an appropriate targeted test, and if this is negative WES with consideration of an array [10, 235, 480-484]. Whole Genome Sequencing (WGS) can then be considered on a research basis if no conclusive variant is found. Up to 35% of individuals with S-CHD with no known genetic diagnosis have a deleterious (usually de novo) variant detected using WES [12, 14, 15].

Testing for those with NS-CHD is more limited. The 100 000 Genomes project offered a panel of NS-CHD genes, but this was on a research basis and recruitment is now closed (https://panelapp.genomicsengland.co.uk/panels/212/). The current UK test directory from Genomics England only offers a microarray for individuals with syndromic CHD (https://www.england.nhs.uk/wp-content/uploads/2018/08/rare-and-inherited-disease-eligibility-criteria-march-19.pdf). WES can identify a deleterious mutation in up to 40% with NS-CHD with a family history of CHD, [485]. For those with a family history of CHD appropriate testing should be offered [10]. For those without a family history testing is not recommended unless the affected individual or their first degree relative is of reproductive age and wants to find a diagnosis, or the individual has a specific type of CHD (TOF, IAA, TA, VSD with ascending aortic aneurysm, anomalous branch pulmonary arteries). Unfortunately NS-CHD with no family history is the most common presentation of CHD, and monogenic causes are only thought to account for around 4% [486]. The European guidelines recognise that genetic testing is more common in those with S-CHD, but that headway is being made in understanding the genetic aetiology behind NS-CHD.

Phenotypic variability, reduced penetrance and identification of variants means that genetic testing can incredibly complex. Current recommendations suggest genetic testing should include pre and post-test genetic counselling and be carried out by a genetic counsellor as part of multidisciplinary team [10, 478,

479]. An individual's plans for family and their perception of risk should be explored during the discussion [10]. Ideally counselling should be carried out prior to conception, but testing during a pregnancy can also be helpful [487]. Van Engelen et al. reported that half of the patients they interviewed had postponed having children until they had been through genetic counselling for CHD, and 13% changed their mind about whether or not to have children based on the genetic information provided [488].

1.13 Hypothesis and aims

Whilst significant progress has been made in identifying novel CHD genes, the reality is that the majority of individuals with CHD do not receive a genetic diagnosis. This thesis concentrates on the identification of novel CHD genes, and animal models of some of these new genes. Identifying and providing evidence for these new CHD genes, will hopefully translate into benefit for patients by increasing diagnostic yield, and by providing a starting point for identifying related genes and pathways that might be important in CHD pathogenesis. The main aims were as follows:

- To conduct a WES based study to identify novel CHD genes.
- To identify pathogenic mutations from this WES data, for correlation with individual phenotype, validation of the mutation and feedback to the individual.
- To perform a preliminary analysis in mice for newly identified novel CHD genes, for embryonic lethality.
- To perform a preliminary analysis of mouse colonies for newly identified novel CHD genes, for cardiac defects using high resolution episcopic microscopy (HREM) and Magnetic Resonance Imaging (MRI) analysis.

Chapter 2 Methods

2.1 Methods Relating to the Whole Exome Sequencing Analysis

2.1.1 Sample Collection for WES project

The study population for this analysis was recruited across a number of centres across the world [12]. Local recruitment was through BHF funded research nurses, most recently C Thornborough, and paediatric cardiologists lead by Dr F Bu'Lock at Leicester Glenfield Hospital. Recruitment was over the preceding 20 years. DNA extraction was carried out by members of the Brook lab, mainly K Setchfield, prior to my arrival in the lab.

2.1.2 Bioinformatic Analysis

WES and bioinformatics analysis was carried out at the WTSI (M Hurles team). The pipeline matched that of the DDD project [489]. Samples with a median coverage below 29.7X were excluded. Agilent SureSelect V5 and V3 were used for exome capture. GATK was used to identify SNPs and indels [490]. DeNovoGear was used to call de novo mutations [491]. TADA was used to study genes with enrichment of both DNMs and rare inherited variants [492]. Denovonear was used to identify mutational clustering (https://github.com/jeremymcrae/denovonear).

PCA analysis was used to characterise and define a suitable control population.

Ensembl Variant Effect Predictor (VEP) was used to classify the mutation consequence [493]. Mutations were classified as either protein truncating mutations (PTV, nonsense, frameshift, splice site variants), missense variants (including in frame indels) and silent (synonymous) mutations. Full methods are reported in appendices 3A and 3B of this thesis.

2.1.3 Characterising the de novo mutation burden in CHD

The de novo mutation analysis was based on a comparison of the actual rate of de novo mutations, with the expected number under a null mutational model [494]. It is a statistical model designed to be used for gene discovery in disorders that exhibit extensive locus heterogeneity. This method has been used by the DDD Study [495] and to investigate autistic spectrum disorders, which share a similar polygenic mechanism to CHD, where de novo mutations are important contributors [14].

2.1.4 Characterising the burden of inherited mutations in CHD

A case control approach was used to determine the burden of inherited mutations. S-CHD and NS-CHD individuals of European ancestry, were compared with 12,031 population matched controls. The control group included unaffected parents recruited as part of the DDD study [495, 496] and healthy blood donors derived from the UK INTERVAL study [497]. This group had had similar exome sequencing processed by an identical pipeline.

2.1.5 Comparison with other disease models

The genome wide excess of de novo mutations found in the CHD cohort [12] was compared with other studies looking at autism and developmental disability without CHD, to compare the relative excess burden of each type of mutation [495, 498, 499].

2.1.6 Evaluation of individual Genes

The same null mutation model was used for this analysis, as was used for the de novo mutation analysis [494]. Genome wide significance was set at $p<1.3 \times 10^{-6}$. Analysis considered S-CHD and NS-CHD trios, those with de novo PTVs, and

then missense variants. This analysis did not include individuals where we had identified a genetic cause of their CHD.

To consider the cohort as a whole, a Bayesian model was used to integrate analysis of the trios and singletons, to identify genes which were enriched for both de novo and inherited variants [492]. This method was developed to take into account de novo mutations with mechanisms other than loss of function and include inherited rare variants, in cases and controls. This is the TADA analysis (Transmission and De novo Association). It has previously been used in autism spectrum disorder analysis [500]. PTVs and missense variants were considered. Candidate genes were stratified by false discovery rate (FDR) of strong (FDR <0.01/1%), intermediate (FDR >=0,01 and <0.05/>=1% and <5%) and weak (>=0.05 and <0.1/ >=5% and <10%).

The evidence for genes identified as possibly significant was reviewed. This included whether it was a known CHD gene, animal models, case reports and any characteristics of the mutations such as clustering to suggest possible gain of function mutations.

2.1.7 Gene Function and pathway analysis

The highest ranked genes from the TADA analysis, with a FDR <50%, were run through InnateDB [501] (accessed November 2015). This programme identified GO terms and pathways that are represented more than we would expect by chance. It uses information from KEGG [502], Reactome NetPath [503], and other databases. A term/pathway was considered to be overrepresented if the Benjamini-Hochberg corrected FDR was less than 10%. High confidence protein-protein interactions (confidence >/= 0.9) were ascertained from STRING (version 10) PPI database [504] to assess for over-represented combinations.

2.1.8 Copy number variation analysis

Copy number variation (CNV) analysis used CoNVex [505]. This analysis is based on read depth information and compares the sample of interest with the median read depth from a group of highly correlated samples. This produces a log2 ratio which is then adjusted using a GC correction. This score is used with the Smith-Waterman algorithm [506]. CIFER[507] was used to determine inheritance pattern. Rare CNVs were identified by comparison with results from other studies including the CNV Project Hapmap Study [508] the Wellcome Trust Case Control Consortium (WTCCC) [509], the 1000 Genomes Project [510] and the DDD normal controls (UKBS and Generation Scotland) [495]. Rare was applied to CNVs that overlapped by 80% or less with other reported CNVs that occur at a frequency greater than 1%.

2.2 Laboratory Methods

2.2.1 PCR

2.2.1.1 PCR for Mouse Genotyping

PCR protocols were supplied by the WTSI with the mice. Primers were purchased from Sigma and diluted to a concentration of 100μM. A full list of all primers used is shown in table 2.1 below.

Mouse	Primer Name	Sequence (5'>3')
*Prkd1*em1(IMPC)Wtsi	Prkd1_DF1	GGCATGGCTGGACCTAATCT
*Prkd1*em1(IMPC)Wtsi	Prkd1_ER1	CACCCATGCCCTCAAGTATC
*Prkd1*em1(IMPC)Wtsi	Prkd1_DR1	GCTTGACACTGGAAATGGAA
*Prkd1*em2(IMPC)Wtsi	Prkd1_Gly592_F	GACTCCGAGACACACAGATTCA
*Prkd1*em2(IMPC)Wtsi	Prkd1_Gly592_R	GCCATTGAAATTTTATCCGACA
*Cdk13*tm1b(EUCOMM)Hmgu	CAS_R1_Term	TCGTGGTATCGTTATGCGCC
*Cdk13*tm1b(EUCOMM)Hmgu	LacZ_2_small_F	ATCACGACGCGCTGTATC
*Cdk13*tm1b(EUCOMM)Hmgu	LacZ_2_small_R	ACATCGGGCAAATAATATCG
*Cdk13*tm1b(EUCOMM)Hmgu	Cdk13_84881_F	GCTCTAAGGGCAACCTTGAA
*Cdk13*tm1b(EUCOMM)Hmgu	Cdk13_84881_R	AGCTGGGAAGATGGTGTTGT

Table 2.1 Primers used to genotype the three mouse strains.

For all PCR reactions, 50-100ng of DNA template was used. The protocol supplied by the WTSI used Invitrogen Platinum Taq DNA polymerase. In this instance, 2µl of 10x buffer, 0.6µl of MgCl (50mM), 0.2µl of Taq, 0.2µl of DNTPs (100mM), 0.4µl of the forward and reverse primers, and 15.2µl SDW were combined.

Promega PCR Master Mix (M7501) could also be used. 1µl of forward and reverse primers (diluted to 10µM) were added to 11.5 µl of master mix. The reaction was made up to 25µl with SDW.

For reactions using NEB LongAmp Taq DNA polymerase, the manufacturer's protocol was followed allowing for the required extension time.

In all cases, amplification was carried out in a 0.2-ml PCR tube, using the T100 Thermal-Cycler (Bio-Rad). A standard protocol of initial denaturation for 5 minutes at 94°C, subsequent denaturation at 94°C for 30 seconds, annealing at

58°C for 30 seconds, extension at 72°C for 1 minute and 30 seconds, repeated 34 times was used. This was followed by a final extension at 72°C for 5 minutes, and hold at 12°C.

2.2.1.2 PCR to confirm mutations identified in human DNA samples

Primers were designed using Primer 3 and Primer 3 Plus freely available design tools on the web (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) [511]. The relevant gene sequence was derived from Ensembl (https://www.ensembl.org/index.html) [512]. NCBI blast primer design/primer designing tool was used to check if the primers might amplify any other products [513] (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). PCR protocols were designed using Optimase Protocol Writer (http://www.mutationdiscovery.com/md/MD.com/screens/optimase/OptimaseInput.html?action=none). A full list of primers is shown below in table 2.2.

Gene	Primer Name	Sequence (5'>3')
GATA6	GATA6_F	GTTGATGACAGGGACAAAATACC
GATA6	GATA6_R	CAAGTGAGCAGAATACATGGCA
PTPN11	PTPN11_308F	ACCAGGGTTGTCCTACACGA
PTPN11	PTPN11_308R	TTTCCTCTCTCCACTTCTTTCG
PTPN11	PTPN11_63F	cttgcctccctttccaatg
PTPN11	PTPN11_63R	ctgacCTTTCAGAGGTAGGATCT

Table 2.2 Primers required to confirm pathogenic mutations identified.

Amplification was carried out in a 0.2-ml PCR tube, using the T100 Thermal-Cycler (Bio-Rad). PCR required an initial denaturation step of 2 minutes at 95°C, subsequent denaturation at 95°C for 30 seconds, annealing at 58.3°C for 30 seconds, 30 seconds at 72°C to allow extension, and repeated for 29 cycles. The final extension time was 5 minutes at 72°C, followed by holding at 12°C.

2.2.2 Agarose Gel Production

1% agarose gels were made with 1g of Sigma molecular biology grade agarose powder, dissolved in 100ml of TAE buffer. 2µl of ethidium bromide (10mg/mL Sigma Aldrich) was added.

2.2.3 Running an agarose gel

To visualise the PCR reaction products, x10 loading dye (0.04g bromophenol, 0.04g Xylene Cyanol, 6.25ml 80% Glycerol and 3.75ml SDW) was added. 10µl of Invitrogen 100bp ladder was used to size PCR products.

2.2.4 Reading an agarose gel

Once the electrophoresis was complete, the gel was placed in the Gel Documentation reader (Gel Doc XR+, Bio-Rad). Images were captured using the Quantity One software (Bio-Rad).

2.2.5 Gel Extraction

Gel extraction was performed using the QIAquick Gel extraction kit (Qiagen). The manufacturer's standard protocol was used.

2.2.6 PCR clean up

PCR clean-up was performed using the QIAquick PCR purification kit (Qiagen) and the recommended protocol from the manufacturer.

2.2.7 Sanger Sequencing

Sanger sequencing was performed by Source Bioscience (https://www.sourcebioscience.com/) and Deep Seq at the University of Nottingham (https://www.nottingham.ac.uk/deepseq/). Purified PCR products (5µl at a concentration of 100ng/µl) and primers (5µl at a concentration of 3.2pmol/µl) were supplied for sequencing reactions.

Results were in .ab1 format and analysed using the free software, Chromas (Technelysium, https://technelysium.com.au/wp/chromas/).

2.2.8 Nanodrop

DNA, and RNA concentrations were measured on the Nanodrop 2000 Spectrophotometer (ThermoFisher Scientific) in accordance with the manufacturer's instructions.

2.2.9 Preparation of Phosphate Buffered Saline (PBS)

A number of different suppliers of PBS tablets were used, and made up as per the manufacturer's instructions. This was usually dissolution of 1 tablet per 100ml of distilled water.

2.2.10 4% PFA Production

400ml of PBS was made using double distilled water. 16g of PFA (Sigma P6148) was added to 384g PBS and mixed at 65°C. If necessary, a few drops of sodium hydroxide was added to help dissolution and to ensure that the pH is between 6.9 and 7.3. 10ml aliquots were stored at -20 °C.

2.3 Mouse Work

Both *Prkd1* mouse models were produced at the Wellcome Trust Sanger Institute (WTSI) and the *Cdk13* mouse was obtained from Infrafrontier [511].

2.3.1 General husbandry

Mouse husbandry was carried out by the staff in the Bio Support Unit at the University of Nottingham, mainly by S. Ecob. This included breeding of the mouse pairs to produce a stable colony and to produce pregnant dams for sacrifice, under a Home office Breeding Licence held by Anne-Marie Kelly initially, and then Ian Devonshire (P375A76FE).

Mice were housed in accordance with UK Home Office legislation in ventilated cages with access to food and water ad liberatum. They were fed standard laboratory mouse high calorie mouse diet and provided with mash in cases of increased stress. Additional tubes were added to the cages to improve their environment and to allow handling with minimal stress to the mice. They were housed in a maximum of 4 mice per cage. No mice were kept alone where possible. A standard 16 hour light, 8 hour dark cycle was followed.

2.3.2 Mouse strains studied and genotyping protocols

The background strain used was C57BL/6N. Three mouse lines were studied, two *Prkd1* lines and a single *Cdk13* line. For the purposes of this thesis, the nomenclature for these mice is outlined in table 2.3 below.

	WT	Heterozygous	Homozygous
Prkd1^{em1(IMPC)Wtsi}	Prkd1$^{+/+}$	Prkd1$^{em1/+}$	Prkd1$^{em1/em1}$
Prkd1^{em2(IMPC)Wtsi}	Prkd1$^{+/+}$	Prkd1$^{em2/+}$	Prkd1$^{em2/em2}$
Cdk13^{tm1b(EUCOMM)Hmgu}	Cdk13$^{+/+}$	Cdk13$^{tm1b/+}$	Cdk13$^{tm1b/tm1b}$

Table 2.3 Nomenclature of mice genotypes employed in this thesis.

The genotypes of the mice sent to us had been previously confirmed by the WTSI. I confirmed the genotypes of the mice we received and optimised the genotyping protocols we had been sent. I genotyped approximately the first 100 mice produced from all three strains. After that I had help from S Buxton, K Setchfield and S Rochette.

2.3.2.1 *Prkd1*^{em1(IMPC)Wtsi}

The *Prkd1*^{em1(IMPC)Wtsi} mouse generated by the WTSI using Crispr/Cas9, deletes exon 2 and some of the surrounding intronic sequence (Figure 2.1). Exon 1 is the first coding exon. Exon 2 codes for part of the cysteine rich domain of the protein. Three primers used for genotyping, they are DF1, DR1 and ER1.

Figure 2.1 Representation of _Prkd1_^{em1(IMPC)Wtsi} mouse with exon 2 deleted.

Exon 2 and the surrounding intronic sequence are shown. Exons are in capitals font, introns are in lowercase. Primer sites are DF1 red text, ER1 yellow text, DR1 green text. RNA guide sites for the Crispr/Cas9 work are in pink text. The deleted section is italicised. _Prkd1_ sequence (GRCm38.5) from Ensembl [512]. The deletion was confirmed to be smaller (Chapter 4, Figure 4.8).

Two PCR reactions were required to genotype the mouse. The primers are shown in table 2.1. In this instance, 1.5µl of 10x Buffer, 0.45µl of MgCl2, 0.15µl of Platinum Taq, 0.15µl of dNTPs, 0.3µl Primer and 11.15µl of SDW were used to complete the PCR reaction.

The combination of DF1 and ER1 only produced a product if the wild type allele was present (233bp). The DF1 DR1 combination produced a short product if the deletion is present (212bp). It could also produce a much longer sequence from the WT allele (1360bp). A *Prkd1*^{em1/em1} mouse should produce a band at 212bp. A *Prkd1*^{em1/+} mouse should produce bands which are 212, 223 and possibly 1360bp in size.

2.3.2.2 *Prkd1*^{em2(IMPC)Wtsi}

Genotyping of this mouse was based on a single PCR reaction and the products were sent for sequencing (section 2.2.7) after PCR clean-up had been carried out (section 2.2.6). The sequence and primers are illustrated in figure 2.2.

Figure 2.2 Relevant sequence for amplification by PCR to genotype the _Prkd1_^{Em2(IMPC)Wtsi} mouse.

Figure 2.2 Relevant sequence for amplification by PCR to genotype the _Prkd1_[Em2(IMPC)Wtsi] mouse.

Mouse sequence (GRCm38.5) from Ensembl [512]. Exons are in blue capitalised font, introns are in lowercase grey letters. Primer sites are highlighted in red and green text. The base that is subject to the mutation is highlighted in pink, the base change is G/A (reverse complement) or C/T.

2.3.2.3 *Cdk13*^{tm1b(EUCOMM)Hmgu} Mouse

This mouse employed a reporter tagged deletion allele (post-cre). A figure illustrating the manufacture of the tm1b allele and a comparison with the tm1a and tm1d alleles is shown in chapter 6, figure 6.5. Figure 2.3 below illustrates the primers and their position in the *Cdk13*^{tm1b(EUCOMM)Hmgu} and wildtype mouse. The section of genomic sequence in the mouse and the cassette, is included in appendix B.
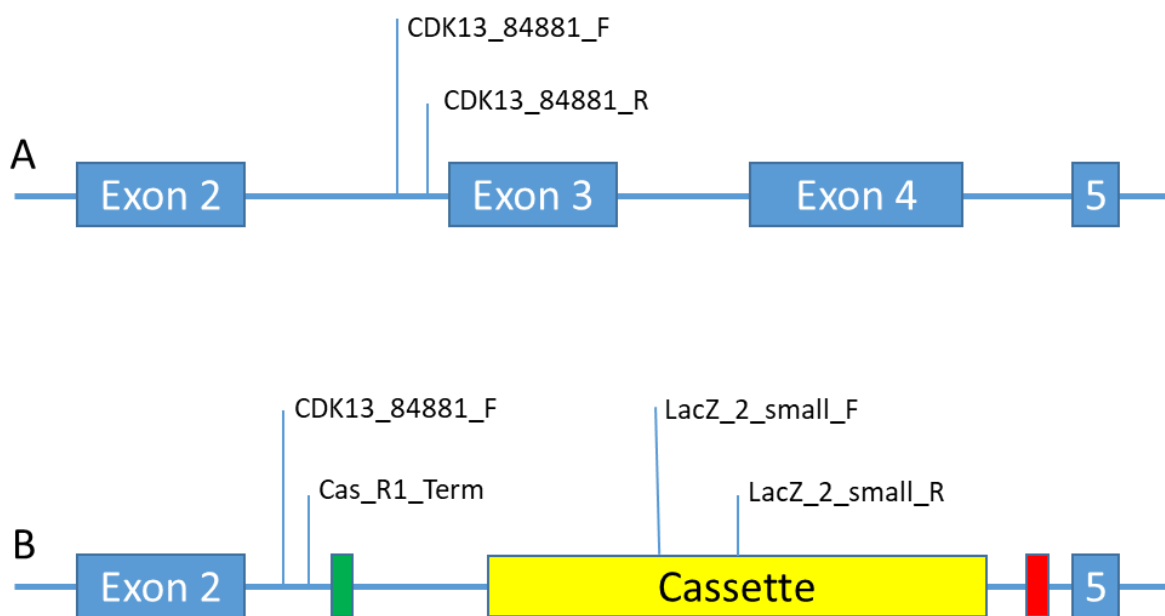
Figure 2.3 Primers to genotype the *Cdk13*^{tm1b(EUCOMM)Hmgu} mouse.

A. Wild type mouse illustrating the position of the CDK13_84881_F and CDK13_84881_R primers, just prior to exon 3. B. The *Cdk13*^{tm1b(EUCOMM)Hmgu} mouse illustrating the position of the cassette and other primers. The green box illustrates the FRT site and the red box, the LoxP site. The wild type sequence finishes just prior to the Cas_R1_Term primer.

The *Cdk13*^{tm1b(EUCOMM)Hmgu} mouse may be genotyped through a combination of separate PCR reactions that detect the cassette (LacZ_2_small_F and LacZ_2_small_R, 108bp) , the gene-specific wild type allele (Cdk13_84881_F and Cdk13_84881_R, 250bp), and a mutant allele-specific sequence

(Cdk13_84881_F and CAS_R1_Term, 200bp) using short range PCR. The standard PCR protocol was used, with the exception of a slightly shorter extension step of 45 seconds at 72°C.

Interpretation of the consolidated results reveals the genotype of the mice. For example, cassette positive, mutant positive, wild type positive represents *Cdk13*^{tm1b/+}.

2.3.3. Timed Matings

Timed matings were carried out by introducing a male stud to a new cage in advance so he could lay down his scent. Dams were then introduced and each morning they were checked for the presence of a cervical plug. If this was present, they were considered to be pregnant with a gestation of E0.5 (0.5 days post conception). Pregnant dams were then sacrificed on the morning of the specific day required. Schedule 1 methods were carried out by the staff in the BSU.

2.3.4 DNA extraction

The following DNA extraction protocol was used for ear notches and embryonic tissue, such as liver for genotyping. The protocol was supplied to me by Dr Bedford at the University of Nottingham.

Mouse tails or ear notches were placed in 700µl (tail) or 500µl (ear) of lysis buffer (0.1 M Tris-HCl, 5 mM EDTA, 0.2 % SDS, 0.2 M NaCl). 35µl (tail) or 25µl (ear) of 10mg/ml proteinase K was added, and the sample was incubated overnight at 55°C. The sample was then vortexed to ensure the tissue had been lysed. The sample was spun down and DNA was precipitated using isopropanol. This was followed by a 70% ethanol wash.

2.3.5 Culling methods used

Adult mice were sacrificed using the schedule one method of cervical dislocation, and death was confirmed by severing the neck vessels to cause exsanguination. For mice due to be imaged using MRI, sacrifice was by exposure to high concentration of carbon dioxide and confirmation of death was by identification of rigor mortis. Embryos in the first two thirds of gestation were sacrificed by disruption of the membranes and decapitation, which severs the spinal cord and allows exsanguination. Embryos more than two thirds of the way through gestation, were sacrificed by cervical dislocation and exsanguination by decapitation. No procedures or interventions were carried out on any mice whilst alive in accordance with our licence.

2.3.6 Embryo harvests for lethality assessment

Assessment of gestational age was based on detection of a mucous plug by a BSU staff member. Gestation was set as 0.5 days on finding a plug. To extract embryos, the pregnant female mouse was culled on the required day. The abdomen was cleaned with 70% ethanol and opened using a midline or V shaped incision. The uterus was isolated in the pelvis and draw upwards out of the body cavity (figure 2.4). The uterus was dissected away from the cervix/vagina. Both ovaries were identified, and evidence of reabsorbed embryos looked for. Each sac was given an identification number and those that had apparently failed to progress were recorded.

Figure 2.4 Removal of the gravid uterus.

Six separate sacs are visible in this picture.

Each uterine cavity was opened and each embryo was removed. The embryo was culled immediately using an appropriate schedule 1 method. The crown rump length (CRL) could be measured at this point by performing the schedule one method over a ruler under the transparent container, so it did not delay culling. At this point and any gross deformities were noted and tissue was removed for genotyping. The heart could be easily dissected out after culling (figure 2.5).
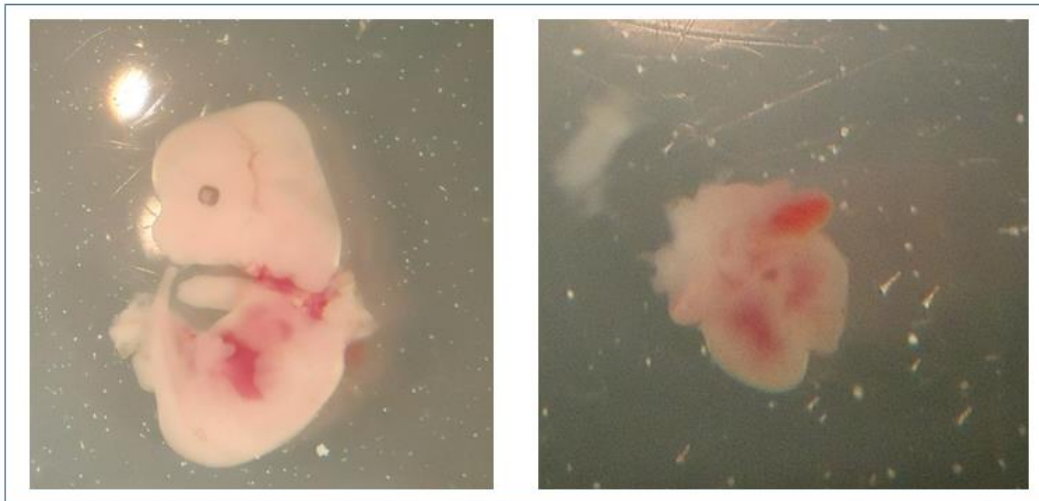
Figure 2.5 E13.5 mouse embryo.

An embryo collected as part of lethality assessment and the heart has been isolated for further work.

2.4 High Resolution Episcopic Microscopy

Protocols for HREM were obtained from https://dmdd.org.uk/hrem/ and Mohun and Weniger [513]. Additional information was added from personal communication with T Mohun and his team at the Crick Institute. In addition I made several trips to the Crick to learn the techniques in person.

2.4.1 Collection of E15.5 hearts for HREM

Embryo harvest was as described in section 2.3.6. Embryos were removed into PBS, which had been maintained at 37°C, and schedule 1 methods applied. CRL should be measured at this point. The placenta was detached and exsanguination was encouraged by repeatedly cutting a distal section from the umbilical vessels to remove clots, and by keeping the embryos in constant motion. The embryo was bisected across the region of the liver, to encourage further exsanguination. External phenotyping and forelimb staging were

performed whilst exsanguination is ongoing (see section 2.5.1). The heart was exposed via sternotomy, dissected from the diaphragm and removed (Figure 2.6). The hearts must remain submerged throughout the whole process, to prevent chambers collapse and damage.
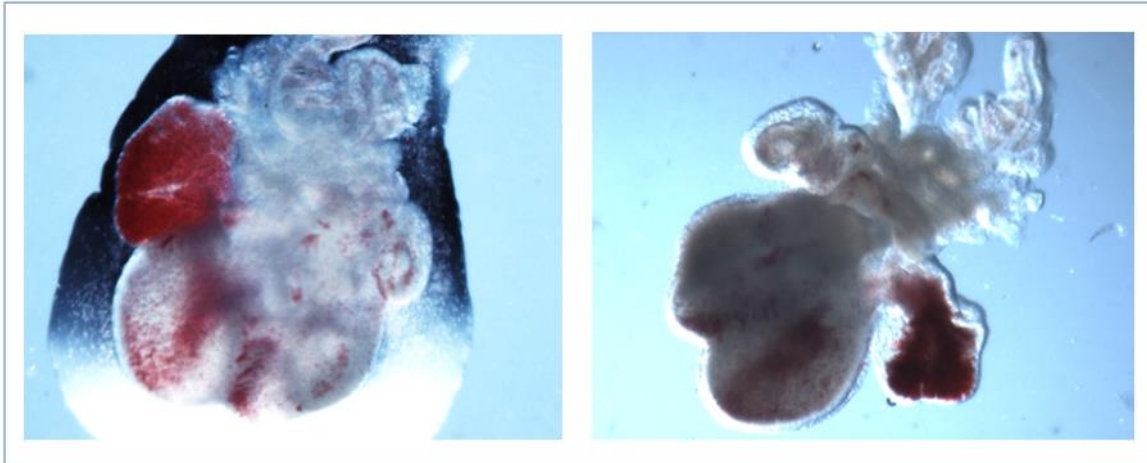
Figure 2.6 Example of E15.5 heart after initial dissection.

Some lung tissue is still attached. Blood is still evident in one of the two atria and some is also seen in the ventricles and will require removal by lysis as described below.

The hearts were then transferred to 4% PFA and left at room temperature for 20-30 minutes. Hearts were then washed extensively in tap water to remove the PFA. They were then placed in a 15ml falcon tube of water and put on roller bars. The water was replaced every 5-10 minutes. This causes lysis of any remaining blood in the heart.

Time in water was mimised to prevent blistering of the external walls of the heart due to local dissection. The hearts were transferred to 4%PFA for overnight fixation, after which additional fine dissection could be carried out.

2.4.2 Tissue Dehydration

The hearts were washed in PBS for 30 minutes at a time, for a total of 4 washes. The tissue was then dehydrated using increasing concentrations of methanol.

2.4.3 Infiltration

The hearts were immersed overnight in a 50:50 mix of methanol and JB-4 dye mix (200ml Solution A, 2.5g Catalyst (benzoyl peroxide, plasticised), 0.55g Eosin B, 0.113g Acridine Orange stirred overnight and filtered through 0.22 μm PES membrane). Samples were briefly rinsed in JB-4 dye mix to remove any residual methanol, submerged in several ml of fresh JB-4 dye mix and left overnight.

2.4.4 Embedding

Samples were embedded in JB4 mix in a custom mould, and a plastic "chuck" was placed on top (figure 2.7). Three hearts were embedded in each resin block.

Solution B and JB-4 dye mix were combined (0.6 ml Solution B per 10 ml JB-4 dye mix) and poured into the mould. 10-30 minutes after mixing, viscosity was sufficient to position the hearts apex up.

Figure 2.7 Embedding embryonic hearts.

The 10mm custom made mould required for embedding of embryonic hearts. The square wells in the centre are for the heart to sit in. The circular depression or "shelf" allows the plastic chuck to sit over the sample and adhere to it. 6 of the 8 wells contain embedded hearts in resin, and the plastic white chuck has been placed on top.

The exposed surface of JB-4 in and around the chuck was covered with a thin layer of mineral oil to allow polymerisation in an oxygen free environment, and left overnight.

2.4.5 Preparing blocks for imaging

After embedding, the polymerised blocks were removed from moulds, trimmed and baked to harden them for sectioning (Figure 2.8).
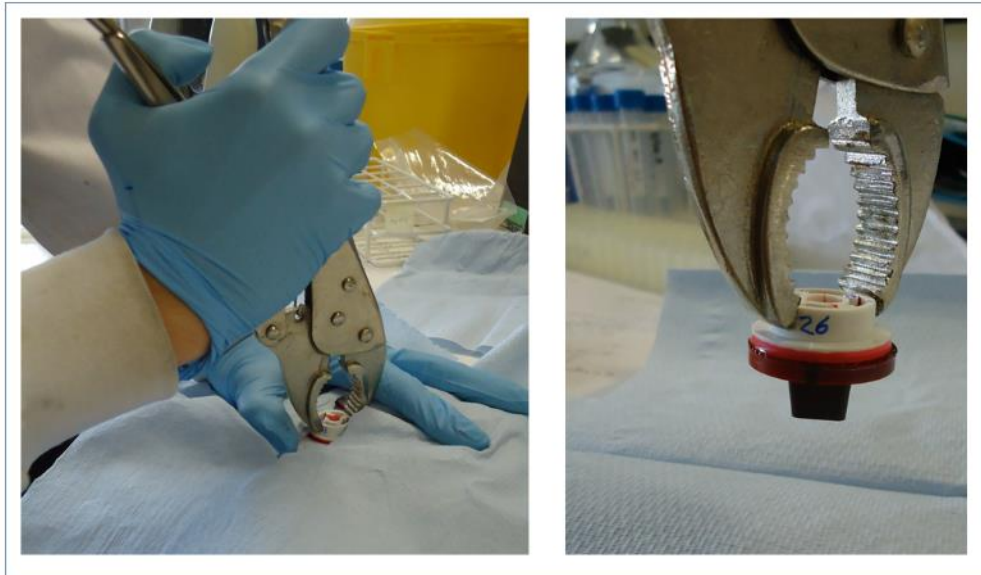
Figure 2.8 Removal of the block from the mould.

An adjustable wrench was used, and the mould covered with tissue paper to prevent any splashing from liquid JB4 and mineral oil. Care must be taken not to damage the chuck, as this is required to fix the block to the sectioning machine.

Excess resin was trimmed from around the circular edge of the chuck with a blunt knife or scissors (Figure 2.9).
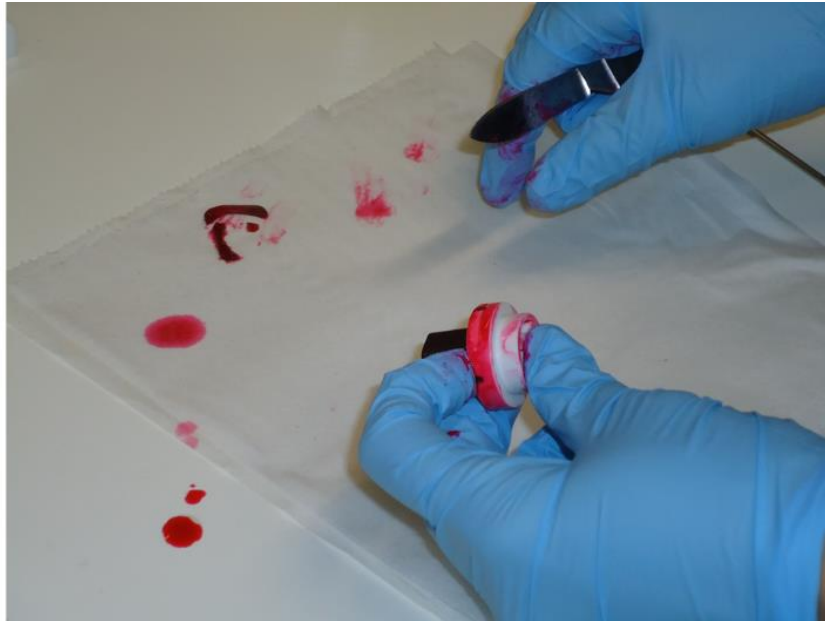
Figure 2.9 Trimming the resin blocks.

A block that has been trimmed with a blunt knife. The pieces of excess resin can be seen on the paper towel. They need to be removed prior to sectioning so that the chuck can be securely attached to the HREM machinery.

The resin blocks were baked at 95 – 100°C overnight, to ensure the blocks were completely hard before sectioning. They were then stored at 4°C (Figure 4.10).
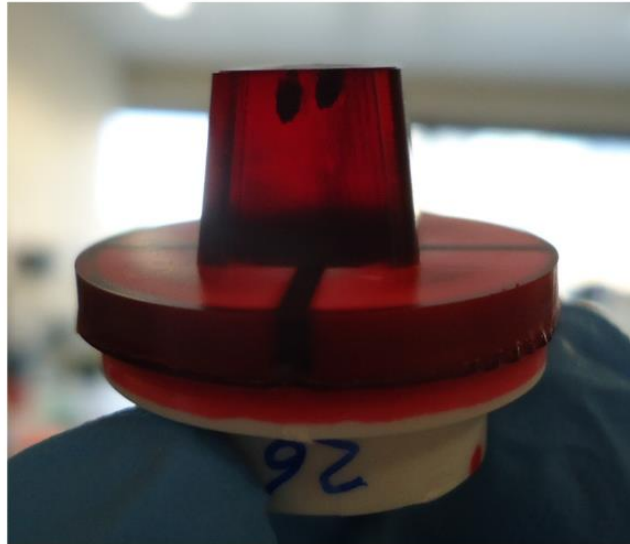
Figure 2.10 A finished resin block.

Two E15.5 hearts are visible in the resin block, with their apex pointing downwards, as the block has been inverted after removal from the mould.

2.4.6 Harvesting and preparing Adult hearts for HREM

This protocol was adapted from the published protocol from https://dmdd.org.uk/hrem/, and personal communication with Dr T Mohun.

The adult mouse was transferred into PBS at 37°C after being culled, and the thorax opened by sternotomy. The thymus and restrictive attachments, such as the diaphragm, were removed. The great vessels were isolated and lifted out of the chest cavity with the heart. To aid exsanguination, any clots forming in the distal vessels were removed, and the heart was gently agitated whilst submerged in the warm PBS. Further excess tissue not required for examination was removed before transferring to 4% PFA for 30 minutes.

The dissected hearts were washed extensively in tap water to remove the PFA. The water was changed every 10-15 minutes until the hearts appeared clear of blood, and then transferred to 4% PFA.

The same protocol was used for dehydration, infiltration and embedding as that detailed for the embryo hearts (sections 2.4.2, 2.4.3, 2.4.4, 2.4.5), with the exception of methanol washes all being one hour steps as opposed to 30 minutes, and infiltration in JB4-dye mix was for a week. A finished adult heart is shown in figure 2.11.
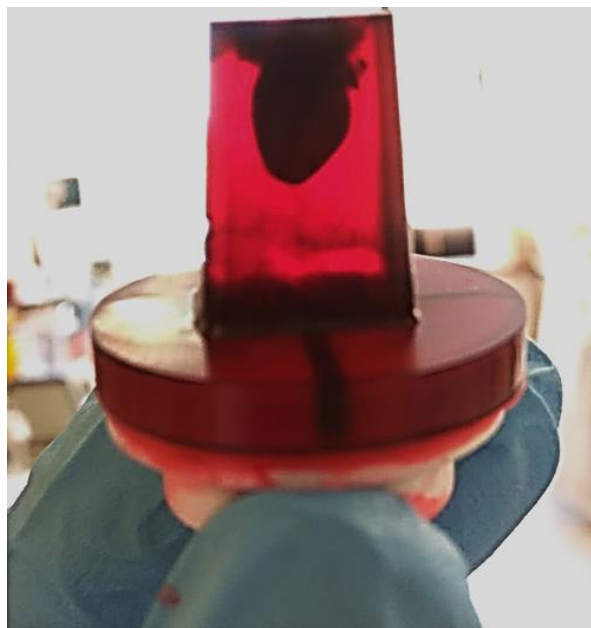
Figure 2.11 Example of an embedded adult heart.

Due to the larger sized heart, this required a larger 17mm depth custom mould.

2.4.7 Sectioning HREM blocks

Sectioning was carried out at the Crick Institute in T Mohun's laboratory. The setup of equipment is shown in figure 2.12.
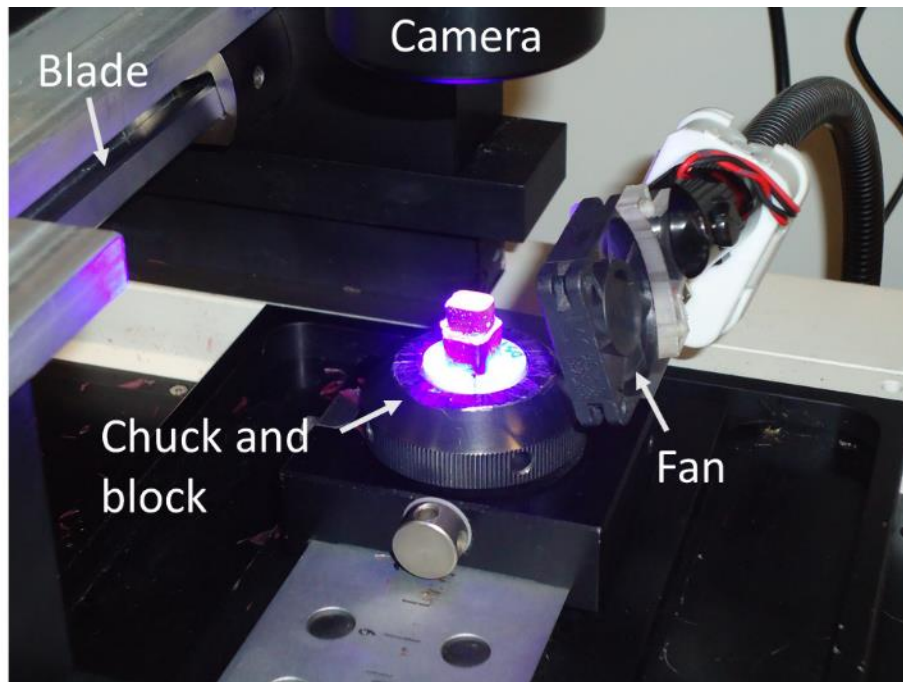
Figure 2.12 Set up of the HREM sectioning equipment.

The camera sits directly above the block and the blade moves forwards and backwards to cut each slice. A fan is required to ensure that the cut section does not remain on the block and obscure the picture. Photograph taken on a visit to T Mohun's laboratory at the Crick Institute.

The top corner of the block was marked to maintain the orientation of the block images once they have been sectioned (Figure 2.13). An initial photograph was taken of the graticule, to calculate the pixel size for image processing. The block was then sectioned (~2 μm thick sections).
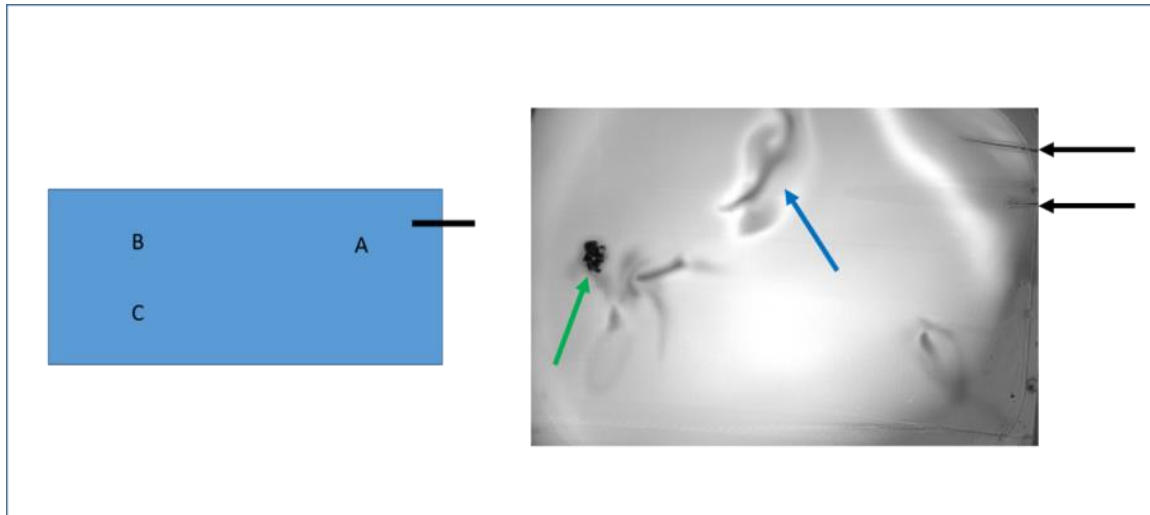
Figure 2.13 Orientation of the hearts in the resin block.

The blue block represents how the three hearts were orientated in the block in relation, to the score mark made in the top right corner. A photograph of a real block is shown. Black arrows indicate the scoring visible in the top right hand corner of the block made prior to sectioning. This indicates that heart A is in the top right hand corner. The blue arrow points to some imperfections on the block surface and the green arrow indicates some tissue that is already visible at this top level of the unsectioned block.

2.4.8 HREM Image processing

2.4.8.1 Preparation of Images

Processing was carried out using Graphic Converter 9 (free software www.lemkesoft.de), Photoshop (Adobe, Knoll and Knoll 1990), ImageJ (free software https://imagej.net), Guideliner (free software https://www.dsd.net/prod/mac/guideliner.php), Quicktime Player (free software[514]) and Osirix ([515]) on a Mac pro. F Prim at the Crick Institute taught me how to carry out the image processing.

Graphic Converter 9 was used to create a video from all the section images. Any substandard images were identified on the video (viewed in Quicktime) and

removed. The remaining images were then imported into ImageJ, and converted to greyscale 8 bit. The proportional scale was then reduced by 50% or less to produce a smaller sized data set.

A Photoshop macro was created, to optimise the contrast between the resin and tissue. The three stages required in the macro were crop, adjust image levels and adjust image curves. The Guideliner app was used to guide image cropping. ImageJ was used to invert the images and reduce the final stack size to around 500-700Mb.

2.4.8.2 Conversion to 3D Model

The ratio of pixels per 1µm was calculated using ImageJ, based on the initial image taken of the graticule. This was used to calculate a scaling factor as shown below, to ensure that the images remain proportional in 3D.

Measurement from graticule:

5500 equates to 5235pixels = 1.05062µm per pixel.

Calculation of scale factor:

$$\text{Scale Factor for 250Mb:} \quad \frac{\sqrt{250}}{\text{Cubic}}$$

For a 450Mb file size, the scaling factor is:

√250/450

= 0.745

Conversion to cubic data:

The image needs to be reduced to 74.5% of its original size. The x and y axis in Osirix was 1, as both were 1 pixel in size. The z axis value, to produce cubic data was calculated as below.

Z proportion = $\underline{\text{section thickness}}$ x   scaling factor.

                Resolution in um/pixel

            = $\underline{2.6 \times 0.745}$

                1.05062

            = 1.844

This proportions used in Osirix were therefore x = 1, y = 1, z = 1.844

The stack was imported as a DICOM into Osirix. The following setting were applied in the 2D viewer; WL/WW = full dynamic, CLUT = No CLUT, Opacity = Linear table and BW / RGB = B&W. The x, y and z values were calculated were entered to produce a 3D image which could be rotated and cropped as necessary.

2.5 Mouse Phenotyping

This section covers phenotyping of both adult mice and embryos, for cardiac and extra cardiac features. All phenotyping in embryos was carried out blind to genotype. In most cases, I knew the genotype of adult mice because they were either heterozygous females being culled to collect embryos, or previously genotyped mice that had to be culled because of welfare issues.

2.5.1 Embryo Phenotyping

I devised a protocol based on those used by EMPReSS [516] and the IMPC [214] for use in the Deciphering Mechanisms of Developmental Disorders (DMDD) pipelines [516]. The protocol I constructed included a more detailed analysis of the heart, based on the features expected at the specific gestation [224]. The expected cardiac development at the stages to be studied are outlined in appendix C. A single protocol was used for embryo phenotyping given that there are no significant differences between male and female anatomy at E15.5, with the exception of the reproductive systems [517]. Embryonic development was staged by measurement of the interdigital web [518]. A copy of the protocol can be found in appendix D.

Cardiac anatomy was assessed using both the raw axial section images, and the 3D reconstructions. Abnormalities were confirmed on orthogonal 2D imaging and 3D reconstructions in Osirix.

2.5.2 Adult Phenotyping

Adult mice were phenotyped to varying degrees. Detailed cardiac assessment was carried out on a group of 2 month old $Prkd1^{Em2/+}$ mice using MRI and a single $Prkd1^{Em2/Em2}$ heart was analysed by HREM.

Mice were examined for gross external malformations as part of normal mouse husbandry processes. Brief internal examinations were conducted on any mice that were culled for any reason. This consisted of an examination of all organ systems with the exception of the brain and spinal cord, by visual inspection only. The phenotyping form used can be found in appendix E.

2.6 MRI Imaging of Adult Mice

2.6.1 Harvesting and preparing Adult hearts for MRI

Surplus two month old male *Prkd1*[Em2/+] mice were imaged using MRI. Mice were sacrificed with overdose of carbon dioxide, and death confirmed by onset of rigor mortis. The mouse was then stored at -20°C until required for scanning.

2.6.2 MRI Protocol

The protocol detailed below produced images with a resolution of 50µm. All imaging was carried out on a 7 T Bruker Avance III (Ettlingen, Germany) MRI system. The protocol was devised by M Prior (Senior Research Fellow, University of Nottingham).

Mice were positioned in the scanner on the animal bed and a small, curved receive-only coil was placed over the heart area for data acquisition. Signal excitation was carried out by a 72 mm internal diameter volume coil. A scout image was acquired showing three orthogonal planes through the chest region to help align the subsequent scans with the heart.

For cardiac imaging, data were acquired using a 3D fast spin-echo (RARE) method with the following parameters: echo spacing of 12 ms, repetition time of 1000 ms and an echo train length of 8 echoes. The phase encoding start position was shifted to allow an effective echo time of 24 ms. Data were acquired with a field of view of 26 x 23.3 x 10.5 mm using 520 x 466 x 210 data points, respectively, resulting in an isotropic image resolution of 0.05 mm.

The method was reliant on the restricted field of view of the surface coil to limit the area that had to be covered by the data acquisition and thereby reduced the total acquisition time. However, it was necessary to place a 7 mm thick saturation band on the edge of the field of view opposite to the coil to reduce the

extent of wrap-around artefacts in the final images. Fat suppression was also employed to reduce image artefacts.

The acquisition time was 3 h 54 min for one image volume and a total of five image volumes were collected from each mouse. This resulted in a total imaging time for each mouse of 19.5 hours. It was anticipated that field drift may occur over such a long acquisition time and that image blurring might be apparent in such high resolution datasets. To allow for correction of this, the acquisition was split up into 5 volumes, so that the acquisition software could recalculate the field value at the start of acquisition for each image volume. Data from five images volumes were added together following rigid body alignment carried out with the FLIRT function in FSL (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki).

2.6.3 Cardiac Assessment of MRI Scans

ITK-Snap software was used for analysis http://www.itksnap.org. Review of orthogonal imaging (Figure 2.14) was carried out by myself and Dr B Erhayiem.
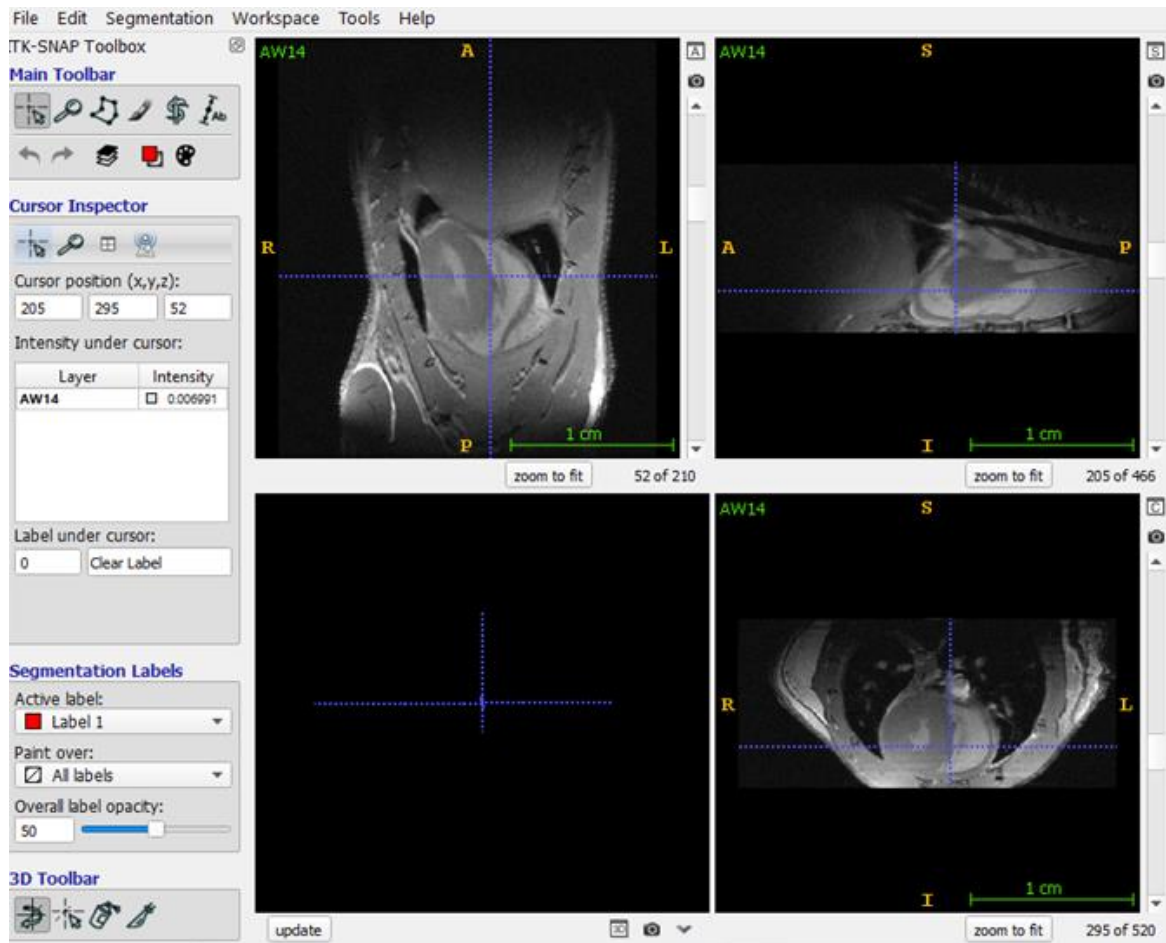
Figure 2.14 MRI images of adult mice.

ITK-Snap was used to view the MRI images produced. It provides simultaneous orientation in axial, sagittal and coronal planes allowing easy identification of structures.

In addition to assessment of cardiac morphology, we took the following measurements:

In the short axis, at the mid-ventricular level where the papillary muscles are present:

- RV free wall thickness
- LV anterior, posterior and lateral wall thickness
- Diameter of the IVS
- Internal diameter of the LV, AP and transverse measurements

Measurements of the aortic root were taken at the level of the sinus of Valsalva (Figure 2.15):

- Three measurements across the commissures
- Diameter measure on long axis image (coronal plane).
- Diameter measure on perpendicular long axis (sagittal plane).

Measurements of the ascending aorta

- Diameter of aorta at level of pulmonary artery in long axis view.
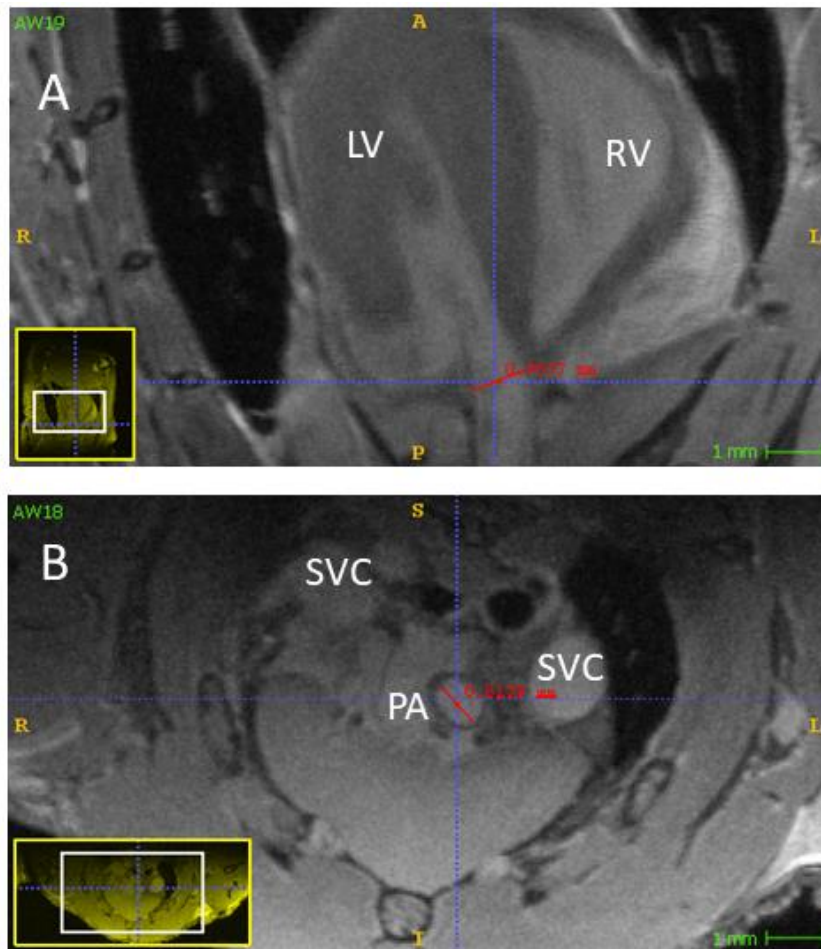
Figure 2.15 Aortic measurements using MRI.

Short and long axis views of *Prkd1*Em2/+ mouse illustrating some of the measurements made.

A. Measurements of the aortic root were taken at the level of the sinus of Valsalva. The maximum measurement was taken from the long axis image (coronal plane). The aortic root at this level was also assessed in the sagittal plane. In addition 3 measurements were taken across the cusps.

B: Diameter of the ascending aorta. The widest diameter of aorta was measured at level of pulmonary artery in the short axis view.

LV: Left ventricle, PA: Pulmonary artery, RV: Right ventricle, SVC: Superior vena cava.

2.7 Disposal of Mice and Tissue

Remaining mouse tissue was stored at -20c in the Bio Support Unit and MRI Unit and then disposed of by the staff there. Tissue no longer required that had been removed from the unit was double bagged and incinerated as per University protocol.

2.8 Identifying Pathogenic CHD Mutations for Feedback

Confirmation of pathogenic mutations identified in known CHD genes through WES [12] was carried out by PCR. The predicted consequence was confirmed using the Ensembl Variant Effect Predictor (VEP) [493]. HGMD [519], the Leiden databases (LOVD) [520], ClinVar [521] and Pubmed (https://pubmed.ncbi.nlm.nih.gov/) were reviewed for additional information about each variant. Clinical records at Glenfield hospital were examined to determine if the phenotype and family history were consistent with the mutation.

If the individual had not already received a genetic diagnosis since recruitment to the study, and if they had consented to receive results, the information was relayed by Dr Bu'Lock (Consultant Paediatric Cardiologist) and I together, or by Dr Bu'Lock alone. The relevant information required for Clinical Genetics review and testing in an NHS laboratory was provided.

PCR was carried out as described in section 2.2.1.2, followed by PCR clean up (2.2.6) and sequencing (2.2.7). A higher annealing temperature of 58.7°C and a slightly longer extension time of 40 seconds was required for PCR to confirm the mutation c.992A>G, (p.Asn308Asp) in *PTPN11*.

Touchdown PCR was required to confirm *PTPN11* mutation *c.188A>G, (*p.Tyr63Cys*).* An initial denaturation step was carried out at 95°C, for 2 minutes. Denaturation was for 30 seconds at 95°C, annealing was for 30 sec at

62.5ºC, with a decrease of 0.5ºC per cycle, followed by 30 seconds at 72°C. 14 cycles were required before a further denaturation step for 30 sec at 95ºC. Annealing was for 30 sec at 55.5ºC, extension was for 30 seconds at 72°C and the cycle was repeated 19 more times, followed by a final extension and infinite hold.

Chapter 3 Using Whole Exome Sequencing to discover novel CHD genes

This chapter documents my involvement in the following publication:

Exome Sequencing reveals distinct genetic architectures for syndromic and non-syndromic congenital heart defects, and three novel disorders [12].

Alejandro Sifrim, Marc-Phillip Hitz, **Anna Wilsdon** et al. Nature Genetics September 2016.

A copy of this paper can be found in Appendix F. This study makes use of data generated by the DECIPHER community. A full list of centres who contributed to the generation of the data is available from https://decipher.sanger.ac.uk and via email from decipher@sanger.ac.uk. Funding for the project was provided by Wellcome. As per the Decipher citation agreement, all recruiting centres for individuals identified through DDD were contacted.

My contribution to this paper included phenotyping, method development and data analysis and writing.

Through bioinformatic analysis of a large cohort of individuals with CHD, we identified three novel CHD genes. This chapter describes my contribution to this work, and one of the new genes we described (*CHD4*). It is included here to present a complete story for the following two chapters, which describe the experimental work I carried out on two further novel genes we identified (Chapters 4 and 5: *PRKD1*, Chapter 6:*CDK13*).

3.1 Whole Exome Sequencing in Congenital Heart Disease

Traditional gene discovery in CHD relied upon identification of new genes in pedigrees with Mendelian patterns of inheritance, followed by targeted sequencing in cohorts with an identical phenotype. Whilst this worked well, these families are rare in CHD. In most instances, there is only a single individual in the family with CHD. This method also predisposes to ascertainment bias, by only including individuals with distinct phenotypes. This means we may not discover the full spectrum of abnormalities associated with that particular gene.

More recently exome sequencing projects in large cohorts with diverse phenotypes, such as the DDD study, have proved successful at identifying novel disease genes in neurodevelopmental disability [522]. Prior to the publication of Sifrim et al. [12], a similar approach using WES in unselected CHD phenotypes to investigate single gene causes of CHD had been tried. Homsy et al. [14] and Zaidi et al.[16] demonstrated the contribution of de novo damaging variants to syndromic-CHD (S-CHD), an important role for histone modifying genes, and identified the novel CHD gene *RBFOX2*. They found that 20% of S-CHD and 2% of NS-CHD could be explained by de novo mutations in genes highly expressed in the developing mouse heart. These two studies went some way to explaining the low recurrence risk in S-CHD, but did not identify any significant results for NS-CHD, which accounts for the majority of people with CHD. These two papers are considered further below.

3.1.1 De novo mutations in histone-modifying genes cause congenital heart disease.

Zaidi et al. was the first study to use WES in a large cohort of unselected CHD and identify that de novo mutations were an important cause of CHD.[16]. They studied 362 trios, who were selected based on an index with severe CHD, no known genetic diagnosis, and no family history of CHD. The control group comprised of trios of parents and unaffected siblings collected as part of an

autism study. They compared the mutations found in cases and controls in a gene set referred to as the high heart expression (HHE) genes. These 4,169 genes are mouse genes with human orthologues that were in the top quartile of expression in an E14.5 mouse heart based on RNA sequencing. There was a significant excess of de novo protein altering mutations in the HHE genes in individuals with CHD, compared to controls.

To isolate those mutations that were most likely to be deleterious, they increased the stringency of variant filtering which resulted in a higher odds ratio (OR) of de novo mutations in these genes in the CHD group, compared with controls. Similar increases in the OR were obtained when using in silico programmes, such as polyphen2, gene expression levels in the developing heart and RNA expression at E9.5 to predict pathogenicity. They also considered CHD by subtype. The OR of de novo mutations in CHD compared to controls was increased for specific CHD groups including left ventricular obstruction, conotruncal defects and heterotaxy.

Evidence that these predicted pathogenic mutations are significant in the CHD cohort is supported by the finding that there was no increased burden in the number of de novo mutations overall, or in genes in the lower quartile of heart expression in the CHD cohort. There was no significant difference in the mutation burden in genes expressed in other tissues such as the adult liver, brain, heart and lung, when compared with the control cohort.

Zaidi et al. identified that abnormal histone modification is likely to be important in the pathogenesis of CHD. This theme has been seen again in subsequent studies [12, 14]. Predicted damaging de novo mutations were identified in eight genes that play a role in histone modification in the CHD cohort, some of which are already linked with CHD. The eight genes were *MLL2* (Kabuki Syndrome), *WDR5*, *KDM5A*, *KDM5B* (AR mental retardation, including CHD), *RNF20*, *UBE2B*, *CHD7* (CHARGE Syndrome) and *USP44*. They are involved in production, removal or reading of methylation of H3K4 (histone H3, at lysine 4). This is a

well-studied acetylation/methylation site on histone 3, which plays an important role in transcription [523]. Genes involved in H3K4me modifications, including *KDM6A* (X-linked Kabuki Syndrome) were the only gene set that was significantly enriched in the CHD cohort, based on an excess of mutations compared to that expected by chance, and absence of mutations in these genes in the control population. There was no consistent CHD genotype-phenotype association, but the majority had S-CHD. This is not surprising given that several of these genes are known to cause S-CHD already.

This study demonstrated the benefits of an untargeted phenotype approach in expanding genotype-phenotype correlations. For example, an atypical non-syndromic phenotype was seen in an individual with a *CHD7* mutation, who was reported to have none of the significant features of CHARGE Syndrome. They were also able to retrospectively diagnose an individual with Kabuki Syndrome, who had a mutation in *MLL2.* This approach may have resulted in an earlier diagnosis for this individual.

A number of other genes harboured multiple mutations, and may be significant. *SMAD2* demonstrated more mutations expected than by chance and compared to the control data set, as did *SUV420H1*, *MED20*, *HUWE1*, *CUL3*, *NUB1* and *NAA15*. Both individuals with *SMAD2* mutations had dextrocardia, in keeping with its known involvement in left-right organisation [524]. The authors felt that this could explain the individuals' phenotypes.

13 further probands were identified with mutations in a predefined set of 277 genes known CHD genes. This included two genes that had been identified by CNV studies (*RAB10*, *BCL6*) [18], providing additional evidence of their role in CHD. An individual with a severe conotruncal defect was found to have a mutation in *NF1*. Mutations in this gene cause Neurofibromatosis type 1, where this type of CHD is unusual. It was not clear from the paper whether the person had other signs of neurofibromatosis or not.

Overall, Zaidi et al. identified an important contribution of de novo mutations to CHD, but the exact contribution to S and NS-CHD was not defined. One limitation is that the cohort was made up of individuals with severe CHD only. Additionally the HHE gene list was based on mouse expression, and we know that mouse genes do not always translate to similar phenotypes in humans.

3.1.2 The role of De novo mutations in Congenital Heart Disease with Neurodevelopmental and other Congenital Abnormalities.

The next major publication using WES in CHD came from Homsy et al [14]. This WES study includes the 355 trios from the previously described Zaidi et al [16], in addition to other CHD trios with no known genetic diagnosis (n=1213). NS-CHD and S-CHD are considered separately, and are referred to as isolated CHD, and CHD with neurodevelopmental disability (NDD) and/or other congenital abnormalities. The control population of 900 healthy trios were collected as part of a study into the genetics of autism. They compared the actual and expected numbers of de novo mutations in each gene by variant class to identify significant genes, the same method reported by Samotcha et al. [494] and used in our publication described in this chapter [12].

Like Zaidi et al.[16] before them, they demonstrated an important contribution of de novo mutations in CHD. A significant enrichment (1.4 fold) of predicted deleterious de novo LoF (loss of function) and missense variants was found in the CHD cohort [525]. All phenotype groups except heterotaxy, displayed this trend. Heterotaxy is a diverse group and includes a number of known recessive causes, so this is perhaps not surprising despite the results from Zaidi et al [16]. In addition, there was a 2.4 fold increase in deleterious mutations in 4420 genes in the HHE list of genes [16]. This was not seen in controls.

Homsy et al. then considered the mutational burden in S-CHD and NS-CHD, to see if there was a difference. There was a significant enrichment of de novo predicted damaging LoF mutations in S-CHD. Enrichment was higher when they considered genes in the HHE list [16] in the most severe phenotypes (CHD with both congenital anomalies and neurodevelopmental disability). They did not demonstrate any significant enrichment in NS-CHD, so the genetic mechanism underlying this group remained unknown.

21 genes were identified with more damaging de novo mutations than expected by chance in the CHD cohort only, and not present in controls. These genes are thought likely to contribute to the pathogenesis of CHD. In keeping with this theory, a number of genes were identified that were already known to cause CHD. These genes are shown below in table 3.1.

Gene	Syndrome	CHD
PTPN11	Noonan Syndrome	Yes
KMT2D (MLL2)	Kabuki Syndrome	Yes
RBFOX2		
KDM5B	AR Mental Retardation	
KRT13	White sponge naevus	
MYH6	ASD, Dilated and Hypertrophic Cardiomyopathy	Yes
CAD	AR epileptic encephalopathy	
NAA15	AD mental retardation	
SMAD2	Identified by Zaidi et al as potentially pathogenic[16]	
RABGAP1L		
POGZ	White Sutton Syndrome	Possibly
JAG1	Alagille Syndrome	Yes
GANAB	AD polycystic kidney disease	
DTNA	LVNC with/without CHD	Yes
PPL		
CHD7	CHARGE Syndrome	Yes
ZEB2	Mowat Wilson Syndrome	Yes
FBN1	Marfan Syndrome	Yes
CHD4	Unknown, but confirmed as CHD gene by Sifrim et al.[12]	Yes
AHNAK		
NOTCH1	Aortic Valve Disease and Adams Oliver Syndrome	Yes

Table 3.1 Genes identified with de novo mutations.

Genes with damaging de novo variants identified in the CHD cohort only, the associated syndrome and whether CHD is a feature or not. A number of these genes are associated with neurodevelopmental disorders.

PTPN11, *KMT2D* and *RBFOX2* had significantly higher numbers of de novo damaging mutations than the others listed in this table. *RBFOX2* is the only novel gene of the three. Three individuals from this study, and an additional patient with a deletion of *RBFOX2 [526],* all had hypoplastic left heart syndrome (HLHS). The authors suggest that given this gene is needed for correct

development of the zebrafish heart [527], and is involved in epithelial to mesenchymal transformation which is linked to HLHS [528, 529], it is likely to be a true CHD causing gene. The mouse model however does not have CHD, although *Rbfox2* is reported to be expressed in the heart [530]. Further research suggests it may play a role in heart failure and [531] and that it is important in controlling the transcriptome of HLHS hearts [532].

Homsy et al. also considered which gene ontology terms were enriched in the CHD cohort [14]. A number of categories were identified as significant based on an enrichment of damaging de novo variants, and included anatomic structure morphogenesis, cardiovascular system development, neurodevelopmental abnormality and chromatin modification. Chromatin modification was highlighted particularly, as this is in line with the results of Zaidi et al [16]. This group of genes were still significant when the previously reported patients from Zaidi et al. were removed from the analysis, suggesting this finding is not just associated with the initial Zaidi et al. cohort. They suggested these genes can cause a high risk of CHD.

More genes were identified as potentially significant based on the number of mutations found involving that particular gene. In the CHD only cohort (NS-CHD), damaging de novo mutations were identified in *SMAD2,* which has been highlighted previously [16]. *KANSL1* was also identified. Mutations (and deletions) in *KANSL1* are known to cause AD Koolen-De Vries Syndrome which includes CHD [533]. This expands the phenotype of this syndrome. Both *KDM5A* (demethylase of H3K4me2/3) and *HLTF* (role in chromatin modification) [534] may be significant. *HIRA, UBE2B, RNF20* and *CTR9* were also highlighted.

In both the CHD and neurodevelopmental disorder cohorts, de novo damaging mutations were identified in *ARID1B, CHD4, TLK2, KDM5B, KMT2C, KMT2D, POGZ, KDM6B, NSD1, WHSC1, ASH1L.* Of these genes *KANSL1, KMT2D, KDM5B* and *NSD1* are associated with S-CHD*. POGZ* causes White-Sutton Syndrome and a single individual has been reported with CHD [535]. *KMT2C* and *WHSC1* have

links with methylation and *KDM6B* has a role in heart development. A single individual has been reported with an abnormality in the SVC and the *KDM6B* monoallelic Neurodevelopmental Disorder with Coarse Facies and mild distal Skeletal abnormalities [536]. More individuals with damaging mutations in these genes will be required to determine if they definitely cause CHD, as was the case with *CHD4,* which we later showed to be a true CHD gene after it reached genome wide significance [12].

The mutational burden in both S-CHD and NS-CHD was compared by Homsy et al, after excluding patients with a known genetic diagnosis identified from the sequencing results. There was enrichment of damaging de novo mutations in genes from the HHE list in the S-CHD cohort. This was 3 fold in those with CHD and NDD or congenital abnormalities, and even higher in those with CHD, NDD and other congenital abnormalities (4.7-fold). There was no enrichment in the NS-CHD or control groups. This confirms the findings of Zaidi et al.[16], that de novo mutations contribute to S-CHD, but also suggests that we are more likely to find de novo mutations in those with the more severe presentations of S-CHD. Homsy et al. predicted that de novo mutations in these HHE genes contributed to CHD in 20% of patients with CHD, NDD and congenital abnormalities, 10% of CHD with NDD, 6% of CHD with congenital abnormalities, and only 2% of NS-CHD. The authors suggested that reduced penetrance in the same genes could lead to both S-CHD and NS-CHD.

We have seen a number of known NDD genes being highlighted as potentially significant in individuals with CHD (Table 3.1). A number of the genes showing enrichment for de novo mutations, were expressed in both the developing heart and the developing brain, suggesting a shared underlying genetic pathogenesis of CHD and NDD. To assess this possible relationship further, Homsy et al. investigated the overlap of genes with de novo damaging mutations in individuals with NDD and CHD, and 1161 genes that were identified with damaging de novo mutations in a population with NDD, but no CHD [14]. 69 genes were shared between the two cohorts and were significantly enriched for de novo damaging mutations (2.6-fold enrichment). They were also significantly

more likely to be in the top quartile of heart and brain development expressed genes. A number of these genes are important in cardiac development pathways including the *WNT* and *NOTCH* pathways. This led to the important suggestion that there is a common aetiology for both CHD and neurodevelopmental disorders, but with variable expressivity of the genes involved.

This was the first study to really consider the contribution of damaging mutations in both S-CHD and NS-CHD, and identify a genome wide significant novel CHD gene. The contribution of de novo mutations in S-CHD was confirmed, but the genetic mechanism behind NS-CHD remained unexplained. They also confirmed the role of chromatin modifying genes in CHD and proposed an overlap between the aetiology of CHD and neurodevelopmental disorders.

3.1.2 Current use of WES in CHD

Unfortunately as things stand, the majority of individuals with CHD do not receive a genetic diagnosis, but studies such as these suggest that WES may be successful in identifying novel CHD genes and diagnoses for patients [537]. Testing within the NHS has moved away from single genes, to panels of multiple relevant genes. This is likely to move towards WGS in the future. In the meantime, it is important that we try and identify novel CHD genes, so that more extensive panels can be provided for use on a clinical basis. The NS-CHD population is the most underserved; they make up the majority of individuals with CHD, but are less likely to receive a genetic diagnosis. Understanding the aetiology of NS-CHD is therefore especially important.

3.2 Aims

1. To accurately phenotype a cohort of individuals with CHD.
2. To produce a list of high confidence genes known to cause CHD in humans based on evaluation of the current available literature.
3. To identify novel CHD genes using WES and established statistical models, which can be put forward for functional and animal studies.
4. To identify pathogenic mutations for feedback to individuals involved in the study locally.

3.3 Results and Discussion

3.3.1 Phenotyping the Cohort for Analysis

WES results from 1891 individuals (1365 trios) with CHD were analysed in Sifrim et al [12]. The majority were NS-CHD (NS-CHD = 1281, S-CHD = 610)) (Appendix G, supplementary table 1). Appendix H (Supplementary table 2) indicates the number of trios and which centres they were collected from.

The individuals with CHD had been collected by a number of different groups and as a result both the cardiac and non-cardiac phenotypes required standardising before any analysis could be carried out. A number of phenotyping systems already exist. These include the Human Phenotype Ontology (HPO) [538] and the European Paediatric Congenital Cardiology Codes (EPCC)[539]. HPO classification has been used as part of Online Mendelian Inheritance in Man (OMIM)[540] and the DDD project [496] and is familiar to Clinical Geneticists. A proportion of the cohort included in this paper are DDD participants, which have HPO coding already. Cardiologists however are more familiar with the EPCC codes, a system designed for phenotyping only cardiac defects.

We transferred all CHD phenotypes into HPO and EPCC terminology. Extra-cardiac information was classified using HPO terms. Additionally, all individuals

were allocated to either syndromic (S-CHD) or non-syndromic CHD (NS-CHD) groups. Additional medical information was gathered from the medical notes when there was insufficient detail, or I consulted the recruiting clinician for the Glenfield Hospital cohort. Information on other contributing factors, family history and other significant information was also coded. We discussed classification systems with Dr Bu'Lock.

We categorised the phenotypes using the following guidelines.

1. Classify the type of CHD as accurately as possible.

Problems arose due to ambiguous cardiac phenotyping in the raw data. For example, some individuals were recorded as having an abnormality of the heart with no further information. Others were recorded as having pulmonary stenosis, which could represent either pulmonary artery stenosis or pulmonary valve stenosis. I reviewed the medical records to confirm the specific lesion before deciding on a final code.

2. Use the most specific IPCC and HPO term possible

The phenotype codes must match the patient's phenotype as closely as possible to facilitate accurate interpretation of the genotype results. Both the HPO and EPCC systems are hierarchical, with a less specific term at the top and an increasing number of subdivides to reach a more specific term. We used the most specific term possible. We also needed to be able to collect phenotypes together in large enough groups to allow a robust and sufficiently powered analysis. Highly specific groups containing only a few individuals, or large groups of individuals with a non-specific phenotype are not ideal. Grouping defects by heart field or lesions with similar developmental mechanisms was thought to be too difficult. We therefore assigned each individual with a detailed and a second more generalised term, such as secundum ASD and interatrial communication, using both HPO and EPCC terminology.

Individuals with multiple cardiac defects and complicated defects were difficult to categorise. For example, Tetralogy of Fallot (TOF) represents a combination of abnormalities. It is broken down into VSD, pulmonary infundibular stenosis, an overriding aorta and right ventricular hypertrophy. The primary pathogenesis is the septal abnormality and anterocephalad deviation of the outlet septum[541], so it may be appropriate to classify it as VSD mainly. However, this is clearly not the same as an individual with an isolated VSD. TOF has different developmental origins from isolated pulmonary stenosis. We therefore coded this lesion as TOF, rather than its separate components.

However, other diagnoses should be categorised as multiple defects to distinguish them from each other. For example, transposition of the great arteries (TGA) with an intact ventricular septum is more common and less likely to be inherited than TGA with a VSD. It would be appropriate to classify these two subtypes of TGA separately as they do appear to be two separate entities. Supplementary table 3 in Sifrim et al [12]. Details the results of the phenotyping we carried out.

3. Classify the individual as either S-CHD or NS CHD

S-CHD was defined as individuals with CHD, who in addition had a distinct facial gestalt and/or at least one extra-cardiac abnormality and/or developmental delay/disability. NS-CHD was those individuals with CHD and no other abnormalities.

Heterotaxia could be classified as either S-CHD or NS-CHD. Ciliopathies are a known cause of heterotaxy and result in a range of abnormalities including retinal degeneration, bronchiectasis, polydactyly, renal and hepatic abnormalities. A transient node in the embryo is responsible for left right patterning and contains motile cilia, therefore a defect in a single cilia gene can result in multiple abnormalities including situs abnormalities in the heart [542].

It was decided to classify heterotaxy malformations as NS-CHD as these malformations can all be accounted for by one well established mechanism. Both isolated dextrocardia and dextrocardia with full situs inversus would both be classified as just heterotaxy. Laterality disorders are less common in our cohort overall.

The cohort comprised of 1365 trios, 518 were S-CHD and 847 were NS-CHD. The most common type of CHD was VSDs (19.4%), followed by aortic valve abnormalities (17.72%) and ASDs (14.86%). CoA, TOF, AVCD accounted for 10% of individuals or more each. Please see appendix I (supplementary table 3) for the full table of results of the phenotyping using HPO codes.

To facilitate discovery of novel CHD genes, phenotypes that might represent different genetic and non-genetic processes were removed from further analysis. This included patent ductus arteriosus (PDA), a common occurrence in preterm births. Often gestation at birth has not been recorded, so it is difficult to interpret if this is a normal finding or not. In addition, patent foramen ovale (PFO) was excluded. The foramen ovale normally closes soon after birth, but is found to be still patent in 20-35% of adults of different age groups [543]. Isolated arterial tortuosity, non-congenital aortopathies and cardiomyopathy were also excluded. In the main these are separate entities to CHD, although there is some overlap with CHD.

Although the Nottingham cohort had some information on maternal diabetes, medications and teratogens in pregnancy, this data was not consistently recorded in all individuals. It did not specify if the mother had longstanding or gestational diabetes, or what glycaemic control was like periconceptually. This made it difficult to exclude a genetic cause of CHD in most individuals, so a decision was made to include them in the study. It would also be extremely difficult to include all these environmental factors into the analysis alongside the genetic analysis.

12,031 control exomes with matched ancestry, were ascertained from the parents of probands without CHD from the DDD project (n=7301) and assumed healthy blood donors from the INTERVAL study (n=4730) [496, 497].

3.3.2 Defining the High Confidence CHD Gene List

For the WES analysis, we produced a list of genes we were confident cause CHD in humans for the following purposes:

- As a test list to see if the analysis picks up common genes such as *PTPN11* to allow necessary quality control.
- For comparison of mutational burden between the S-CHD and NS-CHD groups.
- To identify those with mutations in known CHD genes so that the cohort could be split into resolved and unresolved CHD.
- To identify pathogenic mutations in these genes which can be validated and fed back to that individual.

The list was derived from three primary sources:

- A list compiled by M Hitz causing S-CHD and NS-CHD in humans (n=234)
- Genes causing cardiac phenotypes in the DDG2P genes database produced by the DDD project (n=250) https://decipher.sanger.ac.uk/ddd#ddgenes [495]. The DDG2P list has been produced by Clinical Geneticists using information from OMIM, UniProt, Nature Genetics and American Journal of Human Genetics (http://www.ebi.ac.uk/gene2phenotype/) [496, 540, 544]. These genes are associated with developmental disorders.
- Genes identified in CHD-Wiki, a freely available site set up and curated by J Breckpot https://homes.esat.kuleuven.be/~bioiuser/chdwiki/ [545].

Genes causing a cardiovascular phenotype that we would not want to include in the study including cardiomyopathy and arrhythmia were removed. Perinatal lethal genes were included if they were X-linked or had variable expression that might result in milder phenotypes, however there were no antenatal or post mortem derived DNA samples in the cohort. Genes that cause a cardiac phenotype due to a metabolic process were also not included given that this is a different mechanism.

The genes were then split into two tiers based on the evidence available from the literature and previous animal models. Tier 1 genes are "confirmed" cardiac genes and tier 2 are "probable" and "possible" cardiac genes (Appendix J, supplementary table 19). This follows the same model used by the DDD project [495]. Each gene was annotated with inheritance pattern, whether it causes S-CHD or NS-CHD and the disease mechanism e.g. gain or loss of function. Evidence of links with the heart was collected from OMIM [540], the MGI mouse database [546] and Pubmed (https://pubmed.ncbi.nlm.nih.gov/). Reported mutations in association with PDA in a preterm infant were not counted, as this could represent normal physiology.

A list of 121 tier 1 genes (95 syndromic CHD, 24 non-syndromic and 2 genes that cause both syndromic and non-syndromic CHD) and 70 tier 2 genes (9 non-syndromic CHD, 61 syndromic CHD) was finalised (appendix K, supplementary table 20) [12].

3.3.3 De novo Mutations are an important cause of S-CHD, and NS-CHD to a lesser extent

The de novo analysis was based on a comparison of the actual rate of de novo mutations with the expected number under a null mutational model [494]. This considered all trios, as they did not need to be ethnically matched with a control population.

The number of de novo mutations was compared between S-CHD and NS-CHD and across three sets of genes;

- **CHD-associated genes.** These are genes with good evidence that they cause dominant CHD in humans.
- **Non-CHD developmental disorder (DD) genes.** These are genes known to cause developmental disorders, but not CHD derived from the DDDG2P list (see 3.3.2) [522].
- **All remaining protein coding genes.**

The results were as follows. Please also see Figure 1A in appendix F.

1. Across all trios, there was no difference in the burden of silent mutations in S-CHD and NS-CHD.

2. In the S-CHD group, there was a significant excess of de novo protein truncating variants (PTVs) (27 variants, OR 81, p=1.21x10^{-43}) and de novo missense variants (22 variants, OR 8.6 p=7.35x10^{-15}) in autosomal dominant (AD) CHD-associated genes. The difference between S and NS-CHD was most striking for PTVs, compared with missense variants. The observed difference between de novo PTVs in NS-CHD and S-CHD confirms the findings of Homsy et al. [14]. This burden additionally mirrors that observed in autism between individuals with and without intellectual disability [498].

3. In the NS-CHD group, there was a lower burden of de novo PTVs in AD CHD-associated genes (4 variants, OR 7.3, p=2.61x10^{-4}).

4. In the S-CHD group, there was a significant excess of de novo PTVs in non-CHD AD developmental disorder genes (12 variants, OR 18.4, p=3.49x10^{-13}). This implies that some of the known non-CHD DD genes may also cause CHD.

5. In both S-CHD and NS-CHD, there was a significant exome wide excess of de novo missense mutations, in all other protein coding genes. This suggests that there are further AD CHD causing genes to be discovered.

This confirms the findings of Zaidi et al. [16] and Homsy et al. [14], that de novo mutations are important in the pathogenesis of CHD. The role of de novo deleterious mutations had been predicted because of the low recurrence risk seen, and the likelihood that CHD results in reduced reproductive fitness.

The higher rate of de novo PTVs in S-CHD compared to NS-CHD, confirms the findings of Homsy et al. where they considered CHD with and without NDD. NDD is the most common clinical abnormality seen in combination with CHD. Subsequently Jin et al. analysed WES results from a cohort of 2871 probands with CHD (2645 trios) and found a relatively low rate of de novo mutations in NS-CHD [15]. This cohort includes the individuals analysed in Homsy et al. and is published by the same group. They found that 3.1% of NS-CHD is a result of de novo mutations. This means that the genetic aetiology of NS-CHD remains unexplained in the majority.

3.3.4 Inherited mutations may contribute to the pathogenesis of NS-CHD.

The inherited mutation analysis considered trios of European ancestry only, compared to an ethnically matched control population. The results were as follows. Please also see Figure 1C in appendix F.

1. There was no difference in mutational burden of silent mutations across the NS and S-CHD groups.
2. There was no difference in mutational burden of missense mutations across NS and S-CHD.

3. In the NS-CHD group, there was a significant excess of rare inherited PTVs in AD CHD-associated genes (17 variants, OR=2.67 p=1.1×10^{-4}). This excess was not present in the S-CHD group (p=0.3). This is a novel finding.
4. In the NS-CHD group, there was an exome wide excess of rare inherited PTVs (3318 variants, OR=1.08, p=1.51×10^{-5}) in the remaining protein coding genes (group 3 genes). This was not seen in the S-CHD group and suggests that there are other novel genes showing incomplete penetrance to be discovered.

For the first time, this study identified that different genetic mechanisms contribute to S-CHD and NS-CHD [12]. The inherited burden of reduced penetrance variants is a new finding. This could result from differences in coverage, ancestry and gender between the cohorts. However, principal component analysis (PCA) analysis has shown that both the S-CHD and NS-CHD groups have similar ancestry [12]. There was no significant difference in the PTV burden analysis between males and females. Variant calling used the same pipeline, although there were some differences between the sequencing parameters used to analyse all of the cohorts (sequencing depth and exome capture platform). If there were differences in the methodology that were significant, we should see differences in the burden of synonymous variants between the different cohorts. This was not detected.

Another obvious question to consider is whether a parent is labelled as unaffected, but does actually have CHD. This was considered to be unlikely as many types of CHD would declare themselves before adulthood, and especially in pregnant women. However, it is also not uncommon for a milder lesion to be diagnosed incidentally or in later life. In the Glenfield cohort, family history was discussed as standard and parents were referred for imaging if there was a history of CHD in themselves or they have symptoms suggestive of CHD. Within the Clinical Genetics clinic, some parents may be referred for echocardiograms. However, routinely screening asymptomatic parents raises ethical questions and is not done routinely in all settings.

Under-phenotyping was not the source of this result and the significant exome wide association of rare inherited PTVs remains even after removal of known CHD causing genes and developmental disability genes. This suggests that there are further novel genes to be discovered and incomplete penetrance is present in CHD.

Inherited PTVs were found in genes that cause S-CHD, in the NS-CHD population. The individual genes are shown in table 3.2 following a discussion of possible reasons that might explain why apparently damaging mutations in S-CHD genes were identified in individuals with NS-CHD, and inherited from apparently normal parents.

- **Have the genes been correctly categorised as syndromic or non-syndromic?**

The list of high confidence CHD genes was reviewed. The majority of the genes were clearly associated with either NS-CHD or S-CHD. A few could be classified as either. Deletions containing *ELN* cause S-CHD in the form of Williams Beuren Syndrome. However isolated mutations in *ELN* cause NS-CHD. In addition there is the possibility that disruption of the gene by abnormalities affecting promoter regions and other regulatory elements could also lead to a phenotype as shown for other genes [547, 548] and could produce the same syndromic phenotype as a deletion. Overall, no changes were made to the classification of genes in the list.

- **Does this mean that the patients have been incorrectly phenotyped, and a syndromic phenotype has been missed?**

Possible explanations for this happening include very mild abnormalities, and patients being phenotyped at a very young age, when abnormalities such as developmental delay or internal structural abnormalities would not necessarily

be immediately obvious. A number of these S-CHD genes can produce a mild and more subtle phenotype. We went back to the patient notes where possible to see if they phenotype was recorded accurately. A few Nottingham patients were subsequently found to have S-CHD rather than NS-CHD on review of up to date notes, and were moved in to the syndromic group. This did not change the results however and inherited damaging variants in S-CHD genes remained significant in the NS-CHD individuals.

- **Does this represent an expansion of the known clinical spectrum?**

As clinical testing has mainly been carried out on a gene by gene basis until recently, only patients with a convincing and syndromic phenotype were offered genetic testing. As a result, there is an ascertainment bias and fewer opportunities for us to discover the atypical, milder or non-syndromic phenotypes. Non-targeted sequencing studies like this allow an unbiased approach. It is possible that the phenotype is wider than previously thought and includes a non-syndromic or very mild presentation.

Gene	Associated Disease/Syndrome
ABCC9	Cantu Syndrome, a mild phenotype could be missed. Also causes isolated atrial fibrillation and cardiomyopathy.
ACTC1	NS-CHD. Isolated atrial septal defect (ASD), cardiomyopathy and left ventricular non-compaction.
COL1A1	Ehlers Danlos Syndrome, a mild phenotype can be missed. Dominant negative mutations cause osteogenesis imperfecta, without CHD.
ELN	Mutations are associated with NS-CHD supravalvular aortic stenosis and peripheral pulmonary stenosis and cutis laxa. Deletion of ELN produces a S-CHD phenotype (Williams Syndrome).
FBN2	Beals Syndrome, phenotype can be mild and contractures can be subtle at a young age.
FGFR2	Phenotype of the craniosynostosis syndromes can be mild, and is also usually caused by gain of function mutations. LOF mutations cause LADD syndrome which does not include CHD.
MYH6	NS-CHD. ASD and cardiomyopathy.
NOTCH1	Can cause both S-CHD and NS-CHD. Associated with Adams Oliver Syndrome and isolated aortic valve disease.
NOTCH2	Alagille Syndrome, phenotype can be mild and sometimes only isolated CHD is found in members of families with confirmed Alagille syndrome.
SALL4	Okihiro Syndrome, phenotype can be mild.
SOS1	Gain of function mutations cause Noonan Syndrome. Mild phenotype can be missed.

Table 3.2 S-CHD genes identified in individuals with NS-CHD.

AD S-CHD genes identified in the NS-CHD cohort, with explanation as to why they might be found in this population. Variants in recessive genes were not considered as they are unlikely to be pathogenic unless homozygous, or unless digenic mechanisms are known.

Overall, we ensured that phenotyping was conducted as accurately as possible, which suggests that the phenotypic spectrum of some of these genes listed in table 3.2, might be wider than previously thought.

Another interesting finding was that the S-CHD genes in which we identified inherited mutations in the NS-CHD group, do not overlap with the S-CHD associated genes where de novo PTVs were detected in the S-CHD group (Appendix F: figure 1D in Sifrim et al [12], and table 3.3 below). This could result if the S-CHD genes with de novo mutations caused a more severe phenotype and led to reduced reproductive fitness to the extent that inheriting a deleterious mutation was unlikely to be compatible with life, and in turn the S-CHD genes with inherited mutations produced a mild phenotype. Reviewing the genes and associated syndromes, this does not explain the division between the two groups of genes. For example, only de novo mutations were found in *NSD1*. LoF mutations and haploinsufficiency cause Sotos Syndrome which may be inherited. This could be a reflection of insufficient sample size.

Other explanations for this finding include both expanded phenotype of known syndromes, mild phenotypes and different inheritance patterns, such as biallelic mutations in a normally monoallelic condition. Testing in larger cohorts might provide more information.

Genes with de novo PTVs in S-CHD	
NSD1	Sotos Syndrome
KMT2A (MLL)	Wiedemann Steiner Syndrome
ADNP	Helsmoortel van der Aa Syndrome
CHD7	CHARGE Syndrome
KMT2D	Kabuki Syndrome
ANKRD11	KBG Syndrome
MED13L	Mental retardation and distinctive facial features with or without cardiac defects
EHMT1	Kleefstra Syndrome
CBL	Noonan Syndrome-like disorder with or without juvenile myelomonocytic leukaemia
KANSL1	Koolen-De Vries Syndrome
ASXL1	Bohring Opitz Syndrome
Genes with rare inherited PTVs in NS-CHD	
ELN	Cutis Laxa, supravalvular aortic stenosis
FGFR2	Craniosynostosis syndromes, LADD syndrome
ABCC9	Cantu Syndrome
MYH6	ASD and cardiomyopathy
ACTC1	Isolated atrial septal defect (ASD), cardiomyopathy and left ventricular non-compaction.
COL1A1	Ehlers Danlos Syndrome, osteogenesis imperfecta
SALL4	Okihiro Syndrome
NOTCH2	Alagille Syndrome
SOS1	Noonan Syndrome
FBN2	Beals Syndrome
NOTCH1	Adams Oliver Syndrome and isolated aortic valve disease

Table 3.3 Genes containing rare PTVs.

Genes and related phenotypes, in which either de novo or inherited rare PTVs were identified. None of these genes showed both inherited and de novo rare PTVs.

3.3.5 Comparison with other disease models

This analysis considered all CHD trios and compared them with other disease cohorts without CHD from published autism [498] and developmental disability studies [495] [499]. Across all groups there was no difference in the burden of de novo silent mutations, and there was no difference between those with NS-CHD and controls. However, the excess of de novo missense mutations in S and NS-CHD, is similar to the excess seen in autism with and without intellectual disability [498]. In addition, the excess burden of de novo PTVs in S-CHD is similar to that seen in severe developmental disability without CHD. The rate of de novo PTVs in autism alone was much lower.

There is a higher mutation burden with more severe phenotypes in NDD and autism, similar to S-CHD and NS-CHD. The more severe phenotypes of S-CHD and severe developmental disability show similar excess of de novo PTVs. Autism which would generally be classed as a milder phenotype than severe NDD, shows much lower burden, and NS-CHD and controls showed no excess burden. We see a similar higher burden of de novo missense mutations in S-CHD and autism with ID, compared with NS-CHD and individuals with autism alone. The increased rates of de novo mutations may reflect the reduced reproductive fitness associated with a more severe phenotype.

3.3.6 Gene Function and Pathway Analysis

We sought to identify enrichment of specific gene pathways and functions in CHD. The STRING PPI database [549] was used to determine where high confidence interactions (score > 0.9) existed between genes we had identified as potentially linked with CHD (genes with a false discovery rate of less than 50%). Genes related to chromatin modification, neural tube development, cardiac development and protein phosphorylation were significantly over represented (FDR<10%). Signalling pathways of note are the NOTCH-, IGF1-, HDAC Class II-, ERBB and NFKB- pathways. Many of the top ranking genes could actually be

mapped together as shown in figure 3.1 below. This level of network connection was greater than that expected due to chance alone (p=5.84x10^{-3}). Central hubs included known CHD genes such as *PTPN11*, *NOTCH1*, *SOS1*, *SMAD4* and *EP300*.
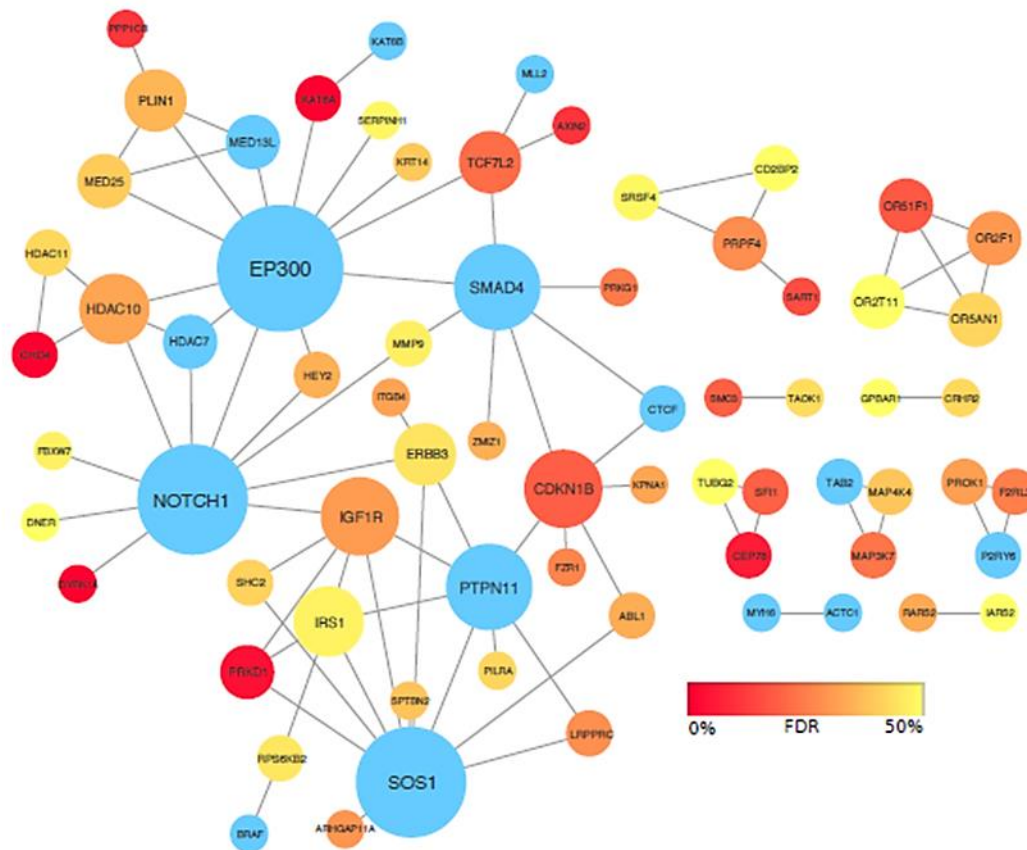
Figure 3.1 Protein-protein interactions between high confidence genes.

Reproduced with permission from Sifrim et al. High confidence interactions found in the STRING PPI database in top-ranking set of CHD-associated genes (false discovery rate <50%) in the integrated de novo and inherited variant analysis. Known CHD-associated genes are coloured blue, other nodes are scaled according to their association false discovery rate. The size of the nodes is scaled by its degree.

This confirms the previous reports of the importance of chromatin modifiers [14, 16] and also suggested HDAC class II genes are important, which is supported by mouse studies [550-552]. Genes related to neural tube development also

appear significant, whether this could relate to folate metabolism or not is unclear [553-555].

The highly connected map of genes we have produced can be used to identify additional genes that might be significant by exploring other related genes and proteins. Priest et al utilised a model previously used in cancer and NDD to identify protein networks that were enriched in individuals with AVSD, to identify novel loci [353].

3.3.7 Expanding the phenotype of known genes; wider phenotype associated CHD and NDD genes

We have been able to expand the phenotypic consequences of mutations in several genes. *TAB2* had only been associated with NS-CHD previously, our results suggest that it is also associated with S-CHD. In particular it appears that a number of genes thought to cause NDD only, also contribute to CHD. *DYRK1A* and *DDX3X* were associated with developmental disability only previously, but may also be associated with CHD. *DYRK1A* is associated with autism, but plays a role in regulation of cardiomyocyte hypertrophy [556] [557]. De novo mutations in *DDX3X* cause X-linked mental retardation, in some cases other additional congenital abnormalities are present [495] and individuals with CHD have been reported since our publication [558].

This highlights the usefulness of testing in cohorts without a narrow phenotypic spectrum. It is important to report instances where an atypical phenotype is found so clinicians can consider testing in individuals who wouldn't be considered "typical" for a certain condition.

3.3.8 The potential shared genetic aetiology between CHD and NDD

Like NDD, only a small proportion of individuals with CHD have a single mutation in a highly penetrant gene. We see a similar complex inheritance patterns and simplex cases in both disorders. Homsy et al suggested a shared mechanism between NDD and CHD [14] and we know that NDD is the commonest additional abnormality found in individuals with S-CHD. We have shown that genes associated with NDD also cause CHD.

Previously NDD was thought to be a consequence of haemodynamic compromise and hypoxia in individuals with CHD, especially in relation to complex surgery and time on bypass. It is common in more severe types of CHD and is thought to affect up to 50% of children who have CHD such as TGA, LVOTO and single ventricle defects [559-562]. However, this does not seem to account for all NDD seen with CHD, as NDD often predates surgical intervention [469, 561]. Homsy et al. showed that certain genes are likely to increase the risk of both CHD and autism [14]. A further recent study also concluded that the presence of CHD increased the odds of autism spectrum disorders by 30% (odds ratio 1.3), with the highest risk possibly being linked to ASD and VSDs [563]. These septal lesions can have important haemodynamic consequences, but equally are less likely to require major surgical intervention that could potentially predispose to neurological changes. This study had also adjusted for genetic conditions, maternal age and gestational diabetes, prematurity and other risks for NDD and CHD.

Jin et al. compared the sequencing results for genes in the top 25% of heart and brain expression, in individuals with CHD and a cohort with autism (n=4778) to identify genes that play a part in the pathogenesis of CHD and NDD[15]. There were 19 genes with damaging de novo LoF mutations common to both cohorts. Chromatin modifier genes made up 40% of these genes and were associated with a higher chance of NDD at the point of recruitment into the study. Chromatin modifying genes like *CHD4*, are known to play a role in the

development of autism spectrum disorders [564, 565], but most of those with autism and a mutation in a chromatin modifying gene do not have CHD.

Perhaps not surprisingly, 67% of the individuals with LoF mutations in genes common to both brain and heart development had NDD in addition to CHD. This was a higher incidence than the cohort as a whole, where only 33% had NDD. It is possible that the same genes are required in both the brain and heart, or regulate other genes that are required for correct organogenesis of both. Both of these organs have high oxygen requirements and could be more severely affected than other less dependent organs as a result of an abnormality in a more ubiquitously expressed gene.

Identifying genes that can cause both CHD and NDD, may provide additional prognostic information for the affected individual. However, it is likely that there are contributions to NDD and CHD from both genetics and environment and genotype is unlikely to predict the full phenotype. It would seem sensible to consider this in selecting individuals who will benefit from surgery in addition to genetic results, but also aim to minimise risks to the brain associated with treatments as much as possible in all.

3.3.9 Individual Gene Analysis

To identify novel CHD genes, we searched for those that were significant at a genome wide level. Across the whole CHD cohort (S-CHD and NS-CHD), 11 genes were significant. Separating the cohort into S-CHD and NS-CHD resulted in no genes reaching genome wide significance in the NS-CHD group, but 12 reaching this threshold in the S-CHD cohort (table 3.4). These included the same 11 genes found to be significant across the whole cohort and one additional gene. 9 of these 12 genes are known NDD genes. 4 had not previously been

associated with CHD. Genes such as *TAB2* had only been associated with NS-CHD previously, and our results suggest that it is also associated with S-CHD.

Gene	Known CHD-associated Gene	Known DD Gene
PTPN11	✓	✓
ANKRD11	✓	✓
CDK13		
ADNP	✓	✓
NSD1	✓	✓
PACS1	✓	✓
KMT2A	✓	✓
TAB2	✓	
DYRK1A		✓
DDX3X		✓
CHD4		
CHD7	✓	✓

Table 3.4 Significant S-CHD genes identified.

Genes identified as significant in the analysis of S-CHD. *CDK13*, *DYRK1A, DDX3X* and *CHD4* are highlighted as they have not previously been linked to CHD.

To look for further novel CHD-associated genes, all individuals whose WES results had identified a de novo mutation that could explain their CHD were removed from further analysis to leave those with "unresolved CHD". This led to the identification of three genome wide significant and novel S-CHD genes; *CDK13*, *CHD4* and *PRKD1*. There were no genes of genome wide significance in the unresolved NS-CHD group.

The TADA (Transmission and De Novo Association) analysis allowed us to identify genes that might be CHD genes based on both de novo and inherited mutations and both PTV and missense mutations, combined with case control data. Candidate genes were stratified by false discovery rate of strong (FDR <0.01/1%), intermediate (FDR >=0,01 and <0.05/>=1% and <5%) and weak

(>=0.05 and <0.1/ >=5% and <10%). Table 3.5 below shows the results for the highest confidence genes. Whilst these results suggests these genes may be involved in the aetiology of CHD, this analysis is not thought not to be sufficient to claim this definitively, due to the risk of false positives.

Gene	Known CHD gene	Known DD gene	Novel CHD gene	Enriched for DNM and inherited variants
ANKRD11	✓	✓		
NSD1	✓	✓		
ADNP	✓	✓		
KMT2A	✓	✓		
KMT2D	✓	✓		
NOTCH1	✓	✓		✓
CHD7	✓	✓		
DYRK1A		✓		
TAB2	✓			
KAT6A	✓	✓		✓
PTPN11	✓	✓		
PACS1	✓	✓		
MYH6	✓			
CHD4			✓	
CDK13			✓	
DIAPH3			✓	

Table 3.5 Likely CHD genes identified.

Genes identified with strong confidence to be CHD causing genes. *CHD4* and *CDK13* are included again.

These results correctly identify a number of genes that are already known to cause CHD. *CHD4* was identified as a novel CHD gene as it reached genome wide significance and is discussed later in this chapter. *CDK13* also reached genome

wide significance and is considered further in chapter 6. *DYRK1A* is found within the trisomy 21 critical region and is thought to contribute to NDD. In addition, heterozygous mutations in this gene cause *DYRK1A* intellectual disability syndrome. Aortic valve insufficiency and stenosis, pulmonary valve abnormalities and VSDs have been reported as part of this syndrome, although this is not a common feature.

DIAPH3 may be a novel CHD gene, but the confidence in this gene was not high enough to consider it definite. It is associated with autosomal dominant auditory neuropathy possibly due to gain of function mutations [566]. Homozygous individuals have a more severe phenotype. Transgenic mice overexpressing *Diaph3* have cardiac defects [566], but no heart abnormalities were seen in the family reported.

NOTCH1 and *KAT6A* were the only two genes enriched for both de novo and inherited variants, as shown above in table 3.5. This is perhaps less surprising in the case of *NOTCH1* as the phenotype can be variable and includes isolated heart defects. Frameshift, nonsense and truncating mutations have been identified in *KAT6A* in individuals with AD Mental Retardation [567-569]. ASD, VSD and PDA have been reported in these individuals. All mutations were de novo. The inherited mutations enriched in this cohort may act by another mechanism.

The phenotypes of the individuals with mutations in CDK13, PRKD1 and CHD4 were reviewed. As all patients were recruited to the DDD study, they had been phenotyped by a Clinical Geneticist already, but the uploaded phenotyping is sometimes brief. We collected additional information and reviewed the clinical phenotypes for similarities including facial features from photographs.

3.3.10 CDK13

7 individuals were identified with S-CHD and missense mutations in *CDK13*.

6 of the 7 mutations were known to be de novo, the seventh was of uncertain inheritance. All mutations clustered in the kinase domain, and four individuals had the same p.Asn842Ser mutation. These mutations are predicted to impair ATP or magnesium binding, or interactions with Cyclin K. There is a recognisable facial gestalt and other important similarities include significant developmental delay, mild to moderate microcephaly and two individuals had agenesis of the corpus callosum. Please see Appendix F (Figure 3A and 3C in Sifrim et al [12]) and Appendix L (Supplementary table 10 from Sifrim et al). *CDK13* is discussed further in chapter 6 along with the relevant experimental work I have carried out.

3.3.11 PRKD1

Three S-CHD individuals with *de novo* missense mutations in *PRKD1* were identified, two individuals had the same mutation. Two individuals have AVSD and the third had PS. Other shared features include severe developmental delay, ectodermal defects and limb abnormalities. Individuals with photos available and mutations in *PRKD1* had a broad forehead with frontal bossing, deep set eyes, recessed nasal root, infraorbital creases and a pointed and prominent chin. Please see appendix M (supplementary figure 3 from Sifrim et al. [12]). *PRKD1* is discussed further in chapters 4 and 5, with the relevant experimental work.

3.3.12 CHD4

Previous studies looking for genetic causes of CHD have identified genes involved in chromatin modification as important in the pathogenesis of CHD [14, 16]. Jin et al. have since also identified a prominence of chromatin modifying genes and loss of function mechanism as significant [15]. *CHD4* is a chromatin remodelling enzyme, and has been identified as a genome wide significant novel

gene causing S-CHD, a syndrome now called Sifrim Hitz Weiss Syndrome. A detailed review of the individuals reported in Sifrim et al. [12] and further publications follows, along with discussion of possible mechanisms by which mutations in *CHD4* result in CHD.

3.3.12.1 Mutations in *CHD4* cause CHD and a recognisable phenotype, which overlaps with other disorders of chromatin modification

We identified five individuals with S-CHD and de novo mutations in *CHD4* (4 missense variants and 1 in-frame deletion). Of the five individuals reported, two had TOF, one had an ASD, VSD and PV abnormality, one had a VSD and one had CoA. Considering these 5 individuals, it is possible that TOF spectrum abnormalities may be a characteristic of this syndrome.

All had significant early NDD, two had Chiari malformations and three of the four males had cryptorchidism or ambiguous genitalia. The individuals with mutations in *CHD4* share a wide, tall forehead with frontal bossing, distinct arched eyebrows, a depressed nasal root, broad nasal tip and low set anteverted ears (Figure 3.2). However, photographs are only available for three individuals so it is difficult to be sure of a distinct phenotype.
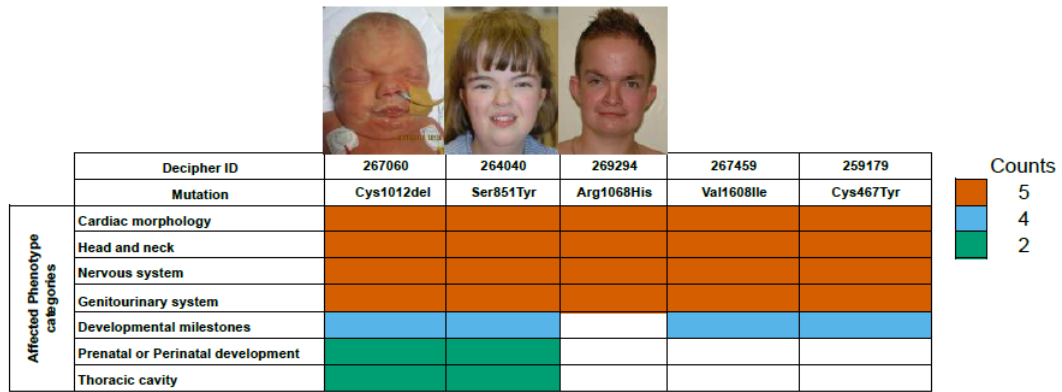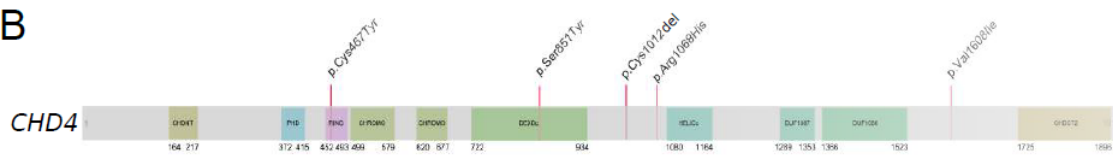
Figure 3.2 Probands with de novo CHD4 mutations.

A. Clinical synopsis of the observed phenotypes across patients carrying *CHD4* mutations. Columns represent single probands, shades of cells represent the number of probands sharing a phenotype in the given phenotypic categories. Photographs of affected probands are shown for which consent could be obtained for publication.

B. Protein plot showing *CHD4* protein domains and the distribution of *de novo* mutations.

This is supplementary Figure 2 from Sifrim et al.

Since the publication of Sifrim et al. changes in genome annotation have meant that two of the mutations have been reclassified on the Decipher Database [12, 522]. The mutation p.Cys1012* has changed from deletion causing truncation, to an insertion of three amino acids (p.Leu1009_Val1011dup) (IDs:267060). This individual has TOF, developmental delay, genital abnormalities and shares similar facial features including the arched eyebrows. The clinical phenotype is in keeping with the other reported individuals.

In the second individual, the mutation has changed from p.Val1698Ile to p.Val1636Ile (ID:267459). This missense variant is reported in gnomAD [570], detected in heterozygous state in Asian and non-Finnish European (allele count

0.000027 in non-Finnish European population and 0.000037 over all populations). The mutation also falls in a region that is not under high missense constraint pressure. The mean allele frequency attached to this variant may still be low enough to mean the variant is pathogenic. Unfortunately, photographs were not available for this individual so we cannot comment on whether they have a similar facial appearance. The clinical phenotype we have does share some similarities with the other reported individuals with *CHD4* mutations including developmental delay and hearing loss. They are the only individual with CoA. It is difficult to assess whether this mutation is definitely pathogenic without further information.

The individuals with *CHD4* mutations showed overlap with features of CHARGE Syndrome and Rubenstein Taybi Syndrome (RTS), caused by heterozygous LoF mutations in the paralog *CHD7* and *EP300*, a histone acetyltransferase respectively. The overlapping features are shown in table 3.6 below and include CHD, limb abnormalities, genitourinary abnormalities, and developmental delay.

Phenotype in *CHD4*	Seen in RTS	Seen in CHARGE
Developmental delay	+	+
Postnatal growth retardation	+	+
Coarctation	+	+
ASD	+	+
VSD	+	+
TOF	Rare	+
Impaired balance	-	+
Arnold Chiari	+	+[571]
Syringomyelia	+	+[571]
Hydrocephalus	+[572]	+[573, 574]
Ptosis	+	+
Hypermetropia	+[575]	-
Astigmatism	-	+[576]
Abnormality of the ear, including low set	+	+
Abnormality of the ear lobe	-	+
Bilateral hearing loss	Conductive	sensorineural
Omphalocoele	-	+
Cryptorchidism	+	+
Renal Abnormalities	+	+
Syndactyly	+	+
Nail hypoplasia	-	+
Low birthweight	-	-
Pulmonary artery abnormalities	-	-
Eyebrows	Thick eyebrows, down slanting palpebral fissures	Down slanting palpebral fissures
Anteriorly placed anus	-	-
Elbow	Dislocation of radial head	-
Broad, short clavicles	-	-
Flat acetabular roof	-	-
Glenoid hypoplasia	-	-
Lunate- triquetral fusion	-	-
Short femoral neck	Slipped upper femoral epiphyses	Bifid femur

Table 3.6 Overlapping phenotypes seen in CHARGE, RTS and CHD4 mutations.

Known core components of CHARGE and RTS are included, with some less common features. This demonstrates an overlap between the individuals with mutations in *CHD4*, with both RTS and CHARGE syndrome. RTS: Rubenstein Taybi Syndrome.

3.3.12.2 Additional Individuals Reported with mutations in *CHD4*

A further fourteen individuals have now been reported with mutations in *CHD4,* and some consistent features are now apparent (table 3.7) [522, 577]. Shared facial features include a wide forehead, telecanthus, dysmorphic ears and arched eyebrows. Additional common phenotypes include developmental delay and intellectual disability, hearing loss, skeletal, palatal and genitourinary abnormalities. Of those where information was available, normal birthweight, weight, height and OFC have been reported along with short stature, micro and macrocephaly. Only one individual had no information about NDD. All 19 other individuals had either developmental delay or ID or both. Where reported this ranged from mild to severe. Three individuals are reported to have specific language deficiencies.

ID	Gender, age	Mutation	Protein Consequence	Inheritance	Domain	CHD Phenotype
267459[1,3]	M 3y	c.4822G>A , changed to c.4906G>A in latest annotation	p.Val1608Ile in Sifrim et al., changed to p.Val1636Ile missense variant Exome MAF:3.65x10-5,	de novo	Neuromodulin and domain interacting with pericentrin	Coarctation
269294[1,3]	M 16y	c.3203G>A	missense variant p.Arg1068His,	De novo	C Helicase domain	ASD and VSD
267060[1,3]	M 1y	g.6701138delCATTCAGCA changed to c.3033-3034insCTGAATGTG	p.Cys1012* in Sifrim et al., changed to inframe insertion p.Leu1009_Val1011dup, 1012V/LNVV	De novo	SNF2_N domain	TOF
264040[1,3]	F 11y	c.2552C>A	p.Ser851Tyr, missense variant	de novo	SNF2 family N-terminal domain and DEXDc domain	ASD, VSD and PV abnormality
259179[1,3]	M 6y	c.1400G>A	p.Cys467Tyr missense variant	De novo	Zinc finger PHD domain	TOF
1[2]	M 10y	c.3380G>A	p.Arg1127Gln, missense variant	De novo	C helicase domain	No CHD
2[2]	F 16y	c.3518G>T	p.Arg1173Leu, missense variant	De novo	After C helicase domain	No CHD
3[2]	M 10y	c.3380G>A	p.Arg1127Gln, missense variant	De novo	C helicase domain	No CHD
4[2]	F 5y	c.3443G>T	p.Trp1148Leu, missense variant	De novo	C helicase domain	PDA requiring ligation, PFO, ASD, VSD

5[2]	M 18y	c.3008G>A	p.Gly1003Asp, missense variant	De novo	SNF2 family domain	ASD, PDA requiring repair, VSD, BAV, mild aortic dilatation
262731[3]	M 6y	, c.3380G>A	missense variant, p.Arg1127Gln,	De novo	C helicase domain	No CHD
265585[3]	F 11y	c.1933C>T	missense variant , p.Arg645Trp	De novo	Chromo domain	No CHD
274073[3]	M 3y	c.2143C>T,	stop gain p.Gln715Ter	De novo,	Between SNF2 family and chromo domain	No CHD
275697[3]	M 6y	c.2860A>G,	missense variant, p.Met954Val	De novo	SNF2 family domain	Secundum ASD and PS
284058	F 10y	c.4019G>A	missense variant, p.Arg1340His	De novo	DUF	No CHD
292555[3]	M 2y	c.3937T>G,	missense variant p.Tyr1313Asp	De novo	DUF	Perimembranous VSD
337310[3]	M	C.3547C>T,	missense variant, p.Arg1183Cys	De novo	After C Helicase domain	No CHD
381781[3]	F Age Unknown	c.3395G>A	Missense variant p.Gly1132Glu (1132G/E)	De novo	C Helicase domain	No CHD
258997[3]	F 3y	c.329A>G	Missense variant p.Tyr110Cys (110Y?C)	Unknown	na	No CHD

Table 3.7 Clinical features of all the individuals reported with mutations in CHD4.

DUF: Domain of unknown function, HH: hypothalamic hypogonadism, IUGR: intrauterine growth retardation, PHD: PHD finger, SNF2: SNF2 family domain, USPFs: up slanting palpebral fissures, UTI: urinary tract infection, VPI/SMC: velopharyngeal insufficiency/submucous cleft, VUR: vesicoureteric reflux, -: no data. References: 1: Sifrim et al. [12], 2: Weiss et al. [577], 3: Decipher Database [522].

ID	Growth parameter	NDD	CNS	Hearing loss	Genital Abnormalities	Palate	Skeletal	Other
267459[1,3]	2.1kg at 38/40 (SD -1.38), H: 50th, OFC SD 2.32	Delayed development	-	Yes	-	-	-	Unilateral ptosis, sacral pit
269294[1,3]	3.4kg at 42/40 OFC: SD 0.62	Gait imbalance, language impairment	Arnold Chiari	-	Cryptorchidism	-	Limited elbow extension, Broad clavicles, Flat acetabular roof, Glenoid hypoplasia, Lunate-triquetral fusion, Short clavicles, Short femoral neck	Gait imbalance, hypermetropia cutaneous syndactyly hands
267060[1,3]	2.635 kg (SD -1.95) OFC 33.3 cm (SD -0.05) at 37/40	Delayed development	-	-	Ambiguous genitalia, Anteriorly placed anus,	-	Wormian bones	Large fontanelle in utero, omphalocoele Poor corticomedullary differentiation on USS of kidneys, small nails
264040[1,3]	2.4Kg at 41/40 (SD -2.32)	Global delayed development especially motor	Arnold-Chiari malformation, Hydrocephal	-	Anteriorly placed anus	-	-	Astigmatism, postaxial polydactyly, recurrent UTIs

	H:-1SD Wt: 0.72 SD OFC: 3.52 SD		us, Syringomyelia					
259179[1,3]	2.7Kg, OFC 33.5 cm (SD -0.87) at 39/40 (SD -1.37) Ht: 117cm (SD -0.09) Wt: 22.25 kg (SD 0.33) OFC: 51 cm (SD -1.47)	Delayed development delayed speech and language, moderate intellectual disability		-	Cryptorchidism	-	Trigonocephaly	VUR, feeding difficulties
1[2]	H: 75th OFC:>98th	Delayed development intellectual disability	Enlarged lateral ventricles, stroke, moyamoya	Yes	Undescended testes, micropenis HH	Bifid uvula	Advanced bone age (2-3y)	Telecanthus, dysmorphic ears
2[2]	H: 40th OFC:>98th	Delayed development intellectual disability	Chiari malformation, enlarged lateral ventricles	Yes	-	-	Tarsal coalition, cervical vertebral fusion	Telecanthus, dysmorphic ears

3[2]	H: 50th OFC:>98th	Delayed development mild intellectual disability	enlarged lateral ventricles	Yes	Undecended testes, micropenis HH	Hypernasal speech, VPI/SMC		Telecanthus, dysmorphic ears, falx calcification
4[2]	H: <3rd OFC:20th	Severely delayed development intellectual disability	enlarged lateral ventricles, basilar invagination and narrow foramen magnum	-	-	Hypernasal speech, VPI/SMC	Scoliosis, C2/3 fusion, bilateral coxa valga, carpal fusion, brachymesophalangia	Telecanthus, dysmorphic ears
5[2]	H: 10th OFC:90th	Delayed development mild intellectual disability	enlarged lateral and third ventricles	Yes	Micropenis HH	Hypernasal speech, VPI/SMC	Diffuse osteopaenia	Telecanthus, dysmorphic ears
262731[3]	Short stature, macrocephaly	Global delayed development	-	-	Abnormality of the male genitalia	-	-	Abnormal face shape, plagiocephaly
265585[3]	-	Intellectual disability, moderate	-	-	-	-	Thoracic kyphosis	Hyperteloric, USPFs, wide nasal bridge. Pica, stereotypy
274073[3]	Macrocephaly	Mild global delayed development delayed speech and language	-	-	-	Cleft palate	-	Micrognathia, upper airway obstruction, synophrys
275697[3]	microcephaly	Mild global delayed	-	-	-	Cleft palate	-	Carious teeth, hyperactivity

		development moderate expressive language delay						
284058[3]	microcephaly	Mild intellectual disability	-	-	-	High palate	-	Anteverted nares, low set ears, overfolded helix, prominent crus of antihelix, telecanthus, U shaped upper lip vermillion branchial cyst, dental crowding, epicanthus,
292555[3]	IUGR, microcephaly	Globally delayed development	-	-	-	-	Bilateral postaxial polydactyly	Epicanthus, single transverse palmar crease
337310[3]	-	Globally delayed development	-	Yes	-	-	-	-
381781[3]	-	-	-	-	-	-	-	-
386103[3]	-	Intellectual disability	-	-	-	-	-	Behavioural abnormality
258997[3]	-	Globally delayed development	-	-	-	-	-	Abnormal facial shape

3.3.12.3 CHD is seen in almost half of individuals with mutations in *CHD4*

The penetrance of CHD is 40% (9/19). The specific types of CHD are outlined below in table 3.8. Septal defects, TOF and TOF like defects predominate.

CHD	Number
ASD and VSD	2
TOF	2
ASD, VSD, PV abnormality	1
ASD and pulmonary stenosis	1
PDA	2
ASD	1
VSD	1
Coarctation	1
BAV	1

Table 3.8 CHD phenotypes reported with mutations in CHD4.

The two PDAs both required ligation and were found in combination with an ASD and VSD in both cases.

3.3.12.4 CHD4 is a chromatin remodelling enzyme

There are four families of chromatin remodelling enzymes; the SWI/SNF, ISWI, INO80 and CHD (Chromodomain-Helicase-DNA binding) families. These ATP dependant chromatin remodelling enzymes control DNA transcription, replication and repair by altering the level of compaction of chromatin. They are important in transcription regulation and many developmental processes. *CHD4* is a member of the CHD gene family. The CHD family comprises three subfamilies. Subfamily I includes Chd1 and 2. CHD3 CHD4 and CHD5 form subfamily II, and Chd6-Chd9 form the third subfamily. There are two important motifs common to the CHD family, tandem chromo domains in the N-terminal region, and a SNF2-like ATPase domain in the centre of the protein.

There is already evidence that these genes play a role in congenital heart disease. The SWI/SNF complex has been linked with congenital heart disease through its interactions with *GATA4*, *TBX5* and *NKX2.5* [578], and loss of function mutations in *CHD7* cause CHARGE syndrome.

3.3.12.5 Structure of CHD4 and other CHD proteins

CHD4 comprises a number of important domains including the central ATPase helicase, PHD and tandem chromo domains shown in figure 3.3 below. CHD3 and CHD4 do not have a DNA binding domain [579] and differ from Chd1 and 2 as they have paired PHD zinc finger like domains in the N-terminal section. These PHD domains bind with histones and may play a role in determining what the CHD4-NuRD complex binds with [580]. They also both have tandem chromodomains and an ATPase motor [581]. In *Chd1*, the chromodomains regulate the ATPase motor activity in the SNF2 domain [582].

CHD4 also has an HMG-box (high mobility group) like domain in the N-terminal domain. HMG boxes are chromatin binding domains, specific for certain nucleic acid structures [583].This domain is important in localising CHD4 to sites of DNA damage, and can preferentially bind PAR (poly ADP-ribose) over DNA in vitro [580]. This results in reduced remodelling function of CHD4. PARylation has an important role in DNA repair and nucleosome remodelling.

There are an additional two domains of currently unknown function in the C-terminal domain. The Chd6-9 subfamily has additional C-terminal motifs, including paired BRK domains, SANT-like, CR and DNA binding domains which set them apart from the other CHD genes [584-586].
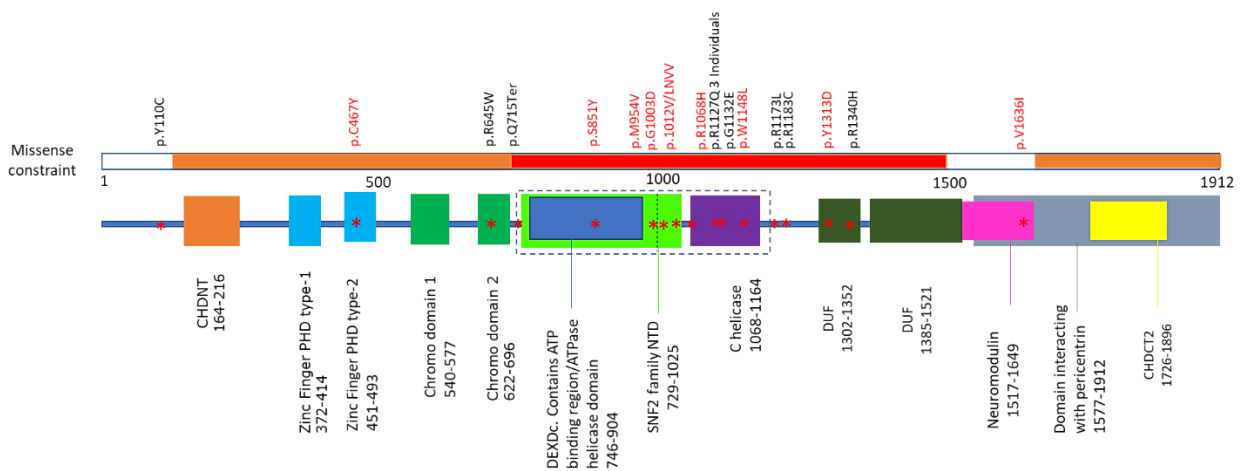
Figure 3.3 Pictorial representation of CHD4 and reported mutations.

Variants in red text are associated with CHD, those in black do not. Areas of missense constraint are a.a. 148-709, 710-1504, 1684-1912 [494]

Domains:

CHDNT (orange): Conserved domain found in PHD/RING finger and chromo domain-associated helicases.

PHD1 (turquoise): PHD finger 1 found in class II Chromodomain-Helicase-DNA binding (CHD) proteins

PHD2 (turquoise): PHD finger 2 found in class II Chromodomain-Helicase-DNA binding (CHD) proteins

Chromo Domain 1 (green): A chromatin organization modifier (chromo) domain is a conserved region of approximately 50 amino acids. This domain is found in a number of chromosomal proteins, which are important for functional organization of the nucleus. This domain may be involved in the binding of these proteins to methylated histone tails and possibly RNA. A tandem arrangement is not uncommon.

Chromo Domain 2 (green): Chromatin organization modifier domain

SNF2 family NTD (lime green): SNF2 family N-Terminal Domain. This domain is found in proteins involved in transcription regulation (including SNF2), DNA repair and recombination and chromatin unwinding.

DEXDc (blue): This domain contains the ATP-binding region. It is seen in the DEAD-like helicases superfamily, which are proteins involved in ATP-dependent RNA or DNA unwinding. Also known as the DEAD/DEAH box helicase

C Helicase: Helicase conserved C-terminal domain. This domain family is found in a wide variety of helicases and helicase related proteins. It is likely to be an important part of the helicase.

DUF (dark green): Domain of unknown function

Neuromodulin (pink): Neuromodulin (Axonal membrane protein GAP-43) sequences. Often found in conjunction with a Calmodulin-binding motif (pfam00612). GAP-43 is a neuronal calmodulin-binding phosphoprotein that is important in axon growth.

CHDCT2 (yellow): This C-terminal domain is found in PHD/RING finger and chromo domain-associated CHD-like helicases.

Black box, dashed outline: ATPase motor Farnung et al.

Domain interacting with pericentrin (grey): Required for anchoring to centrosomes along with CHD3 [587].

Build 38, domain sites from https://www.ncbi.nlm.nih.gov/gene/1108 last updated 12.10.19. Mutations reported by Sifrim et al, Decipher and Weiss et al [12, 489, 588].

3.3.12.6 Function of CHD4

The CHD3 and CHD4 subfamily, are mainly associated with transcriptional repression [589]. CHD3, CHD4 and CHD5 all attach to the NuRD complex [579, 590]. The NuRD complex has an important role in chromatin remodelling and gene transcription [591]. Specifically, it can perform ATP-dependant nucleosome remodelling and deacetylate histones, having these two enzymatic functions sets it apart from other similar complexes [580, 592]. NuRD has a role in gene repression and activation, progression of the cell cycle, response to DNA damage, embryonic development and stem cell renewal and differentiation and

progenitor cell fate [593-597]. Given the importance of NuRD in many cellular pathways and processes, it is feasible that alterations in *CHD4* may have deleterious consequences.

The NuRD complex consists of a number of subunits outlined in table 3.9 below; Gata2a/2b, MBD2/3, MTA1/2/3, Rbbp7, Rbbp4, HDAC1, HDAC2 and CHD3/4 and 5 and possibly an additional subunit CDK2AP1 (DOC-1) [598]. Another gene involved in this complex, *GATAD2B,* is associated with intellectual disability but not CHD [599].

Subunit	Associated functions
HDAC1, HDAC2	Histone deactelylases, remove acetyl groups from lysine residues on histone tails and proteins, including p53 [600]. Not specific to the NuRD complex. Important for regulation and scaffolding.
CHD3 (Mi2-α), CHD4 (Mi2-β), CHD5	Chromodomain-helicase-DNA-binding proteins – helicase and ATPase activity. Repositions nucleosomes on DNA using ATP [601].
MBD2 or MBD3	Methyl-CpG- binding domain proteins, no enzymatic function. They bind cytosine-methylated DNA [602].
MTA1, MTA2 or MTA3	Metastasis associated proteins allow binding to DNA, HDAC1, transcription factors and other substrates of the NURD complex [603, 604].
Rbbp4, Rbbp7	Histone binding proteins. Bind histone H4 and allow formation of larger complexes [605]. Not specific to the NURD complex.
Gata2a, Gata2b	Nuclear zinc finger proteins interact with MBD2/3 [594].

Table 3.9 Components of the NuRD complex.

Different combinations of NuRD subunits mean that the complex can work with multiple transcription and regulatory factors in different tissues to produce different effects. Methylation and demethylation of the MTA1 subunit allows inactivation and activation of genes with CHD4. Methylated MTA1 forms the H3K9me2 histone mark with histone H3, which then recruits CHD4 and creates an inactive chromatin state [597]. When MTA1 is demethylated, it associates

with a different set of factors and CHD4 and the NuRD complex are removed, resulting in active gene transcription [606].

CHD4 plays an important role in cell differentiation and acts by repressing certain cell lineages [607]. NuRD also interacts with SMAD2/3 and the hippo pathway to reduce expression of pluripotency genes and silence mesoendoderm differentiation genes [608]. Specifically *CHD4* has a role in the haematopoietic system, where it drives cells down the lymphoid lineage [609] and conditional deletion in the bone marrow leads to lack of lymphoid and myeloid cells [610]. More recently tbx5 has been shown to interact with the NuRD complex [611]. It represses specific gene programs to allow normal development of the heart. This supports the role of *CHD4* as a novel CHD gene.

Despite the similarities between *CHD3* and *CHD4*, there does not appear to be redundancy between the two genes. Although human CHD3 and CHD4 are co-expressed in many tissues, they form separate complexes with NuRD, and act on some unique and some shared gene targets [612]. Both CHD3 and CHD4-NuRD complexes are found at repair sites of UV-induced DNA damage, but have different nuclear localization patterns in undamaged cells. In vitro there is evidence of both gene repression and activating functions in shared and specific genes of these two CHD proteins.

3.3.12.7 Mouse models suggest *CHD4* is required for normal development of both cardiac and skeletal muscle

Mouse models can help provide supporting evidence for the effects of a gene in humans. The IMPC results for the C57BL/6Ntm1b mouse indicates complete embryonic lethality in the homozygotes, prior to E9.5 [214]. They also demonstrate hypoactivity and decreased lean body mass. Expression is recorded in the embryonic atrium, ventricle, outflow tract and pericardium, Low levels of expression were seen in the adult vascular system, heart and aorta (>50-day old males).

With reference to the heart specifically, the NuRD complex controls cardiomyocyte differentiation and therefore may play an important role in heart development and CHD. It does this by silencing the programme for skeletal muscle lineage in both developing and differentiated cardiomyocytes [613, 614]. Complementary to this, Chd4/NuRD silences cardiac genes in skeletal muscle. Wilczewski et al. found that CHD4/NuRD complex represses non-cardiac myofibril paralogs to allow normal cardiac sarcomere formation [614]. If *CHD4* is lost, an abnormal muscle cell forms in the heart that has components of cardiac, skeletal and smooth muscle cells. The resulting abnormal sarcomeres mean that cardiac function is impaired and the embryo is lost as a result. This team also identified CHD4 binding sites in smooth muscle myosin heavy chain, fast skeletal α-actin, and the fast skeletal troponin complex genes.

Loss of *Chd4* in skeletal muscle causes myopathy in mice [613]. A mouse model with deletion of *Chd4* causes cardiovascular malformations (VSDs), cardiomyopathy and fibrosis, and arrhythmias as well as abnormal skeletal muscle [613]. There is also abnormal mitochondrial function in both cardiac and skeletal striated muscle. The mice showed thinner ventricular compact myocardium due to reduced cardiomyocyte proliferation, more simple trabeculae and abnormal ventricular septum formation at E10.5. Death was seen in almost all by E14.5.

A further mouse model has been produced without endothelial *Chd4*. This results in death at mid-gestation due to vascular rupture [615]. A conditional mouse model was also produced with deletion of *Chd4* in terminally differentiated cardiomyocytes [613]. The mice died after 3 months of life. Histology showed deterioration from grossly normal hearts at 4 weeks old, to atrial dilatation and fibrosis in the interventricular septum and the left ventricle, lower ejection fraction, late arrhythmias, but no hypertrophy. This shows that not only is *Chd4* required in embryonic stem cells, but it also had a role in the postnatal heart.

These mouse models support the role of *CHD4* as an important gene in development of the mouse heart. Abnormal muscle could certainly result in heart failure and death, but how exactly structural defects occur is not certain. VSDs were seen in the mouse and fits nicely with the VSDs and TOF/TOF like phenotypes we have seen in affected individuals.

3.3.12.8 Disease Mechanisms with *CHD4* variants

The majority of the CHD4 mutations reported are missense. Most, but not all fall within characterised domains of *CHD4* (figure 3.2 shown previously). Farnung et al. mapped a number of these mutations onto the protein structure of CHD4 bound to a nucleosome and proposed a mechanism of action in most cases [581]. Information for all the reported mutations are shown below in Table 3.10.

Variant	Proposed mechanism
p.C467Y	The de novo variant p.C467Y falls within the zinc finger PHD type 2 domain. These small protein motifs are important in genes which modify chromatin and control gene transcription [616, 617]. They bind a variety of metals including zinc and can form salt bridges and are involved in a wide variety of processes [618]. Farnung et al. suggest that this mutation results in abnormal binding of zinc ions[581]. This individual has TOF.
p.R645W	This falls within the second of the tandem chromo domains (CHRomatin Organization Modifier). These domains are around 50 amino acids long, and are found in other CHD proteins including CHD3. The chromodomain helps mediate chromatin interactions by binding directly to DNA, RNA, and methylated histone H3, based on functional studies [619]. In the mouse, mutations in the chromodomain of Chd1 cause nuclear redistribution and impaired association with chromatin [620] . Additionally, in Drosophila abnormal nucleosome binding, mobilization, and ATPase functions resulted when the chromo domains of CHD2 and CHD4 were deleted [619, 621]. This suggests a deleterious effect of this variant.

p.Q715Ter	A single individual has the de novo variant p.Q715Ter. They do not have any other CNVs or other potentially pathogenic mutations. They do not have CHD. This would terminate the protein before the ATPase helicase domain. No functional evidence is available, but I assume it would result in loss of function of CHD4.
p.S851Y	This mutation falls in the ATPase lobe 1, Farnung et al. mapped this variant onto their crystal structure, but did not propose a specific method of action[581]. This individual has an ASD, VSD and PV abnormality.
p.M954V	Falls within the SNF2 family NTD and the ATPase lobe 1. This residue is identified as an area of low conservation across the chromatin remodellers, and a mechanism has not been proposed[581]. This individual has an ASD and PS.
G.1003D	G.1003D could result in loss of chromatin remodelling ability of CHD4. It falls within the ATPase lobe 2, and Farnung et al. showed that this variant is close to H3 alpha helix 1[581]. Deletion of this loop in Chd1, resulted in loss of chromatin remodelling ability [622]. This individual has ASD, PDA, VSD, BAV and aortic dilatation.
P1012V/LNVV	P1012V/LNVV falls within the ATPase lobe 2. The effect of this variant is unknown, but it is in close proximity to residues required for H4 and H3 binding, which are important for chromatin remodelling ability (Leu1009) [622]. This individual has TOF.
p.R1068H	p.R1068H is predicted to alter the RecAATPase fold as this arginine residue is important in forming a number of hydrogen bonds to maintain its structure[581]. It falls within the ATPase lobe 2. This individual has an ASD and VSD.
R1127Q	R1127Q falls within the ATPase lobe 2, and disrupts the interactions with the DNA minor groove[581]. Farnung et al. also pointed out that a variant in the equivalent arginine of SMARCA4 results in Coffin-Siris Syndrome (which includes CHD), supporting its pathogenicity [623]. Weiss et al showed that the R1127Q mutant protein localised and interacted with HDAC1 as normal, and predicted that this mutation was likely to disrupt ATPase activity[577]. Three individuals carry this mutation, none have CHD.
p.G1132E	p.G1132E falls within the ATPase lobe 2, within the highly conserved canonical ATPase motif V. The consequence of this variant is unknown, but R1127Q variant is in the same motif. This individual does not have CHD.

p.W1148L	p.W1148L prevents ATP hydrolysis and chromatin remodelling as this residue is part of the ATPase motif Va [624, 625]. This individual has an ASD and VSD, PDA and PFO.
p.R1173L	p.R1173L inserts into an acidic pocket and is predicted to result in abnormal folding of the ATPase lobe 2[581]. Weiss et al showed that the p.R1173L mutant protein localised and interacted with HDAC1 as normal, and predicted that this mutation was likely to disrupt ATPase activity [626]. This individual does not have CHD.
p.R1183C	p.R1183C variant does not fall within a specific domain and is in an area that shows only low conservation across the CHD proteins. However Farnung et al. show this arginine residue as forming a bond with the DNA, which may mean that this residue is important for CHD4-DNA binding[581]. This individual does not have CHD.
p.Y1313D	p.Y1313D falls within exon 26, and a domain of unknown function, the result of which is not clear. This residue is conserved in mouse and rat (UCSC Genome browser https://genome.ucsc.edu), and Decipher confirms Tyrosine in Chimpanzee, chicken, Xenopus, Zebrafish, Drosophila and C. elegans. This variant falls within an area of high missense constraint. This individual has a VSD.
p.R1340H	p.R1340H falls within a domain of unknown function, the result of which is not clear. Decipher confirms Arginine at this position in human, mouse, Chimpanzee, chicken, Xenopus, Zebrafish, Drosophila and C. elegans. This variant falls within an area of high missense constraint. This individual does not have CHD.
p.V1636I	V1636I falls within an area known to interact with pericentrin and the neuromodulin domain. It was initially reported as p.V1636I in Sifrim et al. but has since been revised with more recent annotation [12]. Valine at this residue is present in Chimpanzee and Chicken, but not Mouse or Zebrafish. This individual has CoA.
p.Y110C	The p.Y110C missense variant is of uncertain pathogenicity. It falls within a non-missense constrained section of CHD4, with no specific function and it is not known if it is de novo.

| | This individual also has another variant in SLC6A1 which is associated with epilepsy. The phenotype is abnormal face shape and developmental delay, which is not specific to *CHD4*. This individual does not have CHD. |

Table 3.10 Proposed mechanism of *CHD4* mutations.

All reported CHD4 mutations and proposed mechanism resulting in abnormalities. Mutations reported by Sifrim et al. [12], Decipher [489] and Weiss et al [577].

Proposed mechanisms leading to clinical abnormalities include loss of chromatin remodelling ability, disruption of ATPase activity and prevention of ATP hydrolysis, altered CHD4-DNA binding and abnormal zinc ion binding leading to abnormal function of CHD4 [581, 626]. Farnung et al. suggested that the mutations they reported had a plausible mechanism that would result in abnormal function [581].

The single truncating mutation, which is sited before the helicase and SNF2 like domain is associated with a milder phenotype. The individual has mild NDD, dysmorphic features and palatal abnormalities only. This suggests that there may be a milder phenotype with loss of function mutations, and the more severe phenotype seen in association with some missense mutations may represent another mechanism, particularly those affecting important domains of the protein. *CHD4* does have a probability of loss of intolerance (pLI) of 1, indicating likely haploinsufficiency [627]. *CHD7* also has a pLI of 1, and CHARGE syndrome results from loss of function mutations. Variants are reported along the length of the CHD7 protein that are stop gain mutations and predicted to result in loss of function. However additional missense mutations reported in the chromo domains, SNF2 family like N terminal domain and the C-helicase are reported as pathogenic [521].

To consider haploinsufficiency further, I reviewed a number of databases for CNVs overlapping CHD4. There are no benign CNVs which include *CHD4,* reported in the normal population in the International Standards for Cytogenomic Arrays (ISCA) database [628]. Decipher reports 27 deletions considered pathogenic that overlap *CHD4* [522]. They are all large deletions encompassing multiple genes, and none overlap just *CHD4*. All encompass the 12p13.33 deletion syndrome locus. There are four smaller deletions marked as being of unspecified pathogenicity. All are 3.18Mb in size, contain 110 genes including *CHD4*, and are of unknown inheritance. Phenotype data is available for two individuals. Interestingly one individual has macrotia and protruding ears, short stature, intellectual disability and skeletal abnormalities including an abnormality of the hip bone and scoliosis. The second individual shares a sacral dimple, strabismus, pectus excavatum, a preauricular pit and intellectual disability. All of these features have been seen in individuals with mutations in *CHD4*. Additional phenotypes include strabismus, brachycephaly, cutis gyrata, seizures, hyperextensible skin, constipation, and nasal and columella differences. Given the size of the deletions, it is difficult to attribute the phenotype to *CHD4*, especially given that it includes a number of other morbid and haploinsufficient genes such as *ATN1* (heterozygous mutations cause congenital hypotonia, epilepsy and NDD and *C1S* (heterozygous mutations cause periodontal ehlers danlos).

3.3.12.9 Structural abnormalities are common with mutations in the SNF2 family and C helicase domains

All individuals reported so far had either developmental delay or ID. The majority of individuals have structural abnormalities, especially those with variants in the SNF2 family and C-helicase domain (9/10). No structural abnormalities were seen in association with mutations in the chromo domain. Brain abnormalities were found in all but one individual in these domains, and also in association with the p.R1173L mutation just distal to these domains. Ventriculomegaly and Arnold Chiari malformations predominate.

Genitourinary malformations were associated with variants in the SNF2 family NTD and C-helicase domain (7/10). Half of the individuals with variants here had palatal abnormalities (5/10). Palatal abnormalities were only associated with this domain and a single truncating mutation (p.Q715ter) sited between the chromodomain and SNF2 family NTD.

There does not appear to be any specific genotype phenotype correlation for growth, however all three individuals with p.R1127Q exhibit macrocephaly. Skeletal features were fairly diverse and there did not appear to be any specific genotype phenotype correlation. Hearing loss also did not appear to associate with any particular location.

Three individuals have the p.R1127Q variant. All had developmental delay and genitourinary abnormalities. 2 of the 3 had ventriculomegaly, hearing loss and palatal abnormalities. Interestingly these two individuals were reported by Weiss et al, the third individual who does not show all the features is from DDD [522, 577]. The individual with p.R645W (chromo domain) has a milder phenotype than others. As mentioned previously, the individual with the truncating p.Q715ter mutation also has a milder phenotype.

There does not appear to be a clear genotype-phenotype correlation with respect to CHD. Individuals with CHD have mutations along the length of the protein including the PHD type 2 domain, the ATPase lobes, the first domain of unknown function and the neuromodulin domain. There also does not appear to be a specific region where mutations occur that result is a specific type of CHD. Why some mutations result in CHD and others do not remains unclear. There is some consistency in that all three individuals with p.R1127Q do not have CHD.

Overall, there does not appear to be any clear genotype-phenotype correlation, with the exception of higher likelihood of structural abnormalities with mutations in the SNF2 family and C helicase domains.

3.3.12.10 Sifrim Hitz Weiss Syndrome shares some clinical features with syndromes caused by variants in the other CHD family genes.

Given the similar functions and structures of the genes within the CHD family, it is also helpful to review the clinical phenotypes associated with the other CHD genes, especially *CHD3* given it is in the same subfamily as *CHD4*. A number of syndromes already associated with the CHD gene family are outlined below in table 3.11. Common features include NDD, autism, LD, epilepsy, facial dysmorphism, genitourinary abnormalities, hypotonia, micro and macrocephaly. NDD is not a surprising feature given that *CHD3*, *CHD4* and *CHD5* have non-redundant functions in the brain during development [629].

Gene	Disorder	Features	CHD	Shared features with *CHD4*
CHD1	Neurodevelopmental syndrome [630]	Probable dominant negative mutations cause NDD. Speech apraxia, autism, hypotonia, seizures, immune disorders, facial dysmorphism	No	NDD. Some similarities between facial features, but too few individuals reported to be conclusive.
CHD2	Epileptic encephalopathy [631]	NDD, childhood onset epilepsy	No	NDD
CHD3	Snijders Blok-Campeau Syndrome[632]	NDD Impaired speech and language Macrocephaly, autism, stereotypies, macrocephaly, hypotonia, epilepsy (1)	No	NDD Genital abnormalities, hypotonia, low set simple ears with thick helices, pointed chin
CHD4	Sifrim-Hitz-Weiss Syndrome		Yes	-
CHD5	No associated syndrome. *CHD5* expression seems to be mainly			

	confined to the central nervous system and testes [633, 634].			
CHD6	No syndrome. Knockout mouse shows impaired coordination[635]. This genomic region overlaps with that associated with human ataxias.			
CHD7	CHARGE Syndrome [324]		Yes	NDD, CHD, ear malformations and hearing abnormalities, clefts, genital and anal abnormalities
CHD8	Autism [564, 636, 637]	Autism, distinct facial features, macrocephaly, GI abnormalities, seizures, hypertelorism, tall, ID, regression	No	NDD

Table 3.11 Clinical phenotypes associated with variants in the CHD gene family.

No clinical syndrome has been described for CHD5 as yet, and it is in a different subfamily to CHD3 and CHD4.

Several other genes that interact with the NuRD complex, are linked with syndromes with overlapping features. The interaction of the NURD complex and *SALL4* has also been studied in the context of oncogenesis. Like individuals with *CHD4* mutations, Okihiro Syndrome (SALL4) and Townes-Brocks Syndrome both include ear abnormalities, deafness, CHD, anal and genitourinary abnormalities

and skeletal features. Interestingly SALL1 (Townes-Brocks syndrome) [638] is thought to interact with the NuRD histone deacetylase complex [638]. This is of relevance given that there is some overlap between these two syndromes and the presentation of the individuals reported with mutations in *CHD4*.

CHD4 also forms part of the ChAHP complex, with *ADNP* [639]. Mutations in ADNP are associated with Helsmoortel-Van der Aa syndrome, which includes CHD as part of the phenotype. The ChHAP complex is required for correct cell fate progression and therefore abnormal function was in keeping with the wide range of systems affected in this syndrome. There are a number of similar features between the individuals reported with mutations in *CHD4* and *ADNP*, including developmental delay, intellectual disability, ear abnormalities and less commonly CHD in Helsmoortel-Van der Aa syndrome.

There are other possible routes which may lead to clinical abnormalities when *CHD4* is not functioning as normal. There is evidence that *CHD4* can act independently of the NuRD complex, specifically in the transition to adult haemoglobin via γ-globin gene silencing [640], repression of proneural genes in drosophila [641] and CD4 expression and T cell development [642]. Overall there is evidence of shared features of NDD amongst the CHD genes and CHD in some of these and other linked genes.

3.3.12.11 Clinical Management

Suggested clinical management of children with deleterious *CHD4* mutations might include monitoring of growth parameters, as at least one individual required growth hormone. Development should be monitored to allow early provision of support if required, along with brain imaging if there is evidence of NDD. Particular attention should be paid to speech development and hearing should be tested in all cases. Palatal assessment and genital/anal examination should be carried out in all newborns as part of the normal baby checks. Further examination should be carried out if concerns arise. Additional investigations

that might be considered include renal ultrasound and skeletal imaging if symptomatic.

Given CHD is present in almost half of individuals with mutations in *CHD4*, an echo seems a sensible investigation to include in all individuals, especially given the non-invasive nature. A single individual has aortic dilatation in association with BAV. It is unclear if this is part of Sifrim Hitz Weiss syndrome. It may be that the dilatation is purely related to the BAV rather than any other process in the aorta related to *CHD4*.

There are no specific details about whether the affected individuals in our study had myopathy, but given the role of *CHD4* in skeletal as well as cardiac muscle, it would be worth observing if any of the affected individuals have any features of skeletal myopathy.

3.3.12.12 Cancer

We know that the NuRD complex plays a role in response to DNA damage, including double strand break repair [643], and failure to correct this can result in cancer developing. Mutations are found commonly in certain cancer types [644]. It interacts with both tumour suppressors and oncogenes and so can have a variable effect on tumour growth [600, 645].

CHD4 expression increases in response to UV irradiation [646]. *CHD4* binds PARPs (poly ADP-ribose polymerase) and NuRD, and mobilises to sites of double strand breaks and DNA damage [580, 595]. PAR binding motifs are found in the N-terminal domain, upstream of the PHD binding domains [580].

Somatic mutations have been identified in *CHD4* in nearly 20% of uterine cancers, in the helicase domain and second plant homeodomain which binds methylated H3K9 [647-649], along with somatic CNVs involving *CHD4* and *MBD3* [650]. The exact role of *CHD4* is not clear [643, 651]. Interestingly *CHD3* and *CHD4* are associated with dermatomyositis, and 20-30% of individuals with this condition go onto develop cancer [652-654]. *CHD3* is also linked with Hodgkin's Lymphoma, and *CHD5* with neuroblastoma [633, 655, 656].

A number of variants in *CHD4* have been reported in a variety of cancers, spanning the PHD finger domain 2, the tandem chromodomain and the two lobes of the ATPase motor. Farnung et al. proposed two routes by which these variants can have an effect; changes to ATPase and DNA translocation activity, and alterations to protein stability[581]. These mutations are often within a few residues of those reported in individuals with Sifrim Hitz Weiss Syndrome.

The significance of already having a *CHD4* germline mutation upon the risk of developing cancers is unknown currently. It would seem prudent to promptly review any concerning signs and symptoms and report any occurrences of malignancy to determine if any screening is indicated.

3.3.13 Copy Number Variation Analysis

A single patient with Turner syndrome and five classical 22q11.2 deletions were identified. A single patient with the reciprocal 22q11.2 duplication was also identified, the significance of this is currently unclear. I ascertained the gene content of each CNV for inclusion of potentially significant genes identified in this study. No CNVs overlapped any of the newly identified genes *CHD4*, *CDK13*, *PRKD1*, *DYRK1A*, *DDX3X*, or *DIAPH3*.

3.3.14 Finding further novel CHD genes will require much larger cohorts

Since the publication of Sifrim et al. a number of other studies have identified further novel CHD genes including *GDF1* and *FLT4* [15, 19] and put forward additional candidates including *ACVR1, JARID2, PLRG1, RYR1, ZFPM1, CAMTA2, DLX6*, and *PCM1* [19, 657]. This is in keeping with the findings of Jin et al. which supported the role of *FLT4* in TOF [15].

Whilst our study, and subsequent ones, have increased our knowledge of genes that contribute to CHD, it is apparent from both the de novo and inherited PTV analysis, that there are more novel genes to be discovered. One method would be to explore pathways and protein interactions that we identified as significant (section 3.3.10 and figure 3.3 shown previously). An updated protein interaction map from STRING is shown below in figure 3.4 using the same 374 genes identified as potentially significant in the integrated de novo and inherited variant analysis, with a FDR of <50% [12].
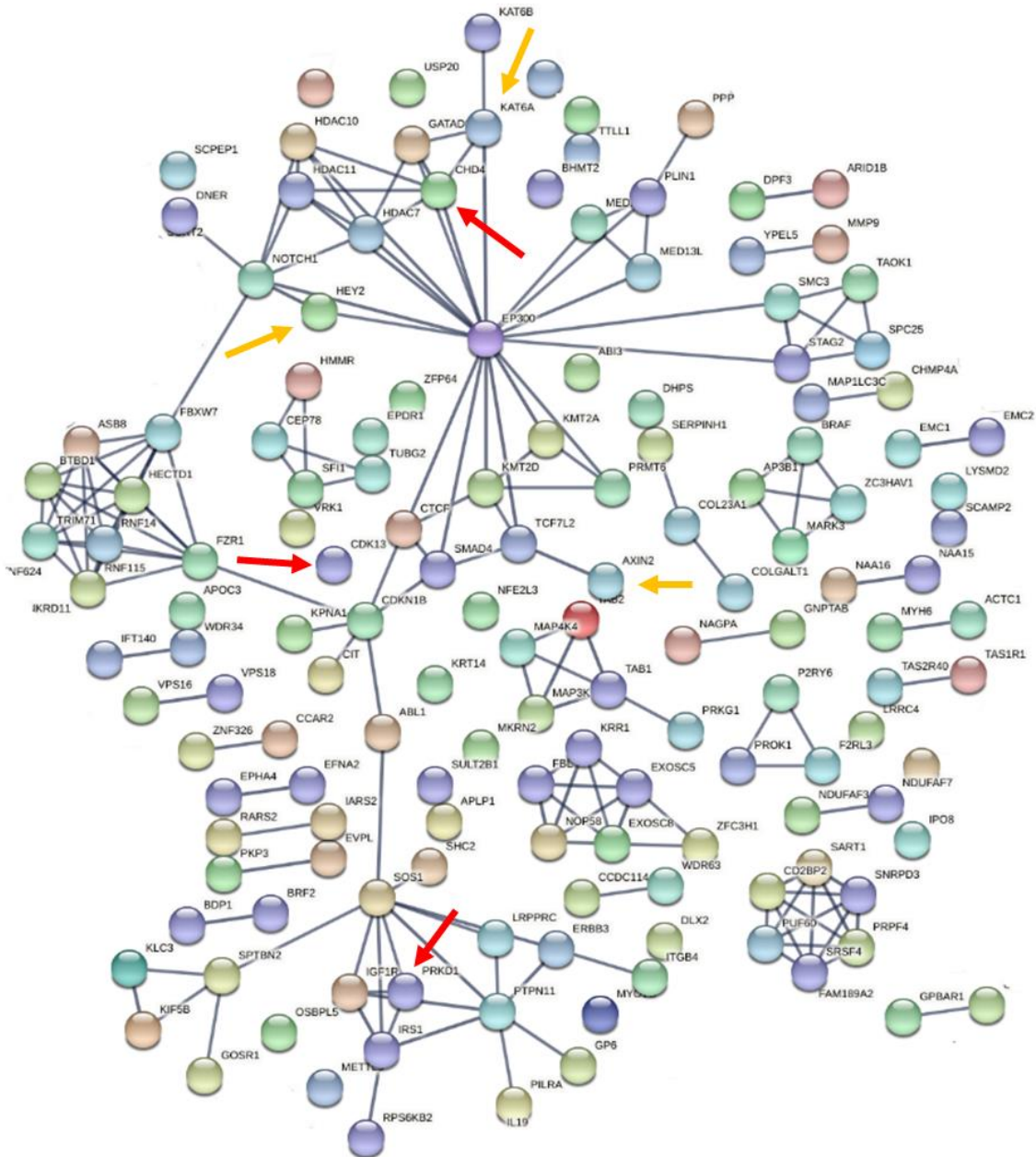
Figure 3.4 Updated protein-protein interactions.

High confidence protein interactions (>0.9) from the STRING PPI database in top-ranking set of CHD-associated genes (false discovery rate <50%) in the integrated de novo and inherited variant analysis (December 2020) [549]. CHD4, CDK13 and PRKD1 are all highlighted with red arrows. Yellow arrows indicate proteins encoded by novel CHD genes reported since, in close proximity to CHD4.

Additional proteins visible in figure 3.4, have since been identified as contributing to CHD, supporting the role of network maps to identify CHD gene candidates. KAT6A (yellow arrow), which is linked with CHD4, has been associated with CHD in some individuals, as part of a neurodevelopemental disability syndrome [567-569]. AXIN2 has been identified as a cause of CHD [658] and variants in HEY2 (yellow arrow) have now been shown to cause both CHD and thoracic aortic aneurysms [659]. It is also interesting that the PRKD1 protein in the figure above, is linked to both SOS1 and PTPN11, which are encoded by two common rasopathy genes. One of the three individuals we reported with mutations in *PRKD1* had pulmonary valve stenosis, a common feature seen in Noonan Syndrome. This could support functional connections between these genes in heart development.

Approaching novel gene discovery in CHD using pathway analysis has recently been used to propose 23 further novel genes [660]. These genes share biological functions and properties, and interact with known CHD genes, and may warrant further exploration to determine if they are true CHD genes. Pathway analyses such as these do not required such large cohorts as required for gene-level discovery and provide an alternative method for gene discovery.

To fully elucidate the genetic aetiology of CHD we will require much larger cohorts. We estimated that these may need to be 20 times the size of this current cohort (Supplementary Figure 5 in appendix N). For case control aspects of the study, much larger control datasets will also be required. Clearly this will require collaboration between research teams.

3.4 Challenges, Limitations and Future Work

As with any other large scale WES study, using data collected from different sources there were a number of challenges relating to standardising the data.

3.4.1. Defining S-CHD and NS-CHD

The definition of S-CHD and NS-CHD presented challenges, particularly in relation to neurodevelopmental disability (NDD). Children undergoing multiple heart surgeries, those who experience post-operative complications, premature babies and those who had additional episodes of illness may exhibit developmental delay due to prolonged hypoxia and complications of surgery, or just extended time in hospital. This means it is difficult to be sure if an individual's developmental delay is genetic or as a consequence of another factor. Recruitment to research at an early stage of life can mean it is difficult to accurately assess a child's development. Whilst additional information can be gained from brain scans, it is not always possible to predict disability from structural lesions alone.

Even with comprehensive medical records and investigations, it would be difficult to determine whether or not NDD was purely genetic or non-genetic. The proposed overlap between the genetic aetiology of NDD and CHD complicates this matter further. For this reason, all patients with developmental delay were considered to have S-CHD. Future studies in larger cohorts may help determine the contribution of genetic and non-genetic factors in CHD with NDD.

3.4.2 Parental Phenotyping

Ideally we would screen all parents with an ECG and echo to be sure that they are unaffected. It is not uncommon for people to be diagnosed with CHD such as haemodynamically insignificant septal defects and BAV in old age. Unfortunately this is cost prohibitive and there is the risk of detecting disease such as cardiomyopathy in asymptomatic individuals, which would require careful counselling prior to imaging. Instead, we assumed all parents did not have CHD as the Cardiologists/Clinical Geneticists would review any parent with a history of CHD or cardiac symptoms.

3.4.3 Sample Size

This study has identified that there are further novel CHD genes to be discovered, but that this requires a much larger sample size. This will require collaboration between researchers, which brings with it the difficulties of standardising platforms and pipelines, and differences in allele frequencies across different ethnic groups.

It was not possible to consider different CHD phenotype groups separately. This is because the number of individuals in each group were too small and genes would not reach genome wide significance. The same applied to analysis for enrichment of certain pathways including genes involved in diabetes, folate processing and alcohol for example.

Increasing the number of individuals with CHD who are sequenced will also identify further individuals with mutations in known CHD genes. This will allow us to describe the phenotypic spectrum and correlation with genotype more accurately. A total of 33 individuals have now been published with mutations in *CHD4* [12, 577, 588, 661]. CHD, in combination with developmental delay or mild to moderate intellectual disability remains a significant feature. 72% (21/29) had CHD, most commonly ASD, followed by VSD and pulmonary stenosis. PDA, TOF, mitral valve abnormalities, EA and TA have also been reported. Cerebral abnormalities are almost always present (92%, 22/24), most commonly ventriculomegaly. A further new feature that has been reported is Moya Moya disease (stenosis and or occlusion of carotid arteries and associated collateral vessels) [662]. Whilst the majority of mutations fall within the ATPase and C terminal helicase domains, genotype-phenotype correlation is still lacking [588].

3.4.4 Reduced Penetrance and Mechanisms other than Monogenic Inheritance

Further evidence supporting the role of inherited variants in CHD has been produced by Jin et al. who found an enrichment of damaging recessive genotypes in individuals with laterality type CHD defects, and conversely no excess of damaging de novo mutations[15]. This is thought to be accurate based on the high recurrence risk in this type of CHD and its prevalence in consanguineous populations [663].

The real challenge is looking for inherited variants with reduced penetrance. Li et al. identified inherited LoF mutations as a cause of left sided CHD in 4.4% of their cohort using WES [657]. Inherited mutations are likely to be important in left sided CHD given the high recurrence risk in children and siblings of up to 20%, albeit more subtle lesions. To understand this mechanism further, larger sample sizes are required. It may also be more successful in types of CHD where recurrence appears to be more common.

It is also important to consider compound mutations, even for genes associated with AD disease as shown by Jin et al [15]. Individuals with compound heterozygosity in known monoallelic CHD genes, presented with a more severe phenotype. Further evidence of reduced penetrance has also emerged. Page et al. identified mutations in *NOTCH1* causing TOF [19]. In two instances, the unaffected parents appeared to harbour the same mutation as the proband (1 maternal and 1 paternally inherited mutation). It is not clear whether the cardiac status of the parents was formally assessed.

Priest et al. found that putative inherited and de novo mutations across genes that cause human and mouse CHD, could explain the occurrence of AVSD in 32% of their relatively small cohort which supports an oligogenic model for CHD [353]. Given the reduced penetrance and variable expressivity we see in CHD,

and likely oligogenic model in most, we could consider using different allele frequency thresholds. It is possible that contributing variants are present at a higher frequency, or even in control cohorts in an oligogenic model. Searching for contributing genes under these conditions is more complex however. We have also not accounted for genetic mosaicism and somatic mutations in the heart. Going forward, it may be possible to collect paired cardiac tissue and blood samples to include analysis for somatic mutations. These approaches could help explain some of the reduced penetrance and variable expressivity.

3.4.5 Environmental and Maternal Factors

Evidence from mouse studies suggests that factors that modify CHD risk can greatly alter the chance of a predisposed individual developing CHD, such as maternal age in mice heterozygous for *Nkx2.5* mutations [664]. It is difficult to determine the importance of genetic variation, without understanding the contribution of additional non-genetic influences to CHD risk. Whilst it would be very difficult to take all environmental and maternal influences into account, future studies may be able to include maternal diabetes or other well established risk factors if prospective recruitment or a CHD registry could be established.

3.4.6 Whole Genome Sequencing

There are relatively few WGS based studies in CHD currently. Bjornsson et al. used WGS to identify common variants contributing to CoA in the Icelandic population [665]. This is a genetically defined and well characterised population. Hauser et al. performed WGS in a small cohort of neonates and children (34 trios) with CHD in America [537]. In 6% of individuals they identified a definite pathogenic variant, 26% had a variant of uncertain significance (VUS). They concluded that the diagnostic rate was insufficient to recommend WGS in routine practice for this group of individuals.

We have used WES in this instance as it is cheaper, was established the Sanger Institute and used for the DDD project, which formed part of our cohort [522]. Moving forward, WGS is likely to become less expensive. Certainly it seems that there is increasing evidence that non-coding and miRNAs could be important in CHD for example[666].

3.4.7 Further Functional and Animal work

Proteomics and nucleosome binding assays are ongoing within the Brook lab for the *CHD4*. A *Chd4* knockout mouse is being examined by S Loughna and J Aparicio within the lab, to study the effects on the heart using High Resolution Episcopic Microscopy (HREM).

3.5 Conclusions

We have shown that there are different mechanisms behind S-CHD and NS-CHD. We have identified three novel genome wide significant CHD genes, and have begun to describe the characteristic clinical features of, and potential mechanisms behind these new syndromes. We intend to complete additional functional work to add further support to the role of these genes in the heart.

This type of study also provides clinically useful results for participants, and other individuals with CHD. It allows expansion of the phenotype of known CHD genes. Results from large cohorts like this also provide clues as to the prevalence of syndromes associated with a particular gene, including a relative lack of genes considered important for heart development.

We can use the results generated from studies like this to construct networks of genes that are important for correct cardiogenesis, and identify further novel CHD genes. Results from studies such as this, and ideally larger cohorts, might be useful to guide clinical testing.

Moving forward, we require a much larger data set. This requires collaboration with other researchers and standardisation of data. Ethical aspects of untargeted sequencing and generation of secondary findings must be addressed as part of this. Data collection should be standardised and phenotyping should be as detailed and as accurate as possible, with consideration for additional maternal and environmental factors.

Following the WES in a cohort with CHD described in chapter 3, *de novo* missense mutations in *PRKD1* were identified as causing S-CHD [12]. Three individuals were reported, two of whom had an identical mutation.

PRKD1 is a serine/threonine protein kinase, which is involved in a number of basic cellular processes and has roles in many pathways including MAPK8/JNK1 and Ras-Raf-MEK-ERK. In the cardiovascular system, it controls cardiac hypertrophy and angiogenesis [667, 668].

This chapter explores what is known about the function of *PRKD1* in relation to the heart. The genotypes and phenotypes of the individuals identified with mutations in this gene are considered, and novel data is presented on two *Prkd1* mutant mouse lines, providing the first characterisation of their phenotypes.

4.1.1 An introduction to the Structure and Function of *PRKD1*

The expression of *PRKD1* is fairly ubiquitous and its functions are diverse, affecting many cellular processes and pathways. It has important roles in cell survival, fate, differentiation, proliferation, adhesion, migration and apoptosis [669-673] as well as basic cellular processes such as maintaining the Golgi apparatus and vesicle trafficking to the plasma membrane [674, 675]. It plays a role in cell response to stress and antigen challenges, as part of our innate immunity and inflammatory responses [670, 672, 676, 677]. It is also required for normal structure of neurones [675, 678-681]. Processes more relevant for cardiovascular development are outlined later in this chapter, but research so far seems to suggest that *PRKD1* has a wide range of functions in many different tissues and that there is some functional redundancy within the gene family.

PRKD1 is one gene in a highly conserved family of large serine threonine protein kinases, consisting of *PRKD1*, *PRKD2* and *PRKD3*. They share similar functional domains and high levels of sequence identity, especially within the kinase domain [673, 682, 683]. PRKD1 and PRKD2 are the most similar, sharing 85% amino acid sequence [684]. The mouse PRKD1, 2 and 3 proteins all share the same catalytic and pleckstrin homology domain sequence [685]. Signalling molecules like PRKD1 are required to move within the cell in response to specific stimuli. PRKD1/2 and 3 all shuttle within the cell once it has been activated, suggesting that all three are important in multiple signalling pathways [686].

There are many mechanisms that regulate the activity of PRKD1, including G protein coupled receptors, antigen binding and growth factors [686]. PRKD1 can be activated by a number of different molecules and physiological stimuli including phorbol esters, DAG, TNF and angiotensin II [673, 687]. Its activity is also influenced by phospholipase C and protein kinase C, and this is one of the more well described sequences [686, 688].

Within the N-terminal domain of the proteins, there are two cysteine rich (CRD) diacylglycerol (DAG) binding domains, and a single Plekstrin homology (PH) domain, which are required for auto inhibition [673]. The kinase domain is contained within the C-terminal region. The four main domains within PRKD1 are shown in Figure 4.1. These domains are all important in controlling the movement and activity of PRKD1.
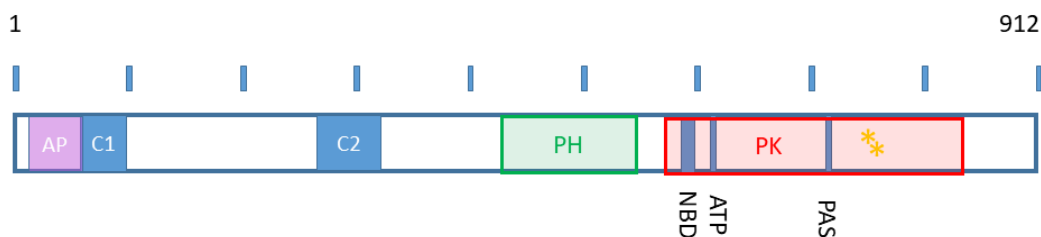
Figure 4.1 Structure of PRKD1.

PRKD1 is 912 amino acids (a.a.) long. Each 100 a.a. division is marked with a blue line across the top of the protein.

AP (lilac): Alanine proline rich region a.a. 17-146, part of the regulatory region.

C1 (blue): Cysteine rich domain 1. Similar to zinc finger domain, binding domain for phorbol esters/DAG a.a. 146/7-196/198.

C2 (blue): Cysteine rich domain 2. Similar to zinc finger domain, binding domain for phorbol esters/DAG a.a. 270/271-320/323.

PH (green): Pleckstrin Homology domain a.a. 423-541.

PK (red): Protein kinase domain: a.a. 584-839.

NBD: Nucleotide binding domain a.a. 589-597.

ATP: ATP binding site at residue 612.

PAS: PAS, proton acceptor site at residue 716.

Yellow asterisk: Serine phosphorylation sites at residue 744 and 748, fall within the catalytic domain.

Other important phosphorylation sites include Ser22 and 23[689], Tyr 95 [690], Ser203 [691], Tyr 463 [692], and Ser 916 [691].

a.a: amino acid

The N-terminal regulatory region of PRKD1 contains the cysteine rich domains (CRD), which are tandem repeats similar to zinc-finger motifs. The motif is structurally similar to others that bind DAG, and not surprisingly DAG binds to PRKD1 at these sites. DAG binding facilitates movement within the cell and the CRD are required for localisation to the Golgi compartment [693]. The two CRDs also appear to have different functional roles, the second CRD has a high affinity for DAG and predominantly binds phorbol esters [694], and is required for localisation to the membrane and nucleus of the cell without contribution from the first CRD [673].

The pleckstrin homology (PH) domain resides between the catalytic domain and CRDs [695]. This type of domain is seen in many proteins. It binds to membrane lipids, other proteins and is common in signal transduction proteins. The PH domain is required to maintain auto inhibition of PRKD1 [673, 682, 696]. Inhibition is removed when there is phosphorylation of serine residues within the activation loop of the kinase domain. This results in activation of PRKD1 [697]. The PH domain is also required for nuclear export of PRKD1 through a CRM-1 nuclear export pathway [698]and influences localisation of PRKD1 within the cell [673].

There is an additional PDZ binding motif at the C-terminal end of PRKD1. This interacts with proteins and may be important in locating specific proteins to the cell surface for correct signalling. It may also contribute to controlling the location of PRKD1 within the cell [686, 699].

4.1.2 Movement of PRKD1 within the cell

PRKD1 shuttles to and from the plasma membrane, and in and out of the nucleus, in response to a stimulus acting on the cell [700, 701]. In an unstimulated cell, PRKD1 mainly resides in the cytosol [700-703], with smaller amounts in intracellular compartments such as the mitochondria, Golgi and

transport vesicles. The amount and location varies depending on the cell population studied [687, 693, 700-708].

A stimulus, such as a molecule binding to a G protein coupled receptor, activates phospholipase C and DAG is produced [673]. Phospholipase C activation causes the second cysteine rich motif in PRKD1 to bind to diacylglycerol (DAG) at the inner plasma membrane [709]. This triggers recruitment of protein kinase Cs at the plasma membrane, and their activation. Movement of PRKD1 to the cell plasma membrane and into the nucleus requires DAG binding at the second CRD only. It is not required to bind to the first CRD. It also occurs independently of both the PH domains and activation by phosphorylation within the catalytic domain [700-703] .

Following its localisation to the cell membrane, PRKD1 is activated. Protein kinase C phosphorylates Ser744 and Ser748 within the activation loop of the catalytic domain [710, 711]. This phosphorylation releases PRKD1 from auto inhibition by the PH domain, and stabilises the active protein [673]. In addition, this phosphorylation is required for movement of PRKD1 back into the cytosol from the plasma membrane [700]. Auto phosphorylation then occurs at Ser 203 and Ser 916 [712] and afterwards, PRKD1 moves into the nucleus via binding at the second CRD [701]. The PH domain is required for subsequent nuclear export and deletion of this domain results in accumulation of PRKD1 within the nucleus [700, 701, 703].

In addition to the protein kinase C pathway of activation, other scenarios have been described. This includes cleavage of the regulatory domains and phosphorylation of Tyr 463 in the PH domain [672, 692]. However, the dominant mechanism appears to be activation by protein kinase C.

Usually PRKD1 locates to the cell plasma membrane and enters the nucleus, but alternative translocation pathways have been described in response to certain

stimuli, such as antigens binding to B cells and mast cells [702, 703]. Another slightly different mechanism appears to take place under conditions of oxidative stress when phospholipase D1 produces mitochondrial DAG and the CRD facilitates PRKD1 localisation to the mitochondrial membrane, rather than the cell membrane [713].

4.1.3 PRKD1 expression

Expression of mRNA of *PRKD* isoforms is fairly ubiquitous. The homology of *Prkd1* and *2* makes it difficult to distinguish their separate expression patterns, but *Prkd1* shows fairly widespread expression in the developing mouse (table 4.1) [714]. This includes the developing cardiovascular system. *Prkd1* and *Prkd2* are also expressed in the adult mouse heart [684].

Prkd1 is expressed slightly earlier than *Prkd2* (E9.5 to 14.5 in the mouse embryo, whereas *Prkd2* wasn't detected until E14.5). *Prkd3* was only weakly present initially and became more abundant after E10.5 [685].

Gene	Embryonic Day	Expression profile
Prkd1	E9.5	Generalised expression, but more highly expressed in the brain and mandibular and hyoid branchial arches, and the apical ectodermal ridge in the limb buds
	E10.5	Generalised expression, but more highly expressed in the brain and mandibular and hyoid branchial arches, Now seen in the whole limb bud, the heart and nasal processes.
	E14.5	Strong expression in parts of the brain, glomeruli, thymus, atria and ventricles of the heart, adrenal cortex and cartilage primordia in

		the axial skeleton. Weaker expression in a number of other tissues including bronchi and intestine
	E18.5	Suggested low level of expression in the heart
Prkd2	E9.5	Not detected
	E10.5	Not detected
	E14.5	High expression in sections of the brain, whiskers, olfactory epithelium, central region of the ventricle of the heart, intestine, thymus, lung mesenchyme, genital tubercle
	E18.5	Strongly expressed in whisker follicles, parts of brain, retina, olfactory epithelium, tooth primordia, tongue, cochlea, lung mesenchyme, kidney, intestine and glans penis
Prkd3	E9.5	Moderate expression in the brain. Also seen in limb buds and branchial arches.
	E10.5	Clear expression in the nasal processes and the heart. Also seen in limb buds and branchial arches.
	E14.5	Fairly ubiquitous, higher in parts of brain, olfactory epithelium, neuronal layer in eye, cardinal veins
	E18.5	Fairly ubiquitous, but levels now slightly reduced. Expression still stronger in parts of the brain olfactory epithelium, thymus, atria of heart, kidney capsule, liver, tongue.

Table 4.1 Expression patterns of the *Prkd* genes.

The different timings of expression of the different *PRKD* genes may mean that they do not exhibit complete functional redundancy, even when expressed in the same tissues [685].

Prkd1, 2 and *3* expression is also fairly generalised in the adult, suggesting they have important roles postnatally too. Two or more isoforms appear to be co-expressed in the kidney, liver, lung, skeletal muscle, spleen and uterus. *Prkd1, 2* and *3* are all expressed in the adult heart [684]. Only *Prkd2* is expressed in lymphoid cells.

4.1.4 Functional redundancy

There is some evidence of functional redundancy within the *PRKD* family across a number of species, but also evidence of different expression patterns and unique functions [678, 685].

Knockout mouse studies provide evidence that *Prkd1* has an essential role in embryogenesis, as the knockout mouse is embryonic lethal in the main [684]. A *Prkd3* gene trap mutant mouse displayed only mild skeletal defects (direct submission to MGI database and MGI30335138 Lexicon Genetics). *Prkd2* must also exhibit an element of functional redundancy as the homozygous knockout mice are phenotypically normal and do not show any evidence of reduced survival, fertility or abnormal embryogenesis [684]. This suggests that *Prkd1* is essential for embryogenesis in the mouse, whereas *Prkd2* and *Prkd3* are not.

Whilst *Prkd1* may be non-redundant in embryogenesis, there is evidence of functional redundancy between the different isoforms in certain cellular functions. PRDK1*, 2* and 3 are all able to phosphorylate class II HDACs to facilitate their nuclear export. In vitro experiments in cardiomyocytes have shown that knock down of *PRKD1* alone is not sufficient to completely abolish export of HDAC5 from the nucleus [715]. In vitro analysis of specific B cells that express both *PRKD1* and *PRKD3,* has shown that phosphorylation of HDAC5/7 was unaltered with knockdown of either *PRKD* isoform, but when both were knocked out there was loss of phosphorylation of HDAC5 and its nuclear export [716]. Matthews et al. summarised reports of functional redundancy seen

between the *PRKD* genes in processes such as protein transport, Golgi organisation and cell survival response, as well as phosphorylation of certain substrates [679, 717-725].

Despite evidence of more than one isoform being able to carry out a specific process, the embryonic lethal *Prkd1* knockout mouse suggests that the isoforms must also have specific essential functions. The CRDs in the different PRKD proteins have different affinities for different ligands [697] and within the same cell type, phosphorylation of a substrate cannot always be achieved by all PRKD proteins [678]. PRKD isoforms also reside in different areas of the cell and perform unique functions. This include vesicle trafficking by the Golgi based isoform and superoxide dismutase 2 expression by the mitochondrial isoform (PRKD1/2) [678, 684, 726]. There are also differences in the movements of the three isoforms within the cell. PRKD1 resides mainly in the cytosol and travels to the cell membrane and is activated in response to the cell being stimulated. It then quickly moves back to the cytosol and organelles [687, 727]. PRKD2 is also mainly cytoplasmic, but locates to the nucleus from the membrane [727]. PRKD3 is found in both cytoplasm and nucleus, even when not active [728].

Overall, despite evidence of functional redundancy in a number of processes including phosphorylation of class II HDACs and the fairly ubiquitous expression of the *PRKD* gene family, *Prkd1* is clearly essential for normal embryogenesis. *PRDK1* may be non-redundant because of its ability to carry out a specific task initiated by a specific ligand, further refined by specific temporospatial expression patterns [685, 729-732].

4.1.5 Mechanism of action

Histone deacetlylases (HDACs) are key proteins involved in transcriptional regulation of many important developmental programs within the body. PRKD1 can phosphorylate and inactivate a number of class II HDACs (HDAC4, 5, 7 and 9) (enhanced by BMP2) [715, 716, 733-735]. Phosphorylation of HDAC4/5 by

PRKD1, allows their export from the nucleus into the cytoplasm. This allows downstream gene expression, by removing gene repression by MEF2A-D transcription factors [736, 737]. This process is outlined in Figure 4.2. Although Calmodulin dependant protein kinase II (CaMKII) and IV (CaMKIV) and PRKD1 all phosphorylate class II HDACs, it appears that phosphorylation by PRKD1 rather than the CaMK genes causes nuclear export of HDAC4/5 and relieves transcriptional repression of the *MEF2* dependent genes.
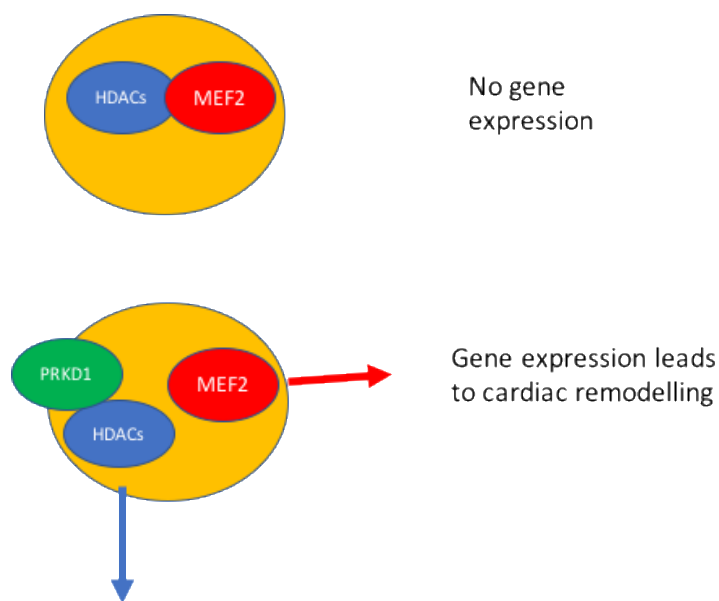
Figure 4.2 Mechanism of action of PRKD1.

Unphosphorylated class II HDACs remain in the nucleus and associate with MEF2 (red), preventing transcription of its targets. PRKD1 (green) phosphorylates the class II HDACS (blue), and they are exported (blue arrow) from the nucleus (yellow). *MEF2* repression is removed (red arrow) and allows downstream gene expression which results in pathological cardiac remodelling.

4.1.5.1 Role of PRKD1 In cardiogenesis

The influence of *PRKD1* on class II *HDAC*s and genes repressed by *MEF2*, is important in the heart. The *MEF2* genes are transcription factors, which are

expressed in cardiogenic precursor cells and differentiated cardiomyocytes during cardiac morphogenesis [738]. *MEF2C* is seen from E7.5 in the mouse cardiac mesoderm, which goes on to form the heart tube. mRNA from *Mef2a*, *2c* and *2d* are found in the myocardium at E8.5. *Mef2c-/-* mice develop an abnormal outflow tract and heart tube looping, and are embryonic lethal by E10 [739, 740]. Given the role of *PRKD1* in alleviating the repression of MEF2C, it is likely that it plays an important role in cardiogenesis, supporting its role in the pathogenesis of CHD.

4.1.5.2 *Role of* PRKD1 *in Cardiac* Hypertrophy and Heart Failure

In addition to formation of the heart, *Prkd1* plays an important role in the response of the heart to stress and resulting hypertrophic response. Cardiac remodelling occurs when the heart is exposed to increased stress from abnormalities such as hypertension, valve defects or ischaemia. *PRKD1* is activated in the failing myocardium of the adult human, rat and rabbit in response to stress from long-term neuro-hormonal signalling, hypertension and pressure overload [715, 741-744]. The role of Prkd1 in cardiac hypertrophy is considered further in chapter 5.

4.1.5.3 *Role of* PRKD1 *in* Angiogenesis and ischaemia

In addition to cardiogenesis and responses in the stressed heart, *PRKD1* is also influential in angiogenesis. *PRKD1* activity is increased by *VEGF* in endothelial cells [745], and a number of other stimuli in vascular smooth muscle cells [746-748]. *VEGF* is required for both normal and pathological angiogenesis [749]and *PRKD1* modulates *VEGF* related *ERK1* signalling and endothelial cell proliferation [745]. *VEGF* stimulates PRKD1 to phosphorylate HDAC7 in endothelial cells, resulting in its nuclear export, subsequent *MEF2* transcriptional activation and expression of genes required for angiogenesis, including cell migration and tube formation processes [667]. *Mef2c-/-* mice exhibit generalised vascular malformations and reduced expression of *Vegf* and *angiopoietin1* [739, 740].

PRKD1 may also be important in cardiac ischaemia. It is activated in response to reactive oxidative species produced in the mitochondria. This results in increased production of proteins required for cell survival under these conditions [690, 692, 726]. A number of the genes which show increased expression in this situation, including mitochondrial manganese dependant superoxide dismutase (*SOD2*), are thought to play a role in ischaemic preconditioning and myocardial protection [750]. PRKD1 is also known to phosphorylate TRPV1 and HSP27 [751, 752]. Both of these proteins are thought to contribute to limiting the extent of ischaemia after a myocardial injury [753, 754].

These studies taken together suggest that *PRKD1* has an important role in the contractile and hypertrophic response of the heart working against increased afterload. Given *PRKD1* induces expression of genes important in ischaemic preconditioning, it may play an important role in the ischaemic heart too. This makes it an interesting target in not just CHD, but also heart failure and ischaemia.

4.1.6 Existing *Prkd1* Mouse Models

A number of mouse models have already been generated to study the effects of abnormal *Prkd1* function. Feilitz et al. generated a cre-loxP mouse which deleted exons 12-14 and disrupts the catalytic domain of *Prkd1* [732]. They showed that the PRKD1 product was degraded. This resulted in embryonic lethality with incomplete penetrance and around a 1% incidence of *Prkd1*-/- mice surviving to term. These findings were corroborated by Matthews et al. [684]. They produced a *Prkd1* mouse without the serine phosphorylation sites required for activation. They found that around 1% of live births were *Prkd1*-/- mice (2/177). Lethality of the *Prkd1*-/- mice appeared to be between E9.5-12.5, normal mendelian ratios were produced prior to E9.5. This indicates that catalytic activity of *Prkd1* is almost universally essential for normal mouse embryogenesis.

In addition to the heart, the effects of *Prkd1* on bone development has been a topic of much research. *PRKD1* has an important role in osteoblast differentiation, mediating BMP2 induced nuclear export of HDAC7 and is also required for nuclear export of HDAC5 [715, 735]. A number of additional *Prkd1* knockout mice have been studied to determine the effects of *Prkd1* on bone density. Ford et al. demonstrated a reduction in bone mineral density [755] and another conditional deletion of *Prkd1* in osteoprogenitor cells results in young mice with reduced bone mass [756, 757]. A further conditional *Prkd1* knockout mouse has been produced and resulted in abnormal bone development and osteoblast differentiation [758]. The mice had craniofacial dysplasia. The scapulae and long bones were also affected.

The mouse models produced so far confirm that *Prkd1* is non-redundant in mouse embryogenesis and is required for cardiac hypertrophic response to increased afterload. It is also required for normal skeletal development. Previous mouse models confirm that no CHD is seen in mice lacking *Prkd2* or *Prkd3*. The mouse models I have studied are novel. The Wellcome Trust Sanger Institute generated the *Prkd1*[em1(IMPC)Wtsi] mouse, which has a different deletion (exon 2) to others reported previously, and the *Prkd1*[em2(IMPC)Wtsi] mouse which is the first humanised mutation mouse designed to replicate the commonest reported mutation causing CHD with ectodermal defects. More details about the mouse models are included in methods section 2.3.

4.2 Aims

The aims of this chapter are:

- To describe the genotype and phenotype of individuals reported by Sifrim et al. [12] and to consider the mechanism by which mutations in *PRKD1* cause abnormalities in humans.
- To assess *Prkd1*[em1(IMPC)Wtsi] and *Prkd1*[em2(IMPC)Wtsi] mice to determine the effects of loss of *Prkd1* on embryonic survival and formation of the heart using HREM.

- To develop an MRI protocol to characterise the cardiac phenotypes of adult *Prkd1*em2(IMPC)Wtsi mice.

4.3 Results and Discussion

4.3.1 Individuals Reported with mutations in *PRKD1*

Three participants were identified with de novo missense mutations in *PRKD1* [12]. Two individuals had identical mutations which fall within the nucleotide binding domain of the protein kinase domain (c.1774G>A, p.Gly592Arg). The third mutation falls within the second CRD (c.896T>G p.Leu299Trp). All three affected individuals have cardiac defects, two have AVSDs and the third has pulmonary stenosis (table 4.2). Because of the prominence of associated ectodermal features, the syndrome has been named CHD and Ectodermal Defects (https://www.omim.org/entry/617364).

ID	Mutation	Inheritance	Cardiac abnormality
268276	p.Gly592Arg	De novo	Pulmonary valve stenosis.
261674	p.Leu299Trp	De novo	AVSD
263568	p.Gly592Arg	De novo	AVSD

Table 4.2 CHD subtypes in *PRKD1*.

Genotypes and associated cardiac malformations identified in the three individuals with variants in *PRKD1*.

All three individuals reported with CHD and Ectodermal Defects are male. They had variable birth weights, but postnatal height/length appears to be within normal limits. Two exhibit microcephaly, and the other has macrocephaly. A single individual has an Arnold Chiari malformation. Developmental information suggests developmental delay in one, and learning disability and ADHD in

another. Two of the three had feeding difficulties, with one individual requiring NG feeding.

Ectodermal features appear to be a feature of this condition and one of the three individuals reported was initially thought to have an ectodermal dysplasia. Two of the three individuals have prominent ectodermal features including sparse scalp hair, dry/thin skin, fragile nails, small widely spaced teeth and premature loss of primary teeth. The third is noted to have a disorganised eyebrow with abnormal flare and widely spaced teeth only.

Digital abnormalities are also prominent and are described in two of the three individuals. These include a broad thumb, short digit, 2, 3 toe syndactyly and finger syndactyly. Photographs were available for two individuals. They share a prominent and wide forehead, with high anterior hair line and a stuck on chin. Ears appear anteverted. Other less common features included stridor, scoliosis, hypotonia, nystagmus and conductive deafness (n=1).

In addition to the three individuals we reported [12], a further individual is included on ClinVar with the diagnosis of CHD and Ectodermal Defects (clinical testing in Brazil) [521]. Unfortunately consent to include the individuals clinical information has not been given. All four mutations fall within important domains of the protein as shown in figure 4.3.
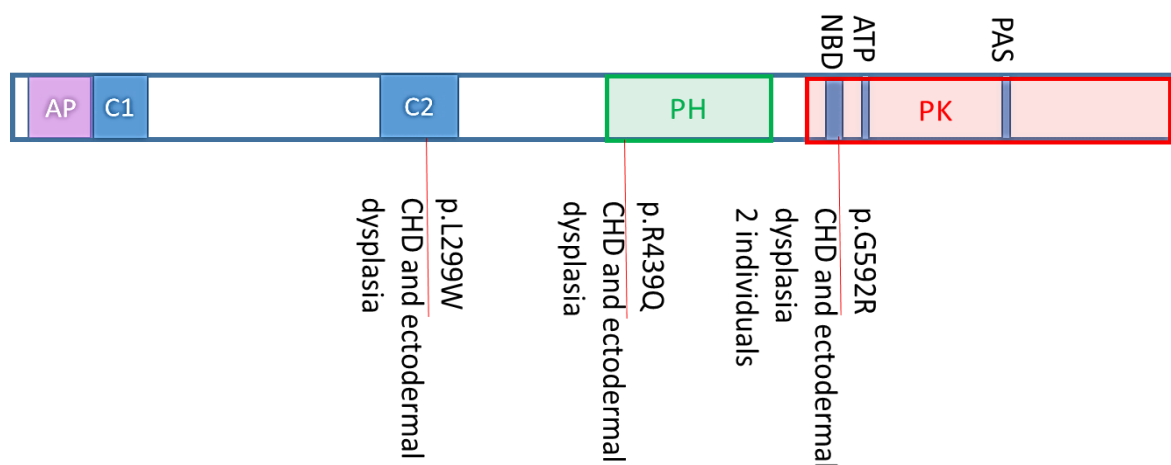
Figure 4.3 Mutations reported in association with CHD and Ectodermal Defects.

All are missense mutations affecting domains required for normal function of PRKD1. P.G592R and p.L299W were reported by Sifrim et al [12]. p.R439Q (c.29632945 G/A) is included on ClinVar [521], and reported by Mendelics Clinical testing Brazil, Rua Cubatao, 86, cj 1202Vila Mariana Sao Paulo Sao Paulo Brazil - 04013-000. All are reported in build 37. Uniprot does not suggest any modification occur at residues 299 and 592 [544].

PRKD1 as a cause of CHD and Ectodermal defects is supported by the correlation between the clinical picture reported in these individuals and its expression patterns during embryogenesis. The de novo occurence of mutations also supports is pathogenicity, although inheritance is unknown in the individual with p.R439Q. The mutations all fall within recognised domains within the protein that are required for normal function. The supporting evidence for *PRKD1* as a cause for CHD and Ectodermal Defects and a discussion of the possible mechanisms of action follows.

4.3.2 Genotype phenotype correlation

The small number of individuals reported with CHD and Ectodermal defects means that it is difficult to assess genotype phenotype correlation. However two individuals have been reported with the same variant. The two individuals with p.G592R both show ectodermal features and limb abnormalities. Both have sparse scalp hair, one has dry skin and the other has thin skin, one had premature loss of the primary teeth and the other had small teeth. One individual has a short digit and broad thumb, and the other has both toe and finger syndactyly. Both had feeding problems. Photographs were available for both individuals. They share similar facial characteristics, with a wide forehead and high anterior hair line, a stuck on chin and anteverted ears.

The individual with the p.L299W mutation differs as they had no digital abnormalities, and were the only individual with nystagmus and structural brain malformation (Arnold Chiari malformation). They had some ectodermal features (a disorganised eyebrow and widely spaced teeth), but not as convincingly as the two individuals with p.G592R mutations.

Features which do not correlate with mutation type include NDD, head circumference (occipitofronto circumference/OFC) and CHD type. The individual with p.L299W had delayed gross motor development, specific learning disability and ADHD. One of the individuals with p.G592R had developmental delay, but the second individual with this mutation does not. One individual with p.G592R has microcephaly and the other has macrocephaly. AVSDs were seen with both p.G592R and p.L299W.

Overall there is some evidence of genotype phenotype correlation with respect to digital abnormalities and possibly ectodermal features, but not CHD type or OFC. There is evidence of shared facial features between the two individuals we have photographs for, but we are limited as to what conclusions we can draw

between facial features and other abnormalities given the very small sample size.

4.3.3 Correlation of the clinical features with *PRKD1* expression

The core features of CHD with ectodermal dysplasia appear to correlate with *Prkd1* expression during embryogenesis. This includes expression in the heart, integument, brain and limb bud. *PRKD1* is fairly widely expressed, so its effects are likely to be wide ranging. Defects might not arise in other organs because of functional redundancy with *PRDK2* and *PRKD3* or other genes. The significance of other less common clinical features is currently unclear, and only further reports of affected individuals will determine whether they are part of the same disease process or not. Expression and clinical abnormalities in the limb, brain and ectodermal tissues are described here. A more detailed discussion considering the heart is included later in section 4.3.4 and 4.9.

4.3.3.1 Limb Abnormalities

Digital abnormalities have been reported in two individuals with CHD and Ectodermal defects, suggesting a role for *PRKD1* in the limbs. One individual has a broad thumb and short digit, and another has syndactyly of the fingers and 2, 3 toe syndactyly.[12]. A short broad thumb might imply radial ray defects and bony involvement. It is not clear if the upper limb syndactyly is bony or just soft tissue. 2, 3 toe syndactyly is relatively common and could be coincidental. Therefore more reports are required to determine the characteristics of associated limb defects.

Mouse models confirm expression in the limb bud of the E10.5 mouse embryo [685] and a role in development of normal bone structure [755, 758]. This supports limb defects as being a true phenotype of abnormal function of *PRKD1*. Whilst the majority of research involving *PRKD1* in the skeleton focusses on bone mineral density, there is also evidence that loss of *Prkd1* can result in other bony

abnormalities. 4 and 10 week old *Prkd1* conditional null mice show craniofacial dysplasia, abnormal scapulae and short long bones compared to WT mice. This is due to reduced osteoblast differentiation [758]. Reduced expression of *Prkd1* was seen in the calvarium and long bones. At 14 days old the mice had abnormal bone nodules in the scapulae and an under developed cranial cap, seen with delayed suture closure and calvarial defects. No specific comment was made upon the limb bones apart from length, and it is not possible to say from the data whether any specific deformities were present similar to those reported in humans. It is also possible that the phenotype in these mice may be in part due to the osterix-Cre transgene used to produce a conditional model, as it can induce a skeletal phenotype of its own [759]. This includes hypomineralisation around the sutures and in a number of craniofacial bones, so the significance of these findings is unclear. Specific craniofacial defects have not been reported in the human population as yet and it is not clear if there are any other similar abnormalities involving the scapula or long bones. Growth appears to be within normal limits in the two individuals we have measurements for, and there are no reports of delayed suture fusion although a single individual has brachycephaly (sifrim). Growth parameters are normal in *Prkd1+/-* pubescent mice so short stature may not occur [755].

Prkd1 is involved in maintaining bone mineral density. It is required for normal osteoblast differentiation and maturation [760]. *Prkd1* is activated by *BMP2* and *IGF1*. A number of mechanisms have been proposed to explain how *PRKD1* regulates bone formation, and include removing reactive oxygen species, its influence on *HDAC7*, activation and expression of osteoblast markers *STAT3* and *p38*, and by influencing β-catenin [756, 758]. No data is available regarding bone mineral density in the individuals with CHD and Ectodermal defects. Lower bone mineral density in *Prkd1* null mice normalises with time suggesting that other mechanisms must compensate for the loss of *Prkd1* [756]. *Prkd2* influences osteoclast differentiation and *Prkd3* knockout mice also display changes in bone density and therefore the action of these genes may compensate (Direct data submission to MGI 2005 from Lexicon Genetics) [546, 685, 758].

The association of *PRKD1* with *TBX5* is also important given that upper limb defects are seen in Holt Oram. This always includes carpal bone abnormalities and variable defects from the thumb, the arm bones, scapula and collar bone. The defects can be unilateral or bilateral. Detailed descriptions, or imaging of the limb defects seen in individuals with CHD and ectodermal dysplasia are not available. It would be interesting to review any imaging for carpal bone involvement for example to see if there is any cross over with features of Holt Oram or not. Holt Oram Syndrome is caused by loss of function mutations. Presumable loss of normal PRKD1 would prevent release of TBX5 from binding by HDACs and therefore reduced action of TBX5. However gain of function mutations in *TBX5* have been associated with atypical presentations [761] and as mentioned previously there are likely to be other compensating mechanisms.

Overall, evidence of expression of *Prkd1* during embryogenesis in the limb bud supports limb abnormalities as a feature of CHD and Ectodermal defects. The full phenotypic spectrum is yet to be determined. *Prkd2*, *Prkd3* and *TBX5* may all contribute to development in the limbs and more information is required about how they all interact. The importance of *Prkd1* in maintaining bone mineral density is uncertain since lower bone mineral density in mice normalises with time [756] and it is unclear whether this will be of relevance to individuals with heterozygous mutations in *PRKD1*.

4.3.3.2 Ectodermal Features

A number of ectodermal features have been reported and one individual was thought to have an ectodermal defect prior to discovery of their pathogenic *PRKD1* mutation. Dry skin, and thin skin with pigmented naevi are reported, along with fragile nails. Scalp hair has been described as sparse. Dental abnormalities reported include premature loss of primary teeth and small widely spaced teeth (n=2).

There is no specific information about dentition and *PRKD1*. *Prkd1* is expressed in the mouse skin at E16-E18 [546] and therefore it is possible that ectodermal defects are a result of abnormal *PRKD1* function. However, mouse models including conditional knockout in keratinocytes reveals that *Prkd1* is not required for normal skin and hair morphology, and numbers of proliferating epithelial cells are normal in conditional knockout models under normal conditions [762]. This suggests that Prkd1 is not required for normal skin development and maintenance. In contrast, *PRKD1* plays a non-redundant role in response of keratinocytes to stress [731, 763, 764]. Wound healing was slower in conditional *Prkd1* knockout mice as a result of reduced proliferation and migration of keratinocytes.

It is unclear why PRKD1 might result in ectodermal defects currently, and more research is needed. Reporting of further affected individuals will delineate the spectrum of associated ectodermal defects.

4.3.3.3 The brain

Prkd1 is strongly expressed in the mouse brain between E9.5-14.5 and has been shown to be important in neuronal links in the hippocampus [675, 765]. Two of the three individuals with phenotype data have features of central nervous system involvement. This includes delayed speech and language (n=1), ID (n=1), ADHD (n=1)) and structural abnormalities (microcephaly (n=2) and Arnold Chiari malformation (n=1)). Given the apparent shared developmental pathways and pathogenesis of NDD and CHD, combined with appropriate expression patterns, it is certainly possible that NDD is a result of the SNVs in *PRKD1*.

4.3.3.4 Other Features

Additional reported features include bilateral conductive hearing loss, stridor, scoliosis, nystagmus and hypotonia, which have all been reported in a single individual only. The significance of these features is currently unclear.

PRKD1 also appears to play an important role in insulin secretion and has been shown to promote its release in response to glucose [766]. Inducible *Prkd1* knock out mice developed hyperglycaemia and hyperinsulinaemia [767]. One individual born at term had hypoglycaemia along with jaundice and hypothermia, and also had a low birthweight which could explain these features.

4.3.4 What is the mechanism causing clinical abnormalities, including CHD?

Three different missense mutations have been associated with CHD and Ectodermal defects in the four individuals reported so far. Whilst *PRKD1* does not show at least two specific regions of missense constraint [768], all mutations fall within functional domains of the protein. The known functions of these domains suggests variants in the PH, second CRD and PK domains may all have a significant effect. Inheritance is de novo in all except for the mutation in the PH domain where inheritance is unknown. These factors all support pathogenicity of the variants, but the mechanism by which they cause clinical abnormalities is currently unknown. The possible consequences of the individual mutations and evidence for specific mechanisms of action are outlined in the following section.

4.3.4.1 p.G592R in the Protein Kinase Domain

The glycine residue affected by the variant p.G592R is conserved in a number of mammals including chimpanzee and mouse, as well as xenopus, c.elegans, zebrafish and chicken. The mutation falls within the protein kinase domain, and more specifically within the nucleotide binding domain (residues 589-597) (Figure 4.3 shown previously). This variant could interfere with ATP binding and normal function of the kinase domain.

Studies involving alterations in the kinase domain have already been carried out to assess the functional consequences [710, 769]. Mutation of both serine phosphorylation sites (Ser^{744} and Ser^{748} in the activation loop) to alanine resulted in an inactive kinase domain. Mutation of both serines to glutamic acid, mimics phosphorylation and produces an activated PRKD1 protein. Mutation of one serine to glutamic acid produces a partially active kinase. The changes in kinase activity is as you would expect, but protein localisation is also altered. Loss of function by mutation of the serine residues to alanine in cell models, resulted in normal translocation of PRKD1 to the plasma membrane, but delayed translocation from the plasma membrane to the cytosol [700]. This suggests that normal function and reverse translocation requires normal catalytic activity of PRKD1.

p.G592R could potentially result in loss of kinase function if it prevents ATP binding. This could result in both loss of kinase activity and potentially abnormal distribution of the protein. Alternatively it could result in a change in the nucleotide binding domain that allows another substrate to preferentially bind. For another substrate to activate the kinase, presumably it would need to act as a phosphate donor or induce a conformational change that results in activation. It would seem more likely that p.G592R would result in loss of kinase function. Functional studies are required to confirm this, but initial work by other members of the Brook laboratory suggests that p.G592R may result in reduced function of the kinase domain (work by T Ghosh and J Aparicio).

4.3.4.2 p.L299W in the Second CRD

p.L299W falls within the second CRD. Leucine at this position is conserved in a number of mammals including chimpanzee, mouse and chicken as well as xenopus and zebrafish, reinforcing the importance of this domain. Although there are two CRDs in *PRKD1*, it is this second CRD that has the higher affinity

for phorbol esters and the two CRDs are not equivalent in function [693, 694, 696].

The second CRD binds DAG at the plasma membrane, and is required for translocation of PRKD1 into the nucleus [700, 701]. Studies have also shown that the two CRDs regulate the activity of PRKD1. Deletion of either CRD, or both together resulted in increased activity of PRKD1 in COS-7 cells [696]. However, mutations of conserved proline residues which contribute to PRKD1's affinity for phorbol esters within the two CRDs, yielded different results. p.P287G in the second CRD resulted in similar kinase activity to the WT PRKD1 protein. However, this mutation also severely reduced the bombesin stimulated translocation of PRKD1 to the plasma membrane in Swiss 3T3 fibroblasts and MDCK cells (Madin-Darby canine kidney cells) [700]. In contrast, p.P155G in the first CRD, resulted in significantly increased activity of the mutated protein compared to WT. This suggests that the two CRDs do not have the same effect on the activity of PRKD1. Given the mutated prolines were chosen specifically, it is not possible to predict what effect p.L299W might have on the kinase activity of PRKD1.

Based on these results, p.L299W could alter DAG binding, movement to the plasma membrane and nucleus, and potentially kinase activity. If any of these functions were reduced, we might see reduced phosphorylation of HDAC targets and continued repression of MEF2 gene targets. Functional studies should be carried out to confirm if like the p.P287G mutation, kinase activity is unchanged. In addition, studies of movement of the mutant protein are needed to determine if p.L299W alters translocation to the plasma membrane and the nucleus.

4.3.4.3 p.R439Q in the PH Domain

p.R439Q falls within the PH domain. PH domains are required for binding to membrane lipids and other proteins [698] and in PRKD1, this domain is responsible for autoregulation [673, 682, 696]. The PH domain is also

necessary for formation of stable complexes with certain PKCs, including PKCη [770], localisation of PRKD1 [698] and its nuclear export [703].

A number of experimental studies have determined the effects of a deleted or mutated PH domain. Deletion of the PH domain altered the resting location of PRKD1 in lymphocytes [703]. Instead of residing in the cytosol, PRKD1 was found in both the nucleus and the cytosol. On activation of the cell, PRKD1 could still locate to the plasma membrane, despite the altered PH domain. This suggests that the PH domain is important in controlling the localisation of PRKD1 in inactivated cells. Deletion of the PH domain also affects activity levels of PRKD1 in COS-7 cells [695]. Deletion of the whole PH domain resulted in very high levels of activity. Partial deletion of the PH domain also increased activity levels, as did mutations p.R447C and p.W538A which affect semi conserved and conserved residues within the domain respectively. Importantly, p.R447C modelled in the COS-7 cells appears to be just two residues from p.R439Q reported in the individual on ClinVar.

Based on this evidence, it seems likely that p.R439Q will alter the protein function and could result in gain of function. The effect on localisation of the protein in unclear. Both kinase studies and localisation studies using a mutated protein are required to determine the effect of the mutation reported in this individual.

4.3.4.4 Additional information to help determine the pathogenic mechanism of *PRKD1* mutations

Considering the domains affected by the three missense mutations reported so far and the functional evidence available, it would seem most likely that p.G592R and p.L299W result in loss of function, but p.R439Q may result in a gain of function. Whether normal shuttling of the mutated proteins occurs is unclear. If translocation of the protein is altered it could increase or decrease the ability of PRKD1 to phosphorylate HDACs based on the resulting levels in the nucleus. To

try and determine how aberrations in PRKD1 result in abnormalities, we can also consider additional reported variants of unknown significance and haploinsufficiency as a mechanism.

Two heterozygous truncating mutations have been reported on public databases, one on Decipher and one on ClinVar [496, 521] (figure 4.30). Both are of unknown significance. One falls within the protein kinase domain (p.Y760*) and was detected during clinical testing in Germany (ClinVar). Unfortunately, despite contacting the laboratory, I have been unable to find out any further information. The other falls between the two CRD (p.Q143*), and is associated with moderate global developmental delay, muscular hypotonia, a depressed nasal bridge, protruding tongue and synophrys. Discussion with the clinician involved has confirmed there are no cardiovascular, limb or ectodermal features and the individual has a primarily neurodevelopmental phenotype. The mutation has been designated a class 3 variant. The individual also has two other variants of uncertain significance, one of which is in a gene associated with NDD.

Whilst NDD does appear to be part of the phenotype associated with missense mutations in *PRKD1*, it has not been seen in isolation in the individuals we have reported. However, the Samocha et al. dataset includes a further mutation in PRKD1; p.Phe306Tyr (figure 4.4) [768]. This was found in a cohort of 5620 individuals with NDD. If this mutation proved to be pathogenic and associated with NDD in isolation, we would have to reconsider if p.Q143* might be significant. However p.Phe306Tyr falls within the second CRD and is therefore more likely to be significant. This variant has a MPC score of 1.1226 and could be pathogenic (compare to p.Gly592Arg which has a MPC of 1.2269, and p.L299W with an MPC score of 1.442). Unfortunately there is no information regarding the phenotype, inheritance or whether they have another potentially pathogenic mutation as the data is anonymised. This means we cannot be sure if isolated NDD is a feature of *PRKD1* mutations and whether either of these mutations are pathogenic.
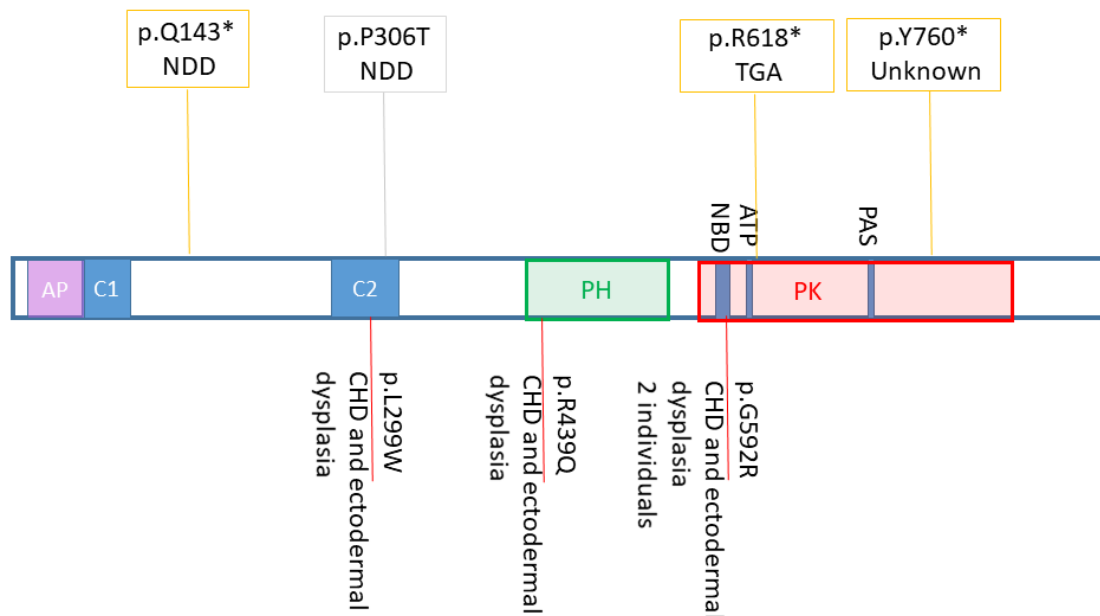
Figure 4.4 Additional variants reported in PRKD1.

Mutations above gene: PTVs of uncertain significance [496, 521] and presumed pathogenic missense mutation associated with NDD only [768]. A further pathogenic homozygous truncating mutation has been reported as causing NS-CHD, this is p.R618* [771].

Mutations below gene: pathogenic missense mutations [12, 521]. Y760X from Clinical testing CeGaT Praxis fuer Humangenetik Tuebingen, PHGT Paul-Ehrlich-Str. 23.

The pLI of *PRKD1* is 0.0 [772], suggesting that this gene is tolerant of LoF mutations and therefore these truncating mutations may be benign in the heterozygous state. However, a family with truncus arteriosus and homozygous PTVs in *PRKD1* has also been reported [771]. Two healthy half cousin parents from a Saudi Arabian family had two affected daughters with TA, and a further six children who were well. The affected daughters had NS-CHD (at the time of reporting they were 12 and 16 years old and had recently been evaluated). Whole exome sequencing was carried out on DNA from the two affected siblings

only and used to identify homozygous variants in coding regions or splice variants with a frequency of less than 0.001 in dbSNP. Autozygosity mapping was then carried out. They then narrowed candidate SNPs down to those identified in autozygous regions between the siblings and removed those that were also seen in a sample of 549 Saudi individuals who had had WES. This filtering process left only a homozygous nonsense mutation p.R618* in *PRKD1* (NM_002742:c.1852C>T). The authors then went on to look at another 17 individuals with TA, of which only 20% had consanguineous pedigrees. They identified three individuals with 22q11.2 deletion syndrome who were excluded and targeted sequencing of *PRKD1* in the rest revealed no further mutations. It is important to note that the parents were not tested for the presence of the mutations. The mutation is recorded on ClinVar as being of uncertain significance.

The p.R618* mutation maps to the protein kinase domain. This begins at residue 595, so truncation occurs relatively near the beginning of the kinase domain. The consequences of homozygosity for p.R618* are unknown. If the mutation resulted in complete loss of function, then presumably the individuals would not have survived (with the obvious caveat that we cannot assume the mouse model always translates directly to humans). This suggests that the protein might retain some function and could represent a partial loss of function mutation. Initial work by T Ghosh and J Aparicio in the Brook laboratory suggests that p.R618* may result in reduced function of the kinase domain. Alternatively, there could be another explanation for the phenotype and the truncating mutations are coincidental.

p.Y760* is also located in the PK domain [521]. Although nearby, more of the kinase domain will be intact (if the protein is not degraded) and obviously this individual has a normal WT protein to compensate, whereas the siblings with p.R618* do not. It is reported as being of uncertain significance. Phenotypic information for the individual carrying p.Y760* is required along with functional studies to determine if there is nonsense mediated decay or any function in a

residual protein to help determine the pathogenicity of these truncating mutations, especially given that the pLI value suggests *PRKD1* should tolerate LoF mutations.

The haploinsufficiency index of PRKD1 is 2.79%, which suggests that it is more likely to exhibit haploinsufficiency [773]. DGV records a number of deletions within *PRKD1* in the normal population. The majority are small and intronic, and therefore possibly benign. This could support haploinsufficiency of *PRKD1* as being detrimental. There are a number of deletions involving half or the whole of PRKD1 in isolation associated with NDD and structural brain abnormalities recorded on decipher [496]. The significance of most is unclear as inheritance is unknown or the individual has other CNVs in addition. Larger deletions including neighbouring *FOXG1* have been reported as causing atypical Rett syndrome, which shares a NDD phenotype. *PRKD1* deletions could affect the function of nearby *FOXG1*. Overall the significance of the deletions including *PRKD1* is unclear at this time.

It is difficult suggest a unifying mechanism to account for all these mutations. Considering the phenotypes it is also difficult to explain why homozygosity for p.R618* results in NS-CHD, but heterozygous missense mutations in the same domain result in S-CHD and a more severe phenotype if both act via the same mechanism. More than one mechanism could account for phenotypic differences. I have so far not considered the possibility of a dominant negative process, by which PRKD1 might be able to sequester HDAC and other proteins resulting in a more severe phenotype. Perhaps homozygosity for p.R618* results in reduced function and has fewer detrimental effects than a single heterozygous missense mutation where the protein retains more function. Different levels of PRKD1 function could result in different levels of phosphorylation of HDAC molecules and different levels of *MEF2* target gene expression.

Considering all the information available currently, the mechanism by which aberrations in *PRKD1* cause disease is uncertain. The haploinsufficiency index

and lack of significant deletions in a normal population suggest haploinsufficiency is plausible, but reports of individuals with deletions of *PRKD1* and CHD and Ectodermal defects are lacking. Both gain and loss of function are possible. Loss of function as a cause of CHD and ectodermal defects would go against the pLI value of 0 however.

It is also possible that simply considering mutations to cause gain or loss of function is too simplistic given that translocation of PRKD1 is likely to be affected in addition, and that it performs many tasks in many different cellular locations under different conditions. Ultimately, further reports of individuals with mutations and deletions in *PRKD1*, paired with functional testing will help unravel the mechanisms behind CHD with ectodermal defects and whether the clinical spectrum includes isolated CHD and isolated NDD.

4.4 Clinical Recommendations

At this stage it is difficult to make clinical recommendations given the small number of individuals reported and the lack of phenotype data about many of them. The first step needs to be a more detailed collection of clinical features. This should be followed by regular review for any new phenotypes. Identification of other affected individuals is also necessary to understand the full spectrum of signs and symptoms. Based on the information we have available I would suggest the following:

Consider genetic testing in individuals with:

- CHD with ectodermal defects.
- CHD with limb abnormalities, after review for features of more well described hand heart syndromes (Holt-Oram, Okihiro, Ulnar Mammary, Rubenstein Taybi, Smith Lemli Opitz and Oral Facial Digital Syndromes) and appropriate testing.

Suggestions for investigations at this stage include:

- Echocardiography

Echocardiography is a low risk and non-invasive investigation that can detect most structural lesions. This should happen at diagnosis and is likely to form part of the ongoing follow up of individuals with CHD. Given the role of *PRKD1* in cardiac hypertrophy and contractility [689, 741, 774], it would be worth noting any evidence of hypertrophy or conduction abnormalities. This may be addressed at cardiac follow up appointments or if symptoms arise.

- Careful review at diagnosis for limb defects and appropriate management.

Currently there is only evidence of mild abnormalities.

- Review for ectodermal defects and appropriate management.

Currently there is no information on whether there is any involvement of sweat glands, which might necessitate additional advice and management.

- Monitor growth parameters and suture closure.

This should include OFC and imaging if required. There is conflicting evidence as to whether *Prkd1* in the mouse causes short stature or not. Mouse models suggest that *Prkd1* function is redundant in maintaining post pubescent bone mineral density, but it might be worth considering DEXA scanning if there are multiple or low impact fractures.

- Monitor and support development and learning as required.

Structural brain malformations have been seen, but given the small cohort of individuals, it is unclear how common a feature this is. It would be appropriate to consider imaging if there are additional neurodevelopmental concerns.

The above suggestions should be refined as more information becomes available. It would be beneficial to report any abnormalities of glucose homeostasis given *PRKD1* is involved in control of insulin release [766].

Much of the research about *PRKD1* involves its role in cancer as *PRKD1* is part of an important pathway regulating cell senescence [775]. A mutation hotspot has been identified in *PRKD1* in polymorphic low-grade adenocarcinoma of the salivary glands. These missense mutations fall within the catalytic loop of the kinase domain (p.E710D) [776]. This mutation results in increased kinase activity and increased cellular proliferation in a cellular model suggesting this mutational hotspot is oncogenic. Genomic rearrangements of the *PRKD* genes have also been described in carcinoma of the salivary glands [683]. The significance of this in individuals with germline mutations in PRKD1 is uncertain. Incidences of cancer should be reported to help determine if there is any risk to individuals with CHD and Ectodermal defects.

4.5 Results from the *Prkd1* Mice

Results from both Prkd1 mouse lines are detailed sequentially here. A discussion of the results of both lines together is included afterwards.

4.5.1 Confirming genotype in a *Prkd1*[Em1(IMPC)Wtsi] mouse

The *Prkd1*[Em1(IMPC)Wtsi] mouse was supplied by the WTSI using Crispr/Cas9, and deletes exon 2 and some of the surrounding intronic sequence. Exon 1 is the first coding exon. Exon 2 precedes the first cysteine rich domain and overlaps its start (CRD residues 123-195, exon 2 residues 87-133).

Results for a wild type mouse are shown in figure 4.5. The 1360bp wild type band did not always amplify. This was detailed in the mouse information sheet from the WTSI. A LongAmp (NEB LongAmp Taq DNA Polymerase M0323)

protocol was used to preferentially amplify this band, following the manufacturer's suggested PCR programme. Gel extraction was then completed and the product sent for sequencing. This confirmed the product produced was the wild type allele.
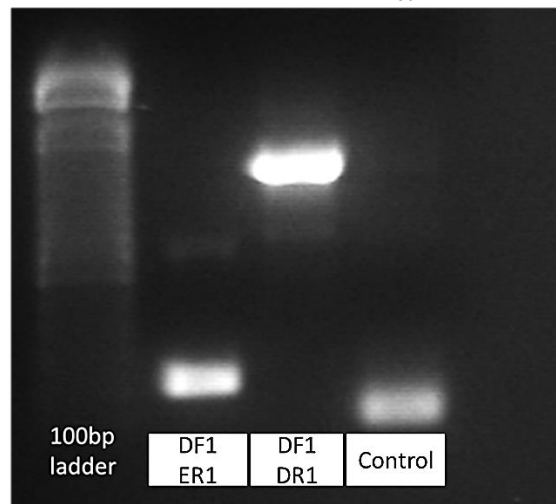
Figure 4.5 *Prkd1*$^{+/+}$ mouse genotype.

PCR results after running primer combinations on *Prkd1*$^{+/+}$ mouse DNA shows the expected band at 223bp and 1360bp. GAPDH primer set used as a control. 1% gel, Invitrogen 100bp ladder.

Confirming the genotype in the *Prkd1*$^{Em1/+}$ and *Prkd1*$^{Em1/+}$ mice proved more difficult initially, due a larger than expected deletion. The results are shown below. As expected for *Prkd1*$^{Em1/+}$ mice, DF1ER1 produced a band at 223bp and DF1DR1 produced a band at 1360bp. However, there was no band at 212bp in DF1DR1 and instead there was an unidentified 400bp band (Figure 4.6).
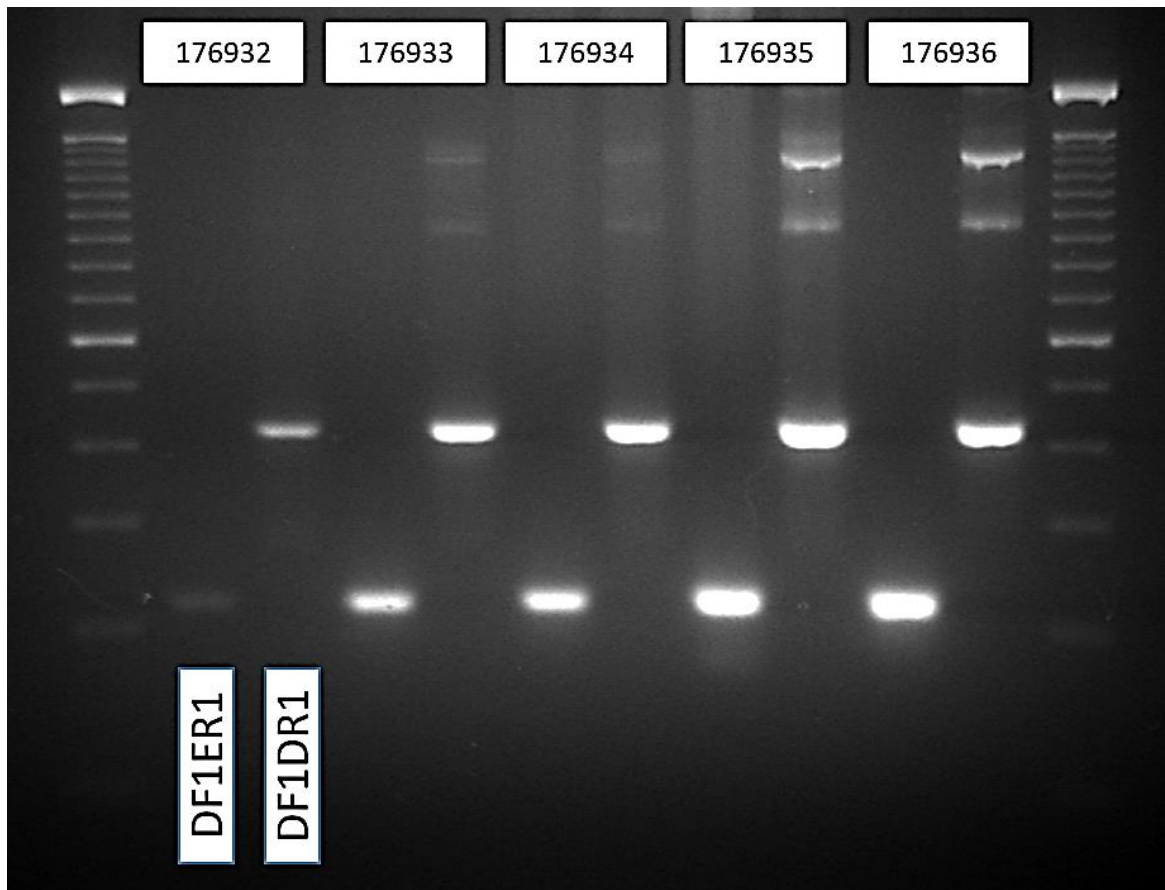
Figure 4.6 _Prkd1_^{Em1/+} genotype.

PCR products from the two primer pairs in 5 _Prkd1_^{Em1/+} mice. 100bp ladder and 3% gel. Each mouse has a six digit ID, shown at the top of the figure. Results for the two primer sets are shown for each mouse in the order of DF1ER1 on the left and DF1DR1 on the right.

To ascertain why the DF1DR1 band was 400bp rather than the expected 212bp, I performed a gel extraction and sent the product for Sanger sequencing. The results for the 200bp (Figure 4.7) and 400bp bands (Figure 4.8) are shown below and confirm that the deletion was smaller than originally documented by the WTSI.

Figure 4.7 Sequence of *Prkd1*^Em1/+ mouse.

Results of PCR using primers DF1 and ER1 produces 200bp band. Gel extraction and sequencing shows that the sequence maps to the correct area, but is slightly shorter that the expected 223bp calculated. Primers highlighted (ER1 orange, DR1 green).
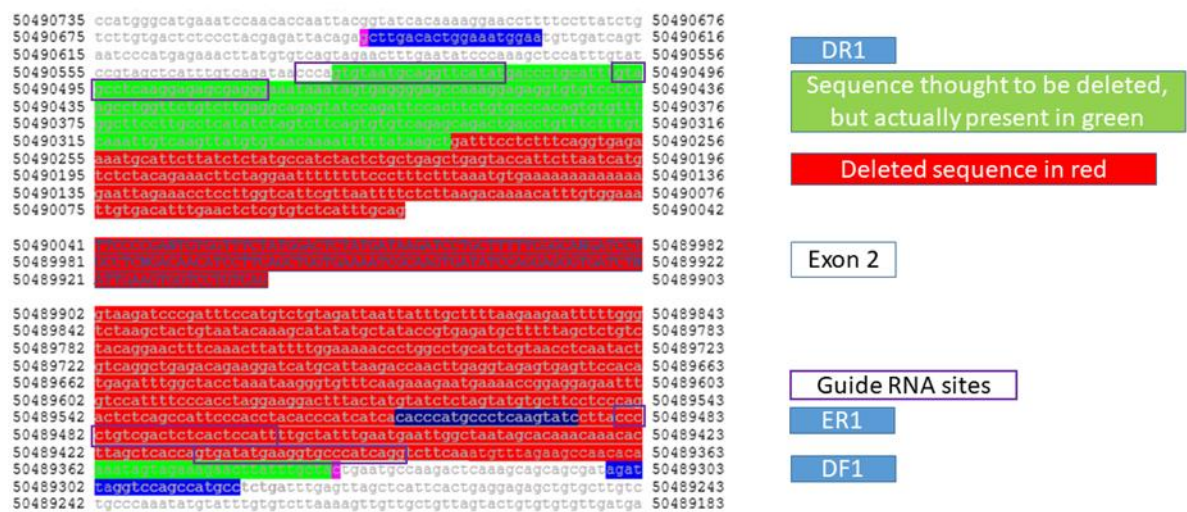
Figure 4.8 Deleted sequence in the *Prkd1*^Em1/+ mouse.

Sequence results produced from PCR using primers DF1 and DR1 of the deleted allele. The deleted section is smaller than expected, which results in a larger band on the gel. The information from the WTSI suggested that both red and green sections would be deleted, whereas only the red section had been deleted. This still contains exon 2 however. Primers are highlighted in dark blue, DR1at the top and DF1 towards the bottom of the sequence. The bases highlighted in

pink show the first and last bases of the sequence obtained. Guide RNA sites used by the WTSI to generate the mouse are highlighted in purple boxes.

In light of the discrepancy in the size of the deletion I contacted the WTSI. Subsequent correspondence confirmed my findings and confirmed the mouse supplied had the smaller deletion. The updated expected band sizes are shown below in table 4.3.

Genotype	Primer pair	PCR Product
Prkd1^{+/+}	DF1ER1	223bp
	DF1DR1	1360bp
Prkd1^{Em1/+}	DF1ER1	223bp
	DF1DR1	397p and 1360bp
Prkd1^{Em1/Em1}	DF1ER1	No product
	DF1DR1	397bp

Table 4.3 Updated PCR for *Prkd1*^{Em1(IMPC)Wtsi} mouse.

The final band sizes for confirmation of genotype in the *Prkd1*^{Em1(IMPC)Wtsi} mouse with the smaller deletion than originally documented.

4.5.2 Ratios of Genotypes Produced in the *Prkd1*^{Em1(IMPC)Wtsi} Mouse

Genotypes were recorded for all mice processed, as part of normal colony maintenance and heterozygous crosses for embryo collections, to observe the ratios of each genotype produced.

Embryos

The results of the embryo genotypes are shown in table 4.4 and figures 4.9 and 4.10. There is evidence of embryonic lethality in the knockout mice, but deaths were also seen in WT $Prkd1^{+/+}$ and $Prkd1^{Em1/+}$ embryos. Genotype was confirmed twice initially to ensure assays were correct.

Gestation	Cross	Expected	Observed
E9.5	$Prkd1^{Em1/+}$ x $Prkd1^{Em1/+}$	Em1/Em1:25% Em1/+: 50% +/+: 25%	Em1/Em1: 0 Em1/+: 3 (25%) +/+: 4 (33%) Unknown: 5 (42%)
E10.5	$Prkd1^{Em1/+}$ x $Prkd1^{Em1/+}$	Em1/Em1:25% Em1/+: 50% +/+: 25%	Em1/Em1: 0 Em1/+: 4 (50%) +/+: 2 (25%) Unknown: 2 (25%)
E10.5	$Prkd1^{Em1/+}$ x $Prkd1^{+/+}$	Em1/Em1: 0% Em1/+: 50% +/+: 50%	Em1/Em1: 0 Em1/+: 2 (20%) +/+: 4 (40%) Unknown: 3 (30%) Failed: 1
E11.5	$Prkd1^{Em1/+}$ x $Prkd1^{Em1/+}$	Em1/Em1:25% Em1/+: 50% +/+: 25%	Em1/Em1: 0 Em1/+: 8 (80%) +/+: 1 (10%) Unknown: 2 (20%)
E12.5	$Prkd1^{Em1/+}$ x $Prkd1^{Em1/+}$	Em1/Em1:25% Em1/+: 50% +/+: 25%	Em1/Em1: 0 (0%) Em1/+: 8 (89%) +/+: 0 (0%) Unknown 1 (11%)
E15.5	$Prkd1^{Em1/+}$ x $Prkd1^{Em1/+}$	Em1/Em1:25% Em1/+: 50% +/+: 25%	Em1/Em1: 0 (0%) Em1/+: 24 (50%) +/+: 15 (31%) Unknown: 9 (19%)
E15.5	$Prkd1^{Em1/+}$ x $Prkd1^{+/+}$	Em1/Em1:0% Em1/+: 50% +/+: 50%	Em1/Em1: 0 (0%) Em1/+: 4 (29%) +/+: 8 (57%) Unknown: 2 (14%)

Table 4.4 $Prkd1^{Em1(IMPC)Wtsi}$ embryo genotypes.

Genotypes of all $Prkd1^{em1(IMPC)Wtsi}$ embryos produced. In some cases the embryo had died at an early stage of development and sufficient tissue to allow genotyping could not be collected. In this instance, the genotype is recorded as unknown. +/+: $Prkd1^{+/+}$, Em1/+: $Prkd1^{Em1/+}$, Em1/Em1: $Prkd1^{Em1/Em1}$.
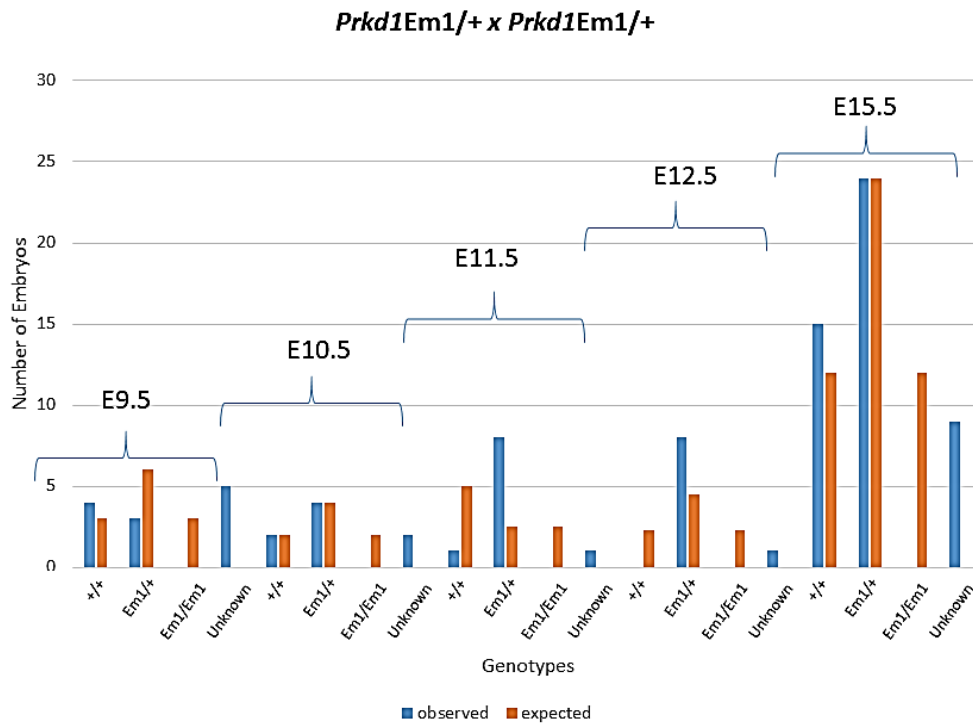
Figure 4.9 Genotypes from *Prkd1*^{Em1/+} crosses.

Numbers of each genotype produced from *Prkd1*^{Em1/+} x *Prkd1*^{Em1/+} matings, compared with expected, at each gestation. *Prkd1*^{Em1/Em1} embryos are not seen at any gestation.

+/+: *Prkd1*^{+/+}, Em1/+: *Prkd1*^{Em1/+}, Em1/Em1: *Prkd1*^{Em1/Em1}.
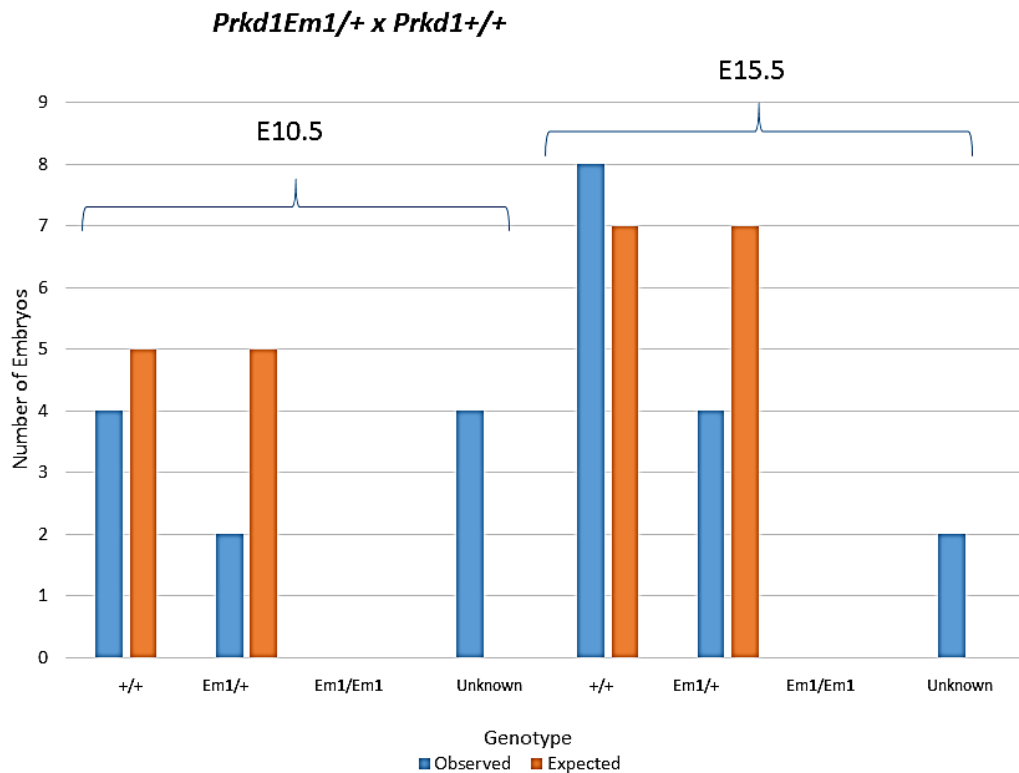
Figure 4.10 Genotypes from *Prkd1*[Em1/+] WT crosses.

Numbers of each genotype produced from *Prkd1*[+/+] x *Prkd1*[Em1/+]crosses, compared with expected over all gestations at which embryos were collected. +/+: *Prkd1*[+/+], Em1/+: *Prkd1*[Em1/+], Em1/Em1: *Prkd1*[Em1/Em1].

No null *Prkd1*[Em1/Em1] embryos were found at any gestation indicating embryonic lethality. It is likely that they die at an early gestation. It is possible that some of the embryos of unknown genotype where insufficient tissue was available for genotyping, were actually *Prkd1*[Em1/Em1]. At E9.5, I was unable to locate an embryo in five instances. There was no clear evidence of an embryo, reabsorption, or necrosis. These are the five embryos with unknown genotypes. The significance of this is unclear.

The numbers of *Prkd1*[+/+]and *Prkd1*[Em1/+] embryos produced at each gestation are variable. The numbers at each gestation are small, with the exception of E15.5. In this group, numbers of both *Prkd1*[+/+] and *Prkd1*[Em1/+]mice are around the expected ratios. This is probably the most accurate, and supports homozygous

lethality with no definite evidence of worsened survival in the heterozygotes. However larger numbers are required to confirm this and more data from E9.5 should be collected.

Live born Mice

Genotype results for live born mice are displayed in table 4.5 below. There were no $Prkd1^{Em1/Em1}$ mice, indicating complete embryonic lethality.

Gestation	Cross	Expected	Observed
Adult	$Prkd1^{Em1/+}$ x $Prkd1^{Em1/+}$	Em1/Em1:25% Em1/+: 50% +/+: 25%	Em1/Em1: 0 Em1/+: 40(62.5%) +/+: 17 (26.5%) Unknown: 7 (11%)
Adult	$Prkd1^{Em1/+}$ x $Prkd1^{+/+}$	Em1/Em1: 0% Em1/+: 50% +/+: 50%	Em1/Em1: 0 Em1/+: 18 (27.3%) +/+: 45 (68.2%) Unknown: 3 (4.5%)

Table 4.5 Genotypes of Prkd1Em1(IMPC)Wtsi live born mice.

The expected ratios depending on parents are indicated. +/+: $Prkd1^{+/+}$, Em1/+: $Prkd1^{Em1/+}$, Em1/Em1: $Prkd1^{Em1/Em1}$.

Assuming homozygous lethality, the ratios are not far off expected from the $Prkd1^{Em1/+}$x $Prkd1^{Em1/+}$ matings. There are more $Prkd1^{+/+}$ than expected from the $Prkd1^{+/+}$ x $Prkd1^{Em1/+}$crosses, but these unexpected ratios may just be a result of relatively small sample sizes.

4.5.3 Forelimb Staging of *Prkd1*^{Em1(IMPC)Wtsi} Mouse Embryos

The forelimb was used to stage all embryos that were harvested at E15.5 (table 4.6). This was to ensure that they had all reached a developmental stage where heart development was complete.

An example of staging of an embryo I collected is shown below in figure 4.11.
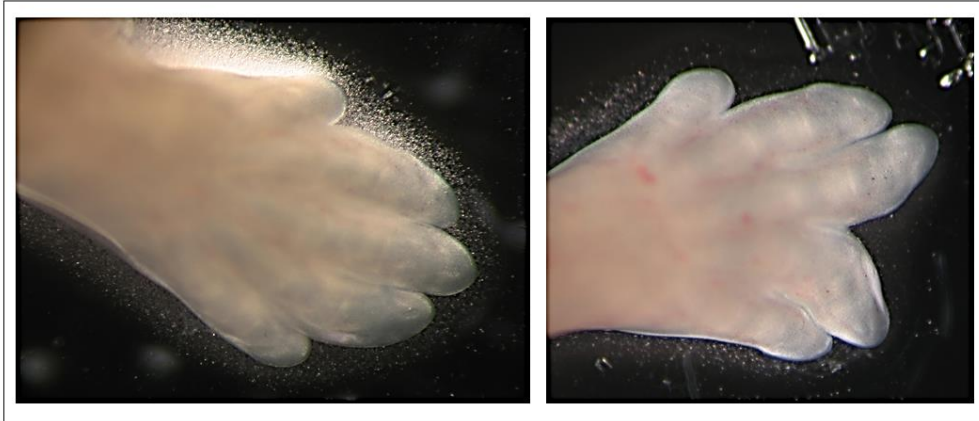
Figure 4.11 Forelimb staging.

WT E15.5 mouse paw. In the right image the digits have been spread to illustrate the interdigital web. This embryo would be stage S22.

Gestation	Genotype	Litter	S22-	S22	S22+	S23-	S23	S23+	Average Stage (Mode)
E15.5	*Prkd1*^{+/+}	1	-	-	-	-	3	-	S23
		2	-	-	-	-	-	1	S23+
		3	-	-	-	-	-	2	S23+
		4	-	-	-	-	-	4	S232+
		5	-	-	-	-	-	2	S23+
		6	-	-	-	-	-	-	-
		7	-	-	-	-	-	3	S23+
E15.5	*Prkd1*^{Em1/+}	1	-	-	-	1	3	-	S23
		2	-	-	-	-	-	3	S23+
		3	-		-	-	-	5	S23+
		4	-	-	-	-	-	3	S23+
		5	-	-	-	-	-	5	S23+
		6	-	-	-	-	-	1	S23+
		7	-	-	-	-	-	4	S23+
E15.5	*Prkd1*^{Em1/Em1}	1	-	-	-	-	-	-	NA
		2	-	-	-	-	-	-	NA
		3	-	-	-	-	-	-	NA
		4	-	-		-	-	-	NA
		5	-	-	-	-	-	-	NA
		6	-	-	-	-	-	-	NA
		7	-	-	-	-	-	-	NA

Table 4.6 Forelimb stage of all *Prkd1*^{Em1(IMPC)Wtsi} mouse embryos.

Forelimb stage is indicated for all the mouse embryos collected at E15.5, by genotype.

The results indicate that the majority of embryos were at S23+ and therefore heart development should be mainly complete. The developmental stage seemed to be comparable within each litter. The only litter showing any variability was litter 1. The numbers are too small to identify any difference in development of the *Prkd1*^{Em1/+} compared to the *Prkd1*^{+/+} group.

4.5.4 Crown Rump Length Measurements of _Prkd1_[Em1(IMPC)Wtsi] Mouse Embryos

Crown rump length (CRL) was measured in all embryos collected at E15.5. The results are shown in table 4.7.

Gestation	Genotype	Litter	CRL (mm)	Average CRL (mm) with stage
E15.5	_Prkd1_[+/+]	1	-	-
		2	15	15 at S23+
		3	-	-
		4	15, 15, 14	14.7 at S23+
		5	16, 17	16.5 at S23+
		6	-	-
		7	-	-
E15.5	_Prkd1_[Em1/+]	1	-	-
		2	15, 15, 16	15.3 at S23+
		3	14, 14, 15, 15, 15	16.6 at S 23+
		4	14, 15, 16	15 at S23+
		5	16, 16, 16, 17, 17	16.4 at S23+
		6	15	15 at S23+
		7	-	-
E15.5	_Prkd1_[Em1/Em1]	1	-	-
		2	-	-
		3	-	-
		4	-	-
		5	-	-
		6	-	-
		7	-	-

Table 4.7 Crown rump length of all E15.5 _Prkd1_[Em1(IMPC)Wtsi] embryos.

The majority of embryos have a CRL between 15-16.5mm. Two tailed T test assuming unequal variances suggests no significant difference (p=0.97). I suspected litter 5 was a bit further on with development than the others as I

noted that the ribcages were a little tougher during dissection. This could explain higher CRL in this litter.

4.5.4 External Phenotype of *Prkd1*[Em1(IMPC)Wtsi] Mouse E15.5 embryos

Phenotypes were evaluated after the forelimb stage of the embryo had been ascertained. Normal features such as herniation of the gut into the base of the umbilical cord were checked, along with development of key structures such as ears and limbs. The embryos had uncovered eyes without eyelids, and the external ear could be visualised. Digits were parallel and nails were visible. The liver is large and easily removed for use in DNA extraction for genotyping.

The results are shown in table 4.8. Phenotyping was carried out before the genotype was known.

Gestation	Genotype	Alive/Deceased	Phenotype
E9.5	PRKD1+/+ (n=1)	Alive: 1	Normal
		Deceased: 0	NA
	Prkd1Em1/+ (n = 4)	Alive: 4	All normal
		Deceased: 0	NA
	Prkd1Em1/Em1 (n=0)	NA	NA
	Unknown (n=5)	Not seen	Not seen
E10.5	PRKD1+/+ (n=6)	Alive: 5	All normal
		Deceased: 1	Embryo reabsorbed, necrotic tissue
	Prkd1Em1/+ (n = 6)	Alive: 4	2 normal 2 slightly delayed, looks like E9.5

		Deceased: 2	Both looked like recent deaths in last day
	Prkd1Em1/Em1 (n=0)	NA	NA
	Unknown (n=6)	Alive: 1	Looks delayed, approx. E9-9.5
		Deceased: 5	All necrotic tissue, small placenta and uterine sac
E11.5	PRKD1+/+ (n=1)	Alive:1	Normal
		Deceased: 0	NA
	Prkd1Em1/+ (n = 8)	Alive: 8	6 normal appearing 1 very small but otherwise normal appearance 1 with small body and tail relative to head
		Deceased: 0	NA
	Prkd1Em1/Em1 (n= 0)	NA	NA
	Unknown (n=1)	Alive: 0	NA
		Deceased: 1	reabsorbed embryo, necrotic tissue
E12.5	PRKD1+/+ (n=0)	NA	NA
	Prkd1Em1/+ (n =8)	Alive:5	Normal
		Deceased:3	All died within last 48 hours
	Prkd1Em1/Em1 (n=0)	NA	NA
	Unknown (n=1)	Alive: 0	NA
		Deceased:1	Necrotic tissue, small placenta and uterine sac
E15.5	PRKD1+/+ (n=23)	Alive: 16	Normal 14 1 with haemorrhage on lung surface

			1 thin, small, cryptophthalmos left, mouth wide open
		Deceased:7	Normal 3 Reabsorbed 4
	Prkd1Em1/+ (n = 28)	Alive: 21	Normal 20 Slightly short snout 1
		Deceased:7	Normal 4 Reabsorbed 3
	Prkd1Em1/Em1 (n=0)	NA	NA
	Unknown (n=11)	Alive: 0	
		Deceased:11	Reabsorbed 11

Table 4.8 Phenotypic analysis of all E15.5 *Prkd1*^{Em1(IMPC)Wtsi}.

Phenotypes based on external inspection at the time of collection.

Abnormalities were seen across both *Prkd1*[+/+] and *Prkd1*[Em1/+] embryos. In the *Prkd1*[+/+] embryos, there is evidence of early demise and abnormal phenotypes were seen in two at E15.5. Cryptophthalmos was seen in one individual; however, eye abnormalities are known to occur in the C57BL6 mice (see Appendix C).

Prkd1[Em1/+] mice appeared to show demise from the early stages, prior to E10.5 through to later stages closer to E15.5. A number of *Prkd1*[Em1/+] embryos were felt to have died at E10.5-11.5 in the E12.5 harvest. It is difficult to stage them very accurately when there is some element of decomposition and maceration. Although the vast majority were grossly normal, abnormalities seen included a shorter snout, general smaller size, possible IUGR phenotype with relatively preserved head size compared to the body.

4.5.5 Phenotype of Adult *Prkd1*^{Em1(IMPC)Wtsi} Mice

126 mice had only general assessment for gross abnormalities as part of routine care in the BSU, and did not have detailed internal and external examinations. No abnormal phenotypes were identified.

A few *Prkd1*^{Em1/+} mice were weighed and their length (crown to anus) measured at 3 months of age (table 4.9). Data available on Charles River and Jackson Laboratory websites suggest the weights of mice are 25g in a WT male mouse and 20g in a WT female. These mice are slightly heavier. However we know that this strain is prone to weight gain and there may have been differences in diet influencing this. I haven't been able to find any data for length of WT mice of the same strain for comparison.

Genotype	Sex	Average weight (g)
Prkd1^{Em1/+}	M (n=5)	26.8
Prkd1^{Em1/+}	F (n=2)	26.3
		Length (mm)
Prkd1^{Em1/+}	M (n=5)	73.6
Prkd1^{Em1/+}	F (n=1)	68

Table 4.9 Weights and lengths of 7 Prkd1Em1+/- mice at three months old.

Of those who had detailed internal and external phenotyping, only a single mouse was identified as having an abnormality (Table 4.10).

Gestation	Genotype		Phenotype
Adult	*Prkd1*^{+/+}	n = 2	All normal
	Prkd1^{Em1/+}	n = 22	1 had absent vibrissae
	Prkd1^{Em1/Em1}	n = 0	NA
	Unknown n = 2		All normal

Table 4.10 Adult Prkd1Em1(IMPC)Wtsi mice phenotypes.

Phenotypic analysis of 26 adult *Prkd1*^{Em1(IMPC)Wtsi} mice by external inspection and internal examination.

Overall, there is no clear abnormal external phenotype in surviving *Prkd1*^{Em1/+} adults compared with *Prkd1*^{+/+} mice.

4.5.6 HREM cardiac assessment of *Prkd1*^{Em1(IMPC)Wtsi} Mouse mice

E15.5 embryos were collected and processed for cardiac phenotyping using HREM. 18 hearts were analysed (*Prkd1*^{+/+}=7, *Prkd1*^{Em1/+}= 11). No *Prkd1*^{Em1/Em1} embryos were identified. Four hearts were excluded from analysis due to damage, leaving 14 for review (2 *Prkd1*^{+/+} and 2 *Prkd1*^{Em1/+}).

All 5 *Prkd1*^{+/+} hearts were normal. In the normal hearts there were a number of consistent features at E15.5 similar to those described by Geyer et al. [224]. This includes the TV sitting slightly inferiorly to the MV and a groove between the ventricles externally at the apex. The PV is superior and to the right of the AV. The pulmonary arteries arise after a short pulmonary trunk. There were no hearts with an immediate bifurcation and no common pulmonary artery, as described by Geyer et al. [224]. The ductus lumen was comparable to the feeding arteries with a normal insertion point. The RV was larger than the LV, with a thinner compact myocardium at the insertion point of the IVS to the anterior RV free wall. Examples of normal features in both *Prkd1*^{+/+} and *Prkd1*^{Em1/+} mice, are shown in figure 4.12- 4.15. Results are summarised in table 4.11 after the figures.

External view of anterior surface of the heart. The pulmonary trunk crosses in front of the aorta and the apex is notched/bifid.

Relative collapse of the right compared to the left atrium
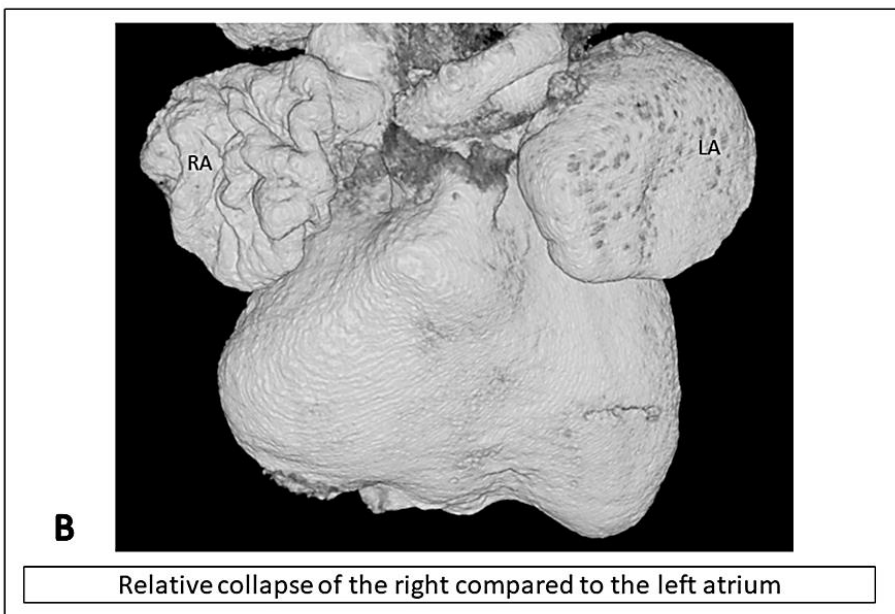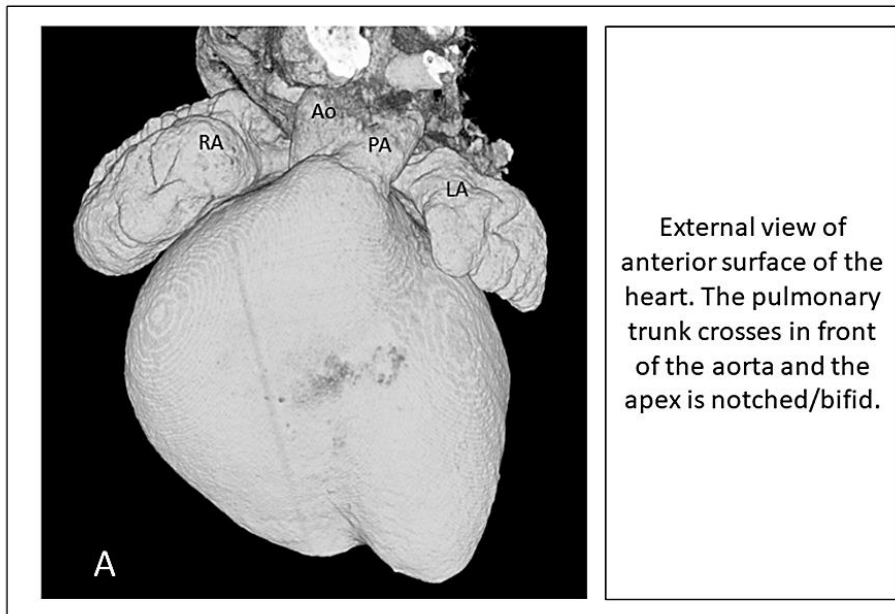
Figure 4.12 3D reconstruction of the external surface of the E15.5 mouse heart, viewed from the front.
A: The two atria are easily determined with their appendages. The pulmonary trunk arises anteriorly to the aorta and passes posterosuperiorly and to the left. At E15.5 the apex has a bifid appearance due to a central groove or notch.
B: The right atrium appears collapsed compared to the left. This is most likely due to the easier emptying and lysis of blood from the right compared to the left atrium. This needs to be taken into account when evaluating the trabeculae in the atria and their appendages.
Ao: Aorta, LA: left atrium, PA: Pulmonary artery, RA: right atrium
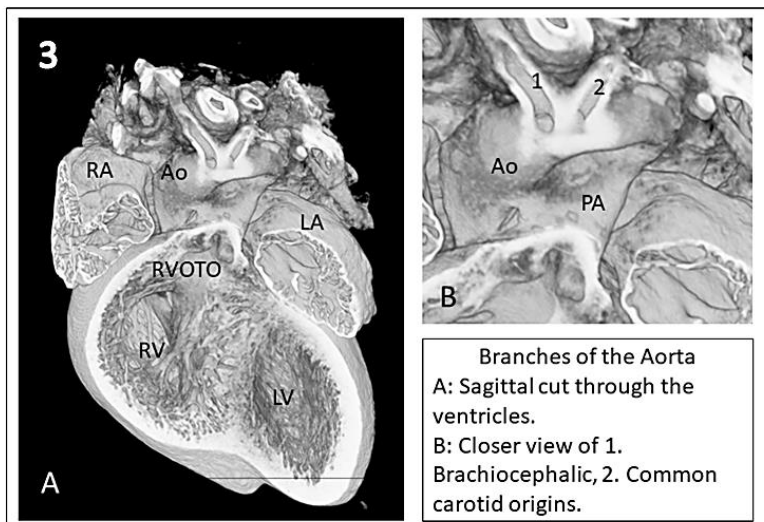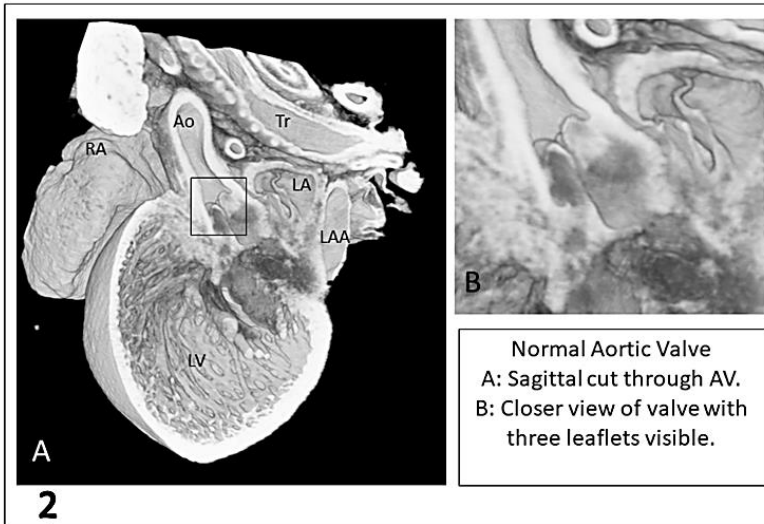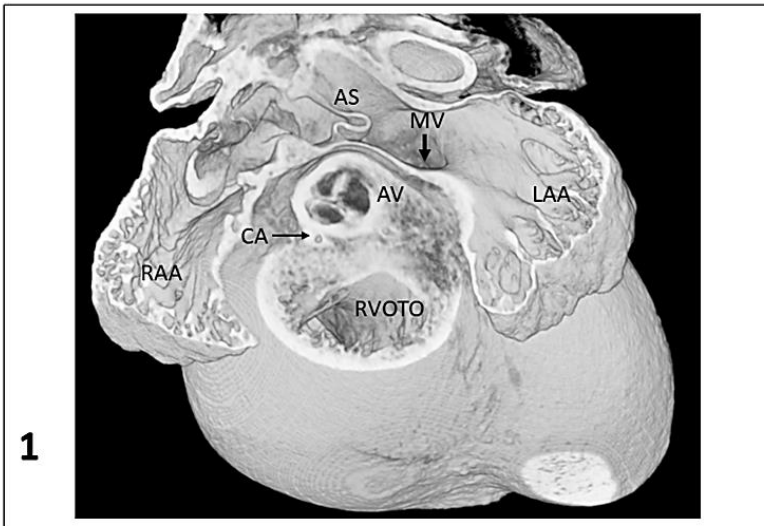
Figure 4.13 3D reconstruction of the E15.5 mouse heart, the aorta and aortic valve

1. **The atria and aortic valve.** Axial cut through the heart. The right and left atrial appendages are shown, demonstrating normal trabeculation. The atrial septum below the level of the foramen ovale is shown. The truncus is cut through at the level of the aortic valve. This is usually sited inferiorly and to the right of the pulmonary valve. The right coronary artery is visible as indicated by the arrow and the site of the mitral valve is also shown.

2. **A. Lateral View of the Aortic Valve.** Trileaflet aortic valve seen from the left side of the heart in sagittal section. The ascending aorta is visible, along with most of the cavity of the left ventricle with normal trabeculation. **B. Enlarged image of aortic valve.** The three valve leaflets are visible.

3. **A. Coronal cut through the heart.** The interventricular septum is visible and the extension of the RVOTO from the RV. **B. Closer image of the first two branches of the aorta.**

AS: atrial septum Ao: Aorta, AV: aortic valve, CA: coronary artery, LA: left atrium LAA: left atrial appendage, LV: left ventricle, PA: Pulmonary artery, MV: mitral valve, RA: right atrium RAA: right atrial appendage, RVOTO: Right ventricular outflow tract, Tr: trachea

Normal Aortic Valve
A: Sagittal cut through AV.
B: Closer view of valve with three leaflets visible.

Branches of the Aorta
A: Sagittal cut through the ventricles.
B: Closer view of 1. Brachiocephalic, 2. Common carotid origins.

Figure 4.14 3D reconstruction of the E15.5 mouse heart. The pulmonary artery and ductus arteriosus.

1. Bifurcation of the Pulmonary Artery. The pulmonary artery travels supero-posteriorly and to the left before bifurcation.

2. Ductus Arteriosus. A: The RVOTO is shown arising from the RV, on this oblique cut through the heart. The leaflets of the pulmonary valve are visible.

B: Closer image of the pulmonary artery and the ductus arteriosus. This is usually of a similar calibre diameter to the PA and Ao. The bifurcation into right and left PA is not seen on this image but the LPA is visible.

Ao: Aorta, BC: Brachiocephalic artery, DA: ductus arteriosus, IVS: interventricular septum, LA: Left atrium and appendage, LPA: Left pulmonary artery, PA: pulmonary artery, PV: pulmonary valve, RPA: right pulmonary artery, , RA: right atrium R SVC: Right superior vena cava, RV: right ventricle, RVOTO: right ventricular outflow tract
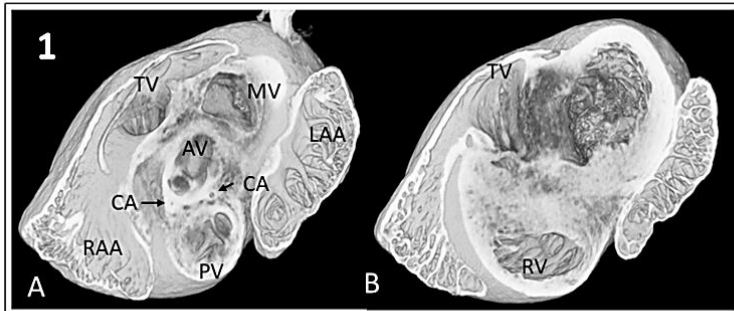
The four valves
A: Axial cut superior to the TV and MV PV and AV with coronary arteries seen and MV
B more inferior cut

Axial cut through atrioventricular connections
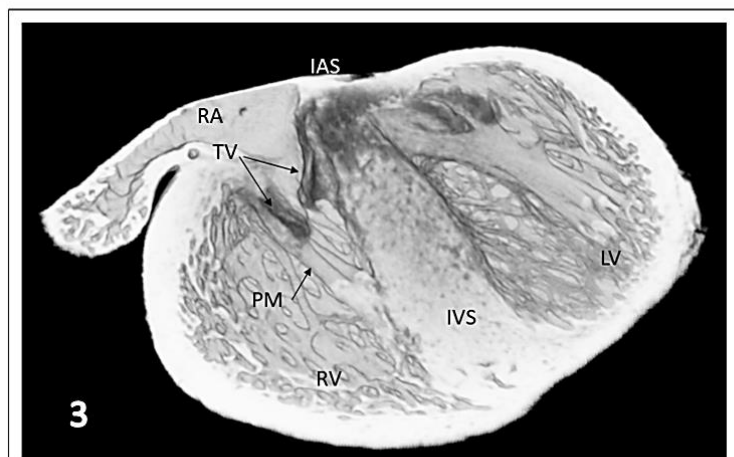
Tricuspid Valve and Papillary muscles

Figure 4.15 3D reconstruction of the E15.5 mouse heart, the four valves.

1. **The four valves.** Axial cut through the heart. A. The three leaflets of the AV and PV are visible. The MV is demonstrated. The TV is always placed slightly inferior to the MV, and only a section of it is visible here. Both left and right coronary arteries are visible (arrows) B. Axial cut through the heart inferior to image A, showing the TV in more detail.

2. **The atrioventricular Valves.** Axial cut through the heart. A more detailed image of the mitral valve.

3. **The tricuspid valve and papillary muscles.** Coronal cut through the heart. The Atrial septum is visible at the top of the image with a cut through the interventricular septum. The leaflets of the TV appear darker (arrows) and the attached papillary muscles are visible (arrow).

AV: aortic valve, CA: coronary artery (arrows), IAS: Interatrial septum, IVS: Interventricular septum, LAA: left atrial appendage, LV: Left ventricle, MV: mitral valve, PA: Pulmonary artery, PM: papillary muscle, PV: pulmonary valve , RA: right atrium, RAA: right atrial appendage, RV: Right ventricle, TV: Tricuspid valve

Genotype	Prkd1^{+/+}	Prkd1^{Em1/+}	Prkd1^{Em1/Em1}
Normal heart	5	8	0
Abnormal heart	0	1 (IAA)	0
Total	5	9	0

Table 4.11 Summary of phenotypes seen in the Prkd1Em1(IMPC)Wtsi E15.5 embryo hearts processed using HREM.

A single *Prkd1*^{Em1/+} heart showed an abnormality, which was an interrupted aortic arch (IAA) (type B). This is shown in figure 4.16. There was no associated VSD or aortopulmonary window. I could not confirm whether the AV in one of the other *Prkd1*Em1+/- mice was definitely tricuspid due to under lysis, but there were no other abnormalities in this heart.

Genotype	$Prkd1^{+/+}$	$Prkd1^{Em1/+}$	$Prkd1^{Em1/Em1}$
Normal heart	5	8	0
Abnormal heart	0	1 (IAA)	0
Total	5	9	0

Table 4.11 Summary of phenotypes seen in the Prkd1Em1(IMPC)Wtsi E15.5 embryo hearts processed using HREM.

A single $Prkd1^{Em1/+}$ heart showed an abnormality, which was an interrupted aortic arch (IAA) (type B). This is shown in figure 4.16. There was no associated VSD or aortopulmonary window. I could not confirm whether the AV in one of the other *Prkd1*Em1+/- mice was definitely tricuspid due to under lysis, but there were no other abnormalities in this heart.

Figure 4.16. *Prkd1*Em1/+ heart 8150C, collected at E15.5 and analysed by HREM. Raw axial images showing an interrupted aortic arch (IAA), type B. A: Axial slice showing the bifurcation of the pulmonary trunk and the Aorta. B: Axial slice showing the three branches of the aorta, the site of the IAA and the descending aorta. The interruption occurs between the left common carotid and subclavian arteries. C:Axial cut inferior to image B showing the three branches of the aortic arch and the site of the interruption. D: Axial cut showing the ductus at the site where it joins the aorta after the interruption. The aortic arch is also visible.

Ao: Aorta, Ao.A: Aortic arch, BCA: brachiocephalic artery, DA: ductus arteriosus, DAo: Descending aorta, IAA: site of interrupted aortic arch, LCC: left common carotid artery, LPA: Left pulmonary artery, LSC: left subclavian artery, PA: Pulmonary artery, RA: Right atrium, RPA: Right pulmonary artery

The disruption of the aorta seen in the *Prkd1*[Em1/+] embryo with an IAA (Figure 4.16), was sited between the LCC and LSC arteries. The ductus arteriosus was continuous with the descending aorta. IAA is a very rare type of CHD, and is usually seen in combination with a VSD or aortopulmonary window, and abnormalities of the head and neck vessels [777]. Other associated features include BAV, TGA and TA. No other abnormalities were seen in this embryo. It is possible that there is a defect that cannot be seen without dynamic imaging, such as a VSD, and there is a degree of under lysis which makes image interpretation more difficult. However, we should also consider the possibility that the IAA seen is iatrogenic given it is at the top of the tissue specimen where the tissue would be most prone to damage, and because this is a rare type of CHD.

There are three subtypes of IAA depending on the location of the disruption (Figure 4.17) [778-780]. Type A, where the disruption is distal to the left subclavian artery accounts for around 13% of IAA, Type B where the disruption falls between the LCC and the LSC arteries (as described above) accounts for around 84% of IAA, and Type C where the disruption is located between the brachiocephalic artery and the LCC artery is the rarest type (around 3% of IAA).

.
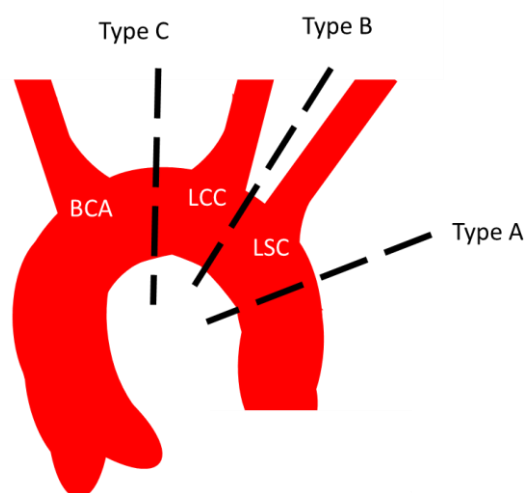
Figure 4.17. Classification of IAA subtypes. Classification is by anatomical location of the disruption.

Type B IAA results from abnormal development of the fourth pharyngeal arch arteries. The left fourth arch artery is required to produce the distal part of the aortic arch, and the right is required for formation of both the brachiocephalic artery, and the proximal part of the right subclavian artery [781]. Type B IAA is associated with 22q11 Deletion Syndrome, and almost 50% of patients with IAA are diagnosed with this condition. *TBX1* is found within this commonly deleted segment, and is required for remodelling and formation of the pharyngeal arch arteries [782]. *Tbx1* null mice develop defects indicative of abnormal forth pharyngeal arch development [783, 784], so it is likely that *TBX1* is responsible for the cardiac phenotype seen as part of this syndrome. There are currently no known interactions between *PRKD1* and *TBX1* to explain why a type B IAA might be seen in this mouse line.

4.6 Results from the *Prkd1*^{Em2(IMPC)Wtsi} Mouse

4.6.1 Confirming genotypes in the *Prkd1*^{Em2(IMPC)Wtsi} Mouse

The WTSI produced a mouse with the equivalent mutation of p.Gly592Arg, detected previously in two of the three individuals with mutations in *PRKD1* reported by Sifrim et al (p.Gly592Arg) [12]. The Variant Effect Predictor (VEP) suggests that the most severe consequence of the base change is a missense mutation, G592R [493]. The mutation is c.G598R in the mouse (Chr 12 50,385,167 G/A -1) and VEP predicts 100% missense mutation with moderate impact based on one transcript ENSMUST00000002765. Both the human and mouse *PRKD1* genes have 19 coding exons and the mutation falls in exon 13 in both species. The alignment around the mutated base is shown below in figure 2.2. The mutation maps to the proximal section of the protein kinase domain.
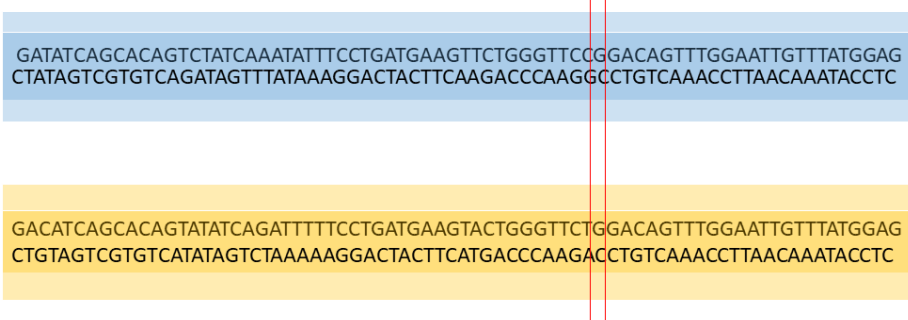
Figure 4.8 Alignment of mouse and human sequence of PRKD1 exon 13 mutation site.

Mouse sequence is in blue, and the human sequence is in yellow. Base affected by mutation is highlighted in red box. Aligned using EMBOSS water alignment tool. https://www.ebi.ac.uk/Tools/psa/emboss_water/nucleotide.html

Example genotype results are shown in figure 4.19 and figure 4.20.
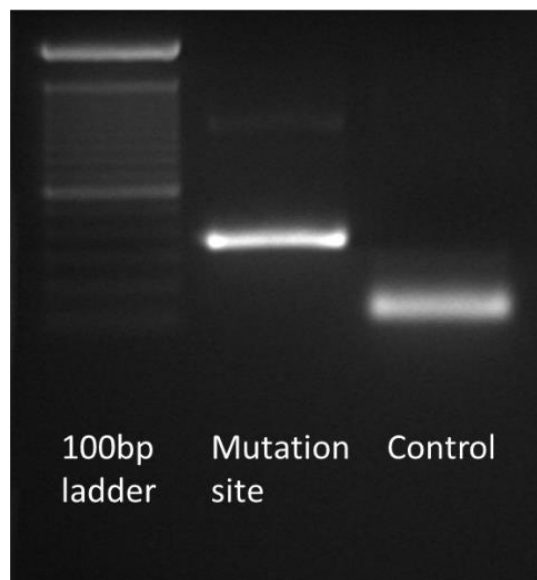
Figure 4.9 PCR for the *Prkd1*[Em2(IMPC)Wtsi] mouse.

A 361bp product is produced. The PCR product was cleaned up and sent for sequencing.
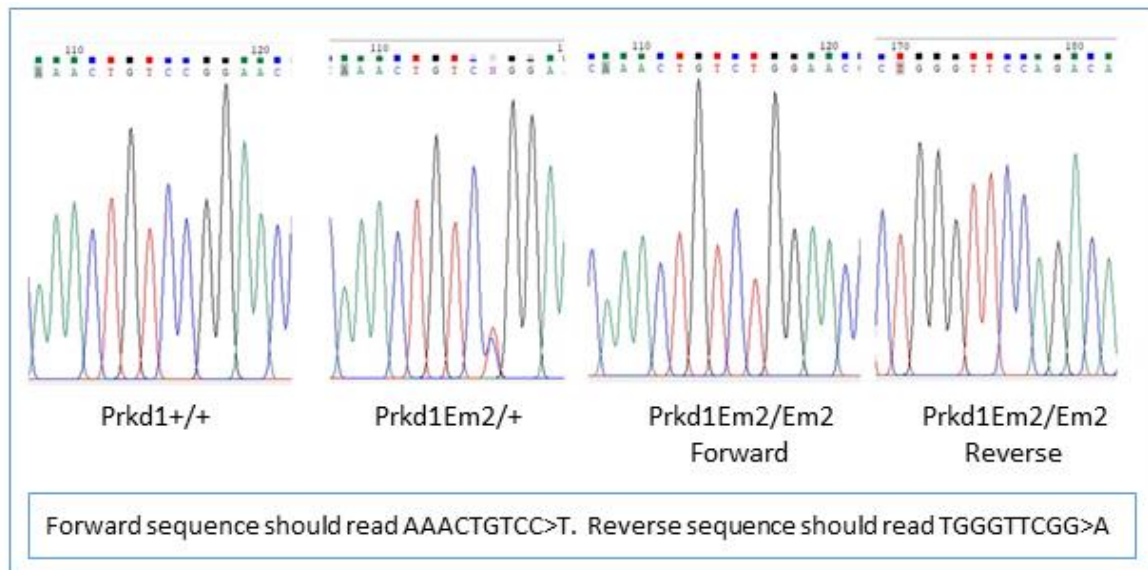
Figure 4.20 Sequencing results for the three genotypes seen in
Prkd1[Em2(IMPC)Wtsi] mice.

4.6.2 Ratios of Genotypes Produced in the *Prkd1*[Em2(IMPC)Wtsi] Mouse

Embryos

All embryos collected were genotyped and the numbers compared to the ratios
expected. The results are shown in table 4.12 and figure 4.21.

Gestation	Cross	Expected	Observed
E9.5	Prkd1Em2/+ x Prkd1Em2/+	Em2/Em2: 25% Em2/+: 50% +/+: 25%	Em2/Em2: 1 (11%) Em2/+: 3 (33%) +/+: 2 (22%) Unknown: 3 (33%)
E10.5	Prkd1Em2/+ x Prkd1Em2/+	Em2/Em2: 25% Em2/+ : 50% +/+: 25%	Em2/Em2: 6 (21%) Em2/+: 15 (54%) +/+: 2 (7%) Unknown: 5 (18%)
E11.5	Prkd1Em2/+ x Prkd1Em2/+	Em2/Em2: 25% Em2/+ : 50% +/+: 25%	Em2/Em2: 1 (10%) Em2/+: 5 (50%) +/+: 3 (30%) Unknown: 1 (10%)

E12.5	Prkd1Em2/+ x Prkd1Em2/+	Em2/Em2: 25% Em2/+ : 50% +/+: 25%	Em2/Em2: 4 (21%) Em2/+: 10(52%) +/+: 2 (11%) Unknown: 3 (16%)
E15.5	Prkd1Em2/+ x Prkd1Em2/+	Em2/Em2: 25% Em2/+ : 50% +/+: 25%	Em2/Em2: 4 (8%) Em2/+: 26 (51%) +/+: 13 (25%) Unknown 9 (18%)

Table 4.12 Genotypes and expected ratios found in the embryos of *Prkd1* Em2(IMPC)Wtsi mice.

+/+: *Prkd1$^{+/+}$*, Em2/+: *Prkd1$^{Em2/+}$*, Em2/Em2 : *Prkd1$^{Em2/Em2}$*
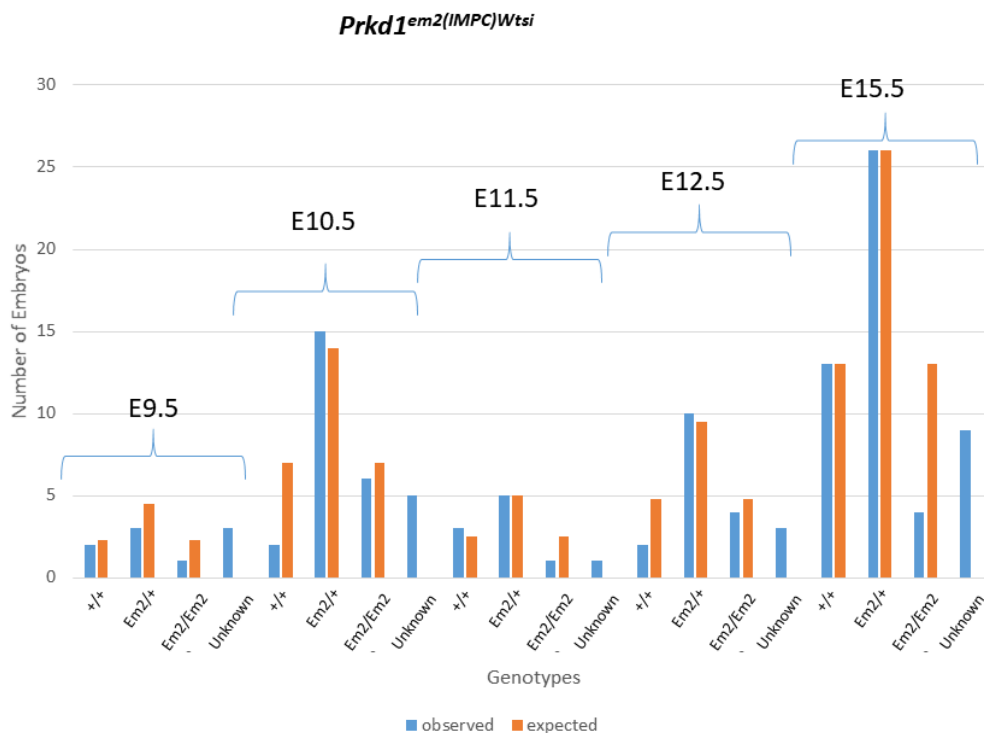
Figure 4.21 Genotypes of *Prkd1*Em2(IMPC)Wtsi embryos.

Graphical representation of numbers of embryos of each genotype identified at each gestation, and compared with the expected numbers. The ratios are fairly well matched with the exception of *Prkd1$^{Em2/Em2}$* embryos.

Prkd1^{Em2/Em2} embryos were found at all stages of embryonic development assessed between E9.5 and E15.5. There were fewer than expected suggesting a degree of lethality. The expected number of *Prkd1^{Em2/+}* embryos were produced, suggesting there is no reduced survival during embryogenesis up to E15.5 of this genotype.

Live born mouse genotypes

The ratios of *Prkd1^{+/+}* and *Prkd1^{Em2/+}* in live born mice are around expected values, suggesting no reduced survival of *Prkd1^{Em2/+}* mice and that they can become healthy adult mice under standard laboratory conditions (table 4.13).

Gestation	Cross	Expected	Observed
Adult	Prkd1Em2/+ x Prkd1Em2/+	Em2/Em2: 25% Em2/+: 50% +/+: 25%	Em2/Em2: 2 (6%) Em2/+: 15 (45%) +/+: 16 (50%) Unknown: 0 (0%)
Adult	Prkd1Em2/+ x +/+	Em2/Em2: 0% Em2/+: 50% +/+: 50%	Em2/Em2: 0 (0%) Em2/+: 39 (39%) +/+: 58 (59%) Unknown: 2 (2%)
Adult	Prkd1Em2/+ x Em2/Em2	Em2/Em2 :50% Em2/+: 50% +/+: 0%	Em2/Em2: 0 (0%) Em2/+: 11 (92%) +/+: 0 (0%) Unknown: 1 (8%)

Table 4.13 Figure 4.21 Genotypes and expected ratios found in the adult *Prkd1^{Em2(IMPC)Wtsi}* mice.

+/+: *Prkd1^{+/+}*, Em2/+: *Prkd1^{Em2+/-}*, Em2/Em2: *Prkd1^{Em2/Em2}*.

Two *Prkd1^{Em2/Em2}* mice were born and survived to adulthood. One was male and one was female. The male proved to be fertile and was used as a stud mouse. In the main homozygosity for this mutation is lethal, with around 1.4% (n=2/144) surviving to adulthood.

4.6.3 Forelimb Staging of Prkd1$^{Em2(IMPC)Wtsi}$ E15.5 Embryos

There was very little variation in the stage of the E15.5 embryos (table 4.14). All reached stage S23- at least, so cardiac development should be mainly complete.

Gestation	Genotype	Litter	S22-	S22	S22+	S23-	S23	S23+	Average Stage (Mode)
E15.5	Prkd1$^{Em2+/+}$	1	-	-	-	1	-	4	S23+
		2	-	-	-	-	-	1	S23+
		3	-	-	-	-	-	3	S23+
		4	-	-	-	-	-	3	S23+
		5	-	-	-	-	-	1	S23+
E15.5	Prkd1$^{Em2+/-}$	1	-	-	-	1	-	2	S23+
		2	-	-	-	-	-	5	S23+
		3	-		-	-	-	5	S23+
		4	-	-	-	-	-	6	S23+
		5	-	-	-	-	-	7	S23+
E15.5	Prkd1$^{Em2/Em2}$	1	-	-	-	-	-	-	-
		2	-	-	-	-	-	1	S23+
		3	-	-	-	-	-	-	-
		4	-	-	-	-	-	1	S23+
		5	-	-	-	-	-	-	-

Table 4.14 Forelimb stage of all E15.5 Prkd1$^{Em2(IMPC)Wtsi}$ embryos and their respective genotypes.

The two homozygotes alive at E15.5 were at the same developmental stage as Prkd1$^{Em2+/+}$ and Prkd1$^{Em2+/-}$ embryos.

4.6.4 Crown Rump Length Measurements of Prkd1$^{\text{Em2(IMPC)Wtsi}}$ E15.5 Embryos

CRL appeared to be comparable across all the embryos examined as shown below in table 4.15.

Gestation	Genotype	Litter	CRL (mm)	Average CRL (mm) with stage
E15.5	*Prkd1*$^{\text{Em2+/+}}$	1	-	-
		2	-	-
		3	-	-
		4	14, 16, 16	15.3 at S23+
		5	15	15 at S23+
E15.5	*Prkd1*$^{\text{Em2+/-}}$	1	-	-
		2	-	-
		3	-	-
		4	14, 14, 15, 15, 15, 15	14.7 at S 23+
		5	9, 14, 14, 15, 15, 15, 16	14 at S23+
E15.5	*Prkd1*$^{\text{Em2/Em2}}$	1	-	-
		2	-	-
		3	-	-
		4	16	16 at S23+
		5	-	-

Table 4.15 CRL (mm) for all *Prkd1*$^{\text{Em2(IMPC)Wtsi}}$ embryos collected at E15.5 and their corresponding genotypes.

Excluding the outlier in the *Prkd1*$^{\text{Em2+/-}}$ embryos (CRL 9mm) brings the mean CRL to 14.75. The single *Prkd1*$^{\text{Em2/Em2}}$ embryo may not be representative. One way anova does not suggest any significant difference.

4.6.5 External Phenotype of Prkd1^{Em2(IMPC)Wtsi} E15.5 embryos

External phenotyping was carried out at the time of embryo collection. The phenotypes are recorded in table 4.16. Phenotyping was carried out before the genotype was known.

Gestation	Genotype	Alive/Deceased	Phenotype
E9.5	PRKD1EM2+/+ (n=2)	Alive: 2	Normal
		Deceased: 0	NA
	Prkd1Em2+/- (n =3)	Alive: 3	Normal
		Deceased: 0	NA
	Prkd1Em2/Em2 (n=1)	Alive: 1 Deceased: 0	Normal
	Unknown (n=3)	No embryo found	NA
E10.5	PRKD1EM2+/+ (n=2)	Alive: 2	Normal
		Deceased: 0	NA
	Prkd1Em2+/- (n=15)	Alive: 14	Normal
		Deceased: 1	NA
	Prkd1Em2/Em2 (n=6)	Alive:6 Deceased: 0	5 normal 1 slight developmental delay and small in size
	Unknown (n=5)	Alive: 0	
		Deceased: 5	All reabsorbed. Small sac, oligohydramnios and discoloured fluid.
E11.5	PRKD1EM2+/+ (n=3)	Alive: 3	Normal
		Deceased: 0	NA

	Prkd1Em2+/- (n =5)	Alive: 5	Normal
		Deceased: 0	NA
	Prkd1Em2/Em2 (n=1)	Alive: 1	Small for gestation
		Deceased: 0	NA
	Unknown (n=1)	Alive: 0	NA
		Deceased: 1	Reabsorbed, necrotic tissue.
E12.5	PRKD1EM2+/+ (n=2)	Alive: 2	Normal
		Deceased: 0	NA
	Prkd1Em2+/- (n =10)	Alive: 10	Normal
		Deceased: 0	NA
	Prkd1Em2/Em2 (n=4)	Alive: 2	Normal
		Deceased: 2	Embryo looks about E11.5, otherwise normal. One looks like death in last 2 days, embryo not overtly necrotic but amniotic fluid is discoloured.
	Unknown (n=3)	Alive: 2	Normal
		Deceased: 1	Necrotic, no embryo identified.
E15.5	PRKD1EM2+/+ (n=13)	Alive: 13	Normal:12 1 with a slightly shorter snout.
		Deceased: 0	NA
	Prkd1Em2+/- (n =26)	Alive: 24	Normal: 23 Splaying of the 3rd and 4th digits in the right upper limb in one.
		Deceased: 2	Normal but small and pale placenta in one.

			Small amount of necrotic tissue only in other.
	Prkd1Em2/Em2 (n=4)	Alive: 2	Normal: 2
		Deceased: 2	1 completely reabsorbed. 1 roughly staged at E10.5, but difficult to assess due to maceration.
	Unknown (n=8)	Alive: 0	
		Deceased: 8	All reabsorbed.

Table 4.16 Phenotype of all *Prkd1*^{Em2(IMPC)Wtsi} embryos collected.

The corresponding genotypes and whether the embryo was alive or deceased at the time of collection is included.

Prkd1$^{+/+}$ embryos were all alive at the time of collection. All were normal with the exception of one who appeared to have a slightly shorter snout. *Prkd1*$^{Em2+/-}$ embryos were alive, with the exception of 3 out of 59 (5%). One had died early in gestation and was necrotic, one had died not long before E10.5 and one had died close to E15.5. This later death was associated with a small pale placenta, which could indicate placental factors contributing to the death.

The majority of the *Prkd1*$^{Em2/Em2}$ embryos appeared normal, although 2 were noted to be slightly smaller in size. *Prkd1*$^{Em2/Em2}$ embryos were found alive at every gestation examined. Death occurred from around E10.5-E12.5, based on an embryo found dead at E12.5 which was somewhat macerated and the amniotic fluid was discoloured. One was completely reabsorbed and may represent an even earlier death. There are a number that we could not genotype due to lack of tissue and these could represent early embryonic death of any genotype, but given the lower than expected numbers of *Prkd1*$^{Em2/Em2}$ embryos, this genotype might represent a significant proportion.

4.6.6 Phenotype of Adult Prkd1^{Em2(IMPC)Wtsi} Mice

No abnormalities were identified in the 188 mice of all genotypes by external assessment during routine care. A single male *Prkd1^{Em2/Em2}* mouse was born and was still alive at the end of my study time. He was externally normal, fertile and produced a number of litters. 20 male *Prkd1^{Em2/+}* and 7 male *Prkd1^{+/+}* mice were collected for MRI studies at 2 months of age. They had detailed external phenotyping. All were normal. No internal phenotyping was carried out as they were required to be intact for the MRI studies.

17 adult mice were examined both externally and internally by dissection. Two *Prkd1^{+/+}* and two *Prkd1^{Em2/+}* mice were dissected for protein studies. Another 8 female *Prkd1^{Em2/+}* mice were phenotyped after being used for embryo harvests. The original four *Prkd1^{Em2/+}* mice received from the Sanger Institute were also examined. Internal and external examination revealed no abnormalities in all cases. A single female *Prkd1^{Em2/Em2}* mouse required schedule 1 killing after being found with a scratch on her cheek. She appeared normal and her heart was explanted for HREM.

4.6.7 HREM cardiac assessment of *Prkd1*^{Em2(IMPC)Wtsi} mice

E15.5 embryos were collected and processed for cardiac phenotyping using HREM (Table 4.18). 21 hearts were analysed (*Prkd1^{+/+}* =5, *Prkd1^{Em2/+}*= 15, *Prkd1^{Em2/Em2}* = 1). 6 hearts were excluded from analysis due to damage, leaving 15 for review (*Prkd1^{+/+}* = 4, *Prkd1^{Em2/+}*= 10 and *Prkd1^{Em2/Em2}* = 1). All hearts appeared normal, and displayed features at E15.5 consistent with those described by Geyer et al. [224] with the exception of one *Prkd1^{Em2/+}* mouse with a muscular VSD (Figure 4.22). Results are summarised in table 4.17.
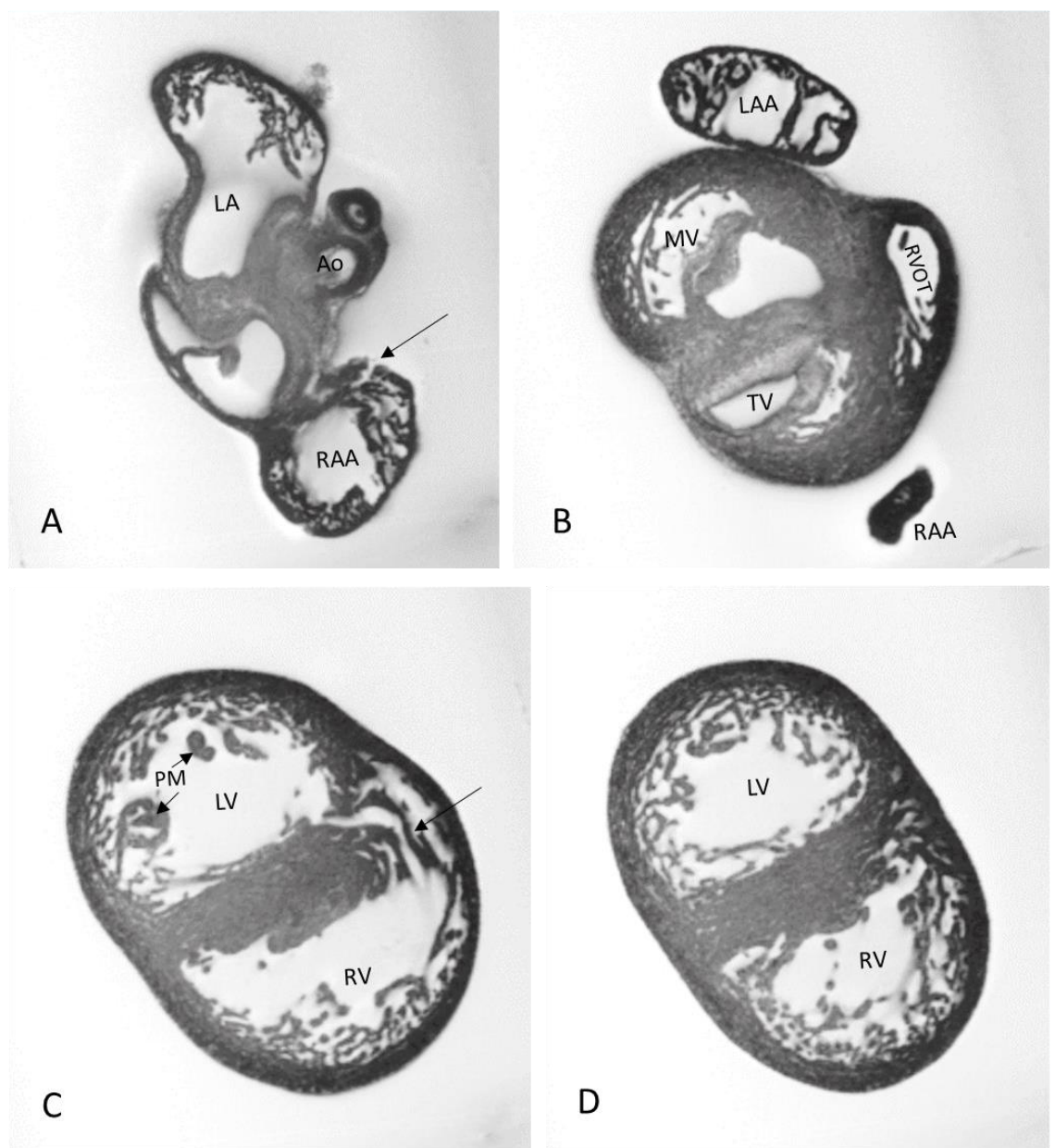
Figure 4.22 Muscular VSD in a Prkd1Em2/+ mouse (8083A) from superior to inferior transverse slices (A-D). HREM raw images.
A: A small section of damage was identified in the anterior wall of the RAA. No other damage was seen in the heart.
B: Transverse section through the heart at the level of the tricuspid and mitral valves.
C: Section through the heart showing the two ventricles at a level of the midpoint of the papillary muscles. A clear connection between the two (arrow) is seen.
D: More inferior section through both ventricles showing an intact IVS and normal morphology of the septum, and its attachment to the anterior wall of the heart.

Ao: Aorta, IVS: Interventricular septum, LA: left atrium, LAA: Left atrial appendage, LV: Left ventricle, MV: Mitral valve, PM: Papillary muscle, RA: Right atrium, RAA: Right atrial appendage , RV: Right ventricle, RVOT: Right ventricular outflow tract, TV: Tricuspid valve.

A small area of damage was identified in the wall of the right atrial appendage of the *Prkd1*^{Em2/+} embryo with the muscular VSD, but nowhere else in the heart. The atria are more prone to damage because of their more fragile and thin wall. I considered whether the defect in the interventricular septum might be mechanical damage. I reviewed the images with Dr B Erhayiem and we think this most likely represents a muscular VSD because:

- There is relatively little damage to the heart. It is not uncommon to see more significant atrial damage than this, with an intact septum.
- Ventricular wall damage, but no septal damage has been observed in other hearts. The free walls of the ventricles are much thinner than the thick muscular septum and are more likely to rupture.
- This appearance would be consistent with a muscular VSD in the septum. It has a very different appearance to those seen in association with AVCD where a very definite gap is seen. This narrower and more winding course is more typical of a muscular VSD.

Isolated VSDs are not uncommon, and muscular VSDs make up around 20% of all VSDs. They may be multiple and often close without intervention. Muscular VSDs frequently close in postnatal life and would be compatible with postnatal survival, so it is possible that this defect may have been present in other apparently healthy surviving *Prkd1*^{Em2/+} mice. VSDs are seen in a number of genetic syndromes including Holt Oram Syndrome, caused by mutations in *TBX5*. *PRKD1* relieves the repression of *TBX5,* mediated by *HDAC4/5* so it is possible that there may be a common mechanism causing VSDs in Holt Oram and with loss of *Prkd1* [785]. In addition, mice with loss of both *Hdac5* and *Hdac9* have lethal VSDs [550] and knockout of *Mef2c* in the second/anterior heart field results in nearly 60% of embryos affected with VSDs, or outflow tract abnormalities in combination with a VSD similar to TOF [786]. The possible mechanisms by which loss of *Prkd1* might result in CHD have been discussed in section 4.3.4, and are also reviewed later in section 4.9.

Genotype	$Prkd1^{+/+}$	$Prkd1^{Em2/+}$	$Prkd1^{Em2/Em2}$
Normal heart	4	9	1
Abnormal heart	0	Muscular VSD	0
Total	4	10	1

Table 4.17 Summary of phenotypes seen in the E15.5 embryo hearts processed using HREM.

A single adult female, and single male $Prkd1^{Em2/Em2}$ mouse were born and apparently both were healthy. The female mouse had to be culled as a result of a scratch. Although HREM is not optimal for adult mouse hearts, I collected the heart and processed it in accordance with suggestions from T Mohun. This included increasing the length of the infiltration stage to a week. Unfortunately the images are suboptimal owing to lack of infiltration of the ventricular septum and insufficient lysis. This is evident as a lack of staining in the LV and septum. However, some useful information can still be gained (Figure 4.23). The atria, atrial connections and atrial septum are normal. The TV, MV and AV can be visualised, but the PV is not seen clearly. The PA branches normally and the aorta appears to have a normal position. The IVS appears intact. The ventricular cavities are small, but appear to have a normal shape and relationship. Unfortunately there is no comparison series for this adult heart, as the remaining adult hearts I processed have not been sectioned. This is because they are also likely to show inadequate lysis and infiltration.
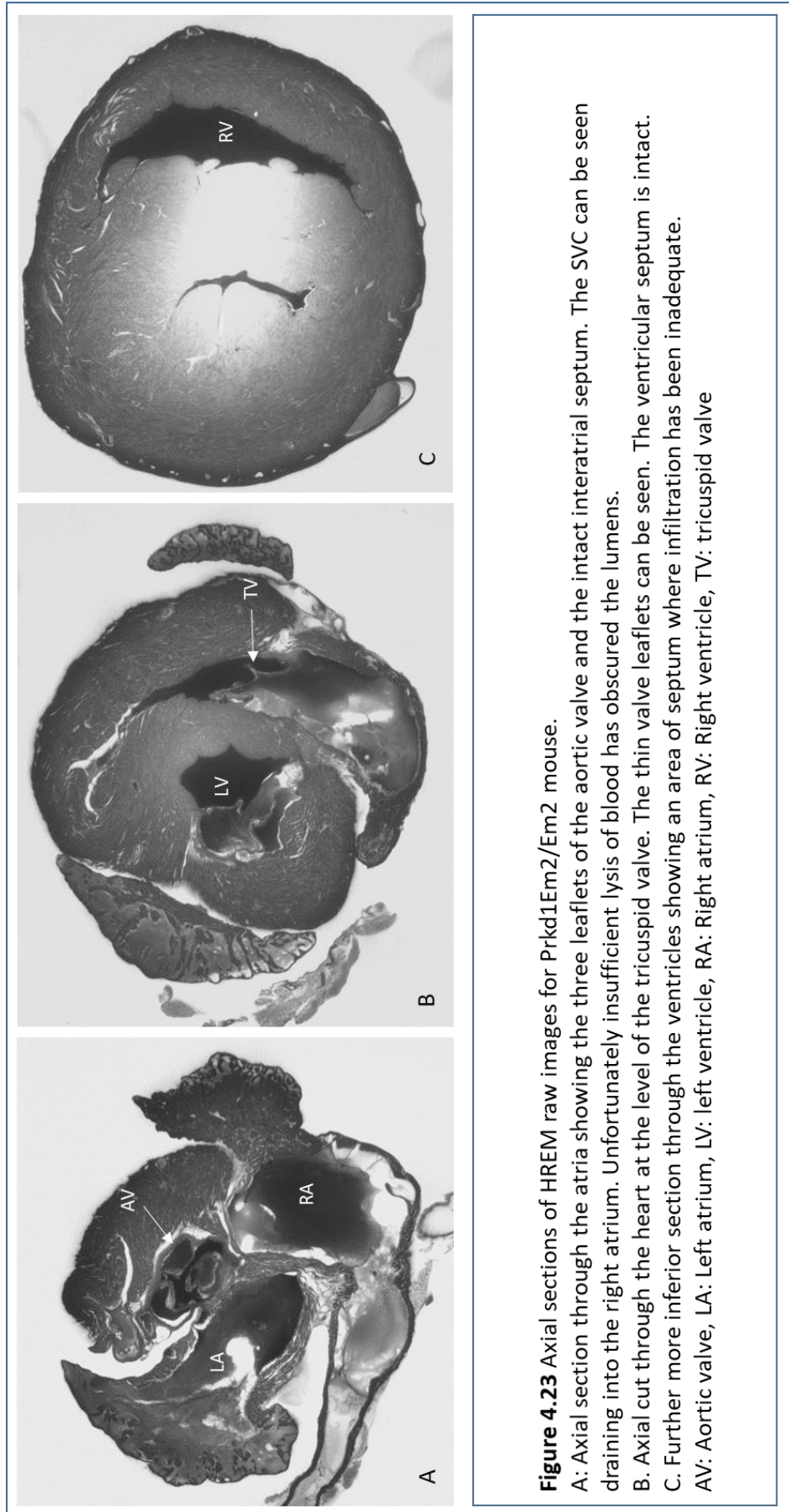
Figure 4.23 Axial sections of HREM raw images for Prkd1Em2/Em2 mouse.

A: Axial section through the atria showing the three leaflets of the aortic valve and the intact interatrial septum. The SVC can be seen draining into the right atrium. Unfortunately insufficient lysis of blood has obscured the lumens.

B. Axial cut through the heart at the level of the tricuspid valve. The thin valve leaflets can be seen. The ventricular septum is intact.

C. Further more inferior section through the ventricles showing an area of septum where infiltration has been inadequate.

AV: Aortic valve, LA: Left atrium, LV: left ventricle, RA: Right atrium, RV: Right ventricle, TV: tricuspid valve

4.7 Comparison of Embryonic Lethality in the *Prkd1*^{Em1(IMPC)Wtsi} and *Prkd1*^{Em2(IMPC)Wtsi} lines

Mouse models reported previously have confirmed that *Prkd1* is required for embryogenesis [684, 732]. Conditional models have shown that absence of *Prkd1* in cardiomyocytes prevents remodelling in response to stress [732]. These results suggest complete embryonic lethality in *Prkd1*^{Em1/Em1} embryos prior to E10.5, which is consistent with previous reports [684, 732]. No *Prkd1*^{Em1/Em1} embryos were identified in collections from E9.5-15.5. In contrast, the *Prkd1*^{Em2/Em2} mice were embryonic lethal with reduced penetrance from a later gestation (E10.5 onwards). Many survived for much longer and died after E15.5. *Prkd1*^{Em2/Em2} mice were live born with an incidence of 1%, and appeared healthy and fertile.

Given *Prkd1* appears to have a wide range of functions, is involved in many different signalling pathways, and shows fairly ubiquitous expression in embryogenesis, it is not surprising that absence results in embryonic lethality. Differing timing of expression and localisation may mitigate any potential functional redundancy within this gene family. *Prkd1* is the dominant *Pkd* isoform expressed in early mouse embryogenesis, with *Prkd2 only* becoming obvious from E14.5, and *Prkd3* being expressed weakly from E10.5 [685]. Embryonic lethality in *Prkd1*^{Em1/Em1} mice here and previous studies confirm that *Prkd1* is important in embryogenesis.

Both previously reported null models (deletion of exons 12-14 and substitution of serine with alanine at the phosphorylation sites) resulted in a 1% incidence of live homozygous null mice [684, 732]. The *Prkd1*Em1 mouse studied here, has a deletion of exon 2. It appears to be universally embryonic lethal in the homozygous null state and no *Prkd1*^{Em1/Em1} mice were born. One explanation for this is that my sample size is too small to detect live born *Prkd1*^{Em1/Em1} mice. Matthews et al. identified 2 *Prkd1*^{SSAA/SSAA} homozygous null mice in 177 births.

157 mice were genotyped and no live *Prkd1*^{Em1/Em1} mice were found. Recent communication with Professor Brook's lab reveals that this is still the case, and no *Prkd1*^{Em1/Em1} mice have survived despite continued breeding.

If it is not related to sample size, then it is possible that there is a different effect from deletion of exon 2, which is more severe. The protein product from the mouse without exons 12-14 was degraded [732]. It is not clear what happens to the protein product in the *Prkd1*Em1 mouse and work is ongoing to determine if a truncated protein is produced, or if nonsense mediated decay is occurring. This may help clarify if the different deletions have different effects. The consequences of a deletion preceding and including part of the first CRD is unclear, but potentially significant given that the second CRD is required for normal movement of PRKD1.

Functional studies might help clarify the effect of the exon 2 deletion. They are also necessary to assess why the protein kinase missense mutation in the *Prkd1*Em2 mouse line results in lethality at a later gestation and occasional survival of homozygous knockout mice.

A number of empty decidua were detected in both *Prkd1*Em1 and Em2 cohorts, with only necrotic tissue present. This suggests post-implantation death (E5.5 onwards) or possibly peri-implantation death. It is also possible for embryos compromised at an earlier stage (E4.5-5.5, peri-implantation) to create an identifiable implantation site, as long as they have been able to shed the zona pellucida and form a trophoblast layer [787]. In most of these cases, sufficient embryonic tissue could not be collected which meant that the genotype was unknown. Given that deaths occur in 5-15% of WT embryos at an early gestation, we cannot assume that peri or post implantation deaths were related to alterations in *Prkd1* [787-789]. No pre-implantation studies were carried out so we can only conclude that some deaths may occur peri or post implantation in both *Prkd1*Em1 and Em2 mice.

The cause of embryonic lethality has not been determined. Collection of hearts for HREM requires swift explantation, and detailed examination of the embryo at the time of dissection is not possible. Partial destruction of the embryo also makes phenotyping afterwards unreliable. The widespread expression of Prkd1 makes it difficult to assign a cause of lethality as any number of organ systems could be involved.

The earlier deaths in the *Prkd1*Em1 mice, suggest that placental factors or cardiovascular malformations may play a role. Causes of death within the first half of gestation include placental failure, cardiovascular insufficiency, abnormal haematopoesis and abnormal angiogenesis. Placental insufficiency usually results in death before E10-E13 and is usually seen in combination with IUGR [790]. Small pale placentas were seen in a few embryos, mainly in combination with an empty decidua. A single *Prkd1*$^{Em1/+}$ embryo appeared to have a small body with relatively preserved head circumference, features that are seen in IUGR. *PRKD1* is expressed in the human placenta [791], so it is possible that placental insufficiency contributed, and further studies should include histology of the placenta. *VEGF* and other stimuli increase the activity of *PRKD1* in endothelial and vascular smooth muscle cells [745-748], therefore angiogenesis may also be a contributing factor in their early embryonic demise.

Abnormalities of the brain, kidney, liver and lung can result in survival in utero and death in the peri or postnatal period [790]. Therefore these are less likely to be the cause of lethality in the *Prkd1* mice. Deaths after E9.5 can result from defects that initially occurred at an earlier stage, and in the majority of cases embryos dying after E12 have defects that occurred prior to this [790]. Therefore, it is possible that defects occur in the same organs at the same early time point in both *Prkd1*Em1 and Em2 lines, but are more severe in *Prkd1*Em1 mice resulting in earlier death.

The variable time of death in the *Prkd1*Em2 line embryos suggests that there is reduced penetrance and variable expressivity for any related defects. For

example, an early death might be a result of a severe cardiac defect, later death related to a less compromised heart, and survival to term seen in those with a normal heart or very mild defect. Common causes of embryonic death after E12 include haemopoetic defects [792] and abnormalities in the heart related to defective septation or cardiac conduction problems [217]. HREM only identified a single mouse with a septal defect (*Prkd1*[Em2/+]) and one embryo with IAA (*Prkd1*[Em1/+]). Conduction abnormalities could not be assessed, but could be important given the role of *Prkd1* in maintaining normal contractility in the heart [689, 742, 774].

To determine the cause of death, whole embryo HREM should be carried out at multiple gestations. Additional examination for placental abnormalities and for signs of cardiac insufficiency such as pooling of blood in the heart and hepatomegaly should be included [787]. Ideally ECGs, echos and MRI should also be carried out to assess function and conduction.

4.8 Comparison of Extra cardiac Phenotypic Abnormalities in the *Prkd1*[Em1(IMPC)Wtsi] and *Prkd1*[Em2(IMPC)Wtsi] lines

Few abnormalities were noted in the embryos from either *Prkd1* line, although as discussed in the previous section, time for examination was limited. A few *Prkd1*[Em1/+] embryos were smaller in size (not significant), and one had a shorter snout. This might be significant given Li et al. reported smaller size and craniofacial abnormalities in a conditional null postnatal mice [758], however a slightly shorted snout was also seen in a WT mouse (table 4.8) and the sample size is too small to determine if short snouts are a feature of loss of *Prkd1* in these mouse lines.

No digital abnormalities were noted except for unilateral increased splaying of the third and fourth digits of the forelimb in a single *Prkd1*[Em2/+] embryo at E15.5.

This could be in line with the digital abnormalities reported in CHD and ectodermal defects [12]. It is interesting that the abnormality was seen in a *Prkd1*^{Em2/+} embryo and not a *Prkd1*^{Em2/Em2} embryo, and given that limb abnormalities were only seen once I cannot be sure that is related to loss of *Prkd1*.

In addition to *Prkd1, Prkd3* is expressed in the limb bud, cartilage and axial skeleton. *Prkd3* knockout mice have reduced cortical thickness in the femur and reduced trabecular bone in the vertebrae, so it is likely that *Prkd3* also contributes to maintaining a normal skeletal phenotype and may mitigate some of the effects of loss of *Prkd1* [546]. Bony imaging at this stage would be difficult due to an immature cartilaginous skeleton, although MRI may helpful in detect abnormalities at this stage. It is not known however, what the underlying bony malformations in the individuals with CHD and Ectodermal Defects are, which makes it challenging to decide what further investigations might be helpful in a mouse model.

A single adult *Prkd1*^{Em1/+} mouse had no whiskers. There was no other evidence of barbering in this mouse or the cage. This might be significant given the ectodermal features seen in humans, but a single report is not enough to say anything conclusive and larger numbers of mice will need to be examined. Whiskers are visible at E17 (Theiler stage 25), so could not easily be seen at the embryonic stages I was examining. *Prkd2* is expressed at site of whisker development at E14.5 so development is unlikely to be solely dependent on *Prkd1* [546, 685].

Two *Prkd1*^{Em2/Em2} mice were born and survived to adult age. Both were apparently normal and the male mouse was fertile. The male mouse was still breeding successfully after 7 months. The female mouse had to be culled at five months of age due to a scratch, but appeared normal on external examination and internal dissection. This suggests reduced penetrance and variable expressivity of the protein kinase missense mutation. Further functional work is

required to understand the fate of the abnormal mouse protein and its activity, to understand why it can is associated with apparently normal survival in homozygous states.

Previously published mouse expression data suggests that *Prkd1* is expressed in the heart, brain, limb buds and other organ systems during embryogenesis. Therefore these organs are of interest. Further work should consider imaging or histology of the brain and imaging of the limbs and skeleton. Postnatal examinations could include behavioural and intelligence tests in addition.

4.9 Significance of Cardiovascular Abnormalities identified in the *Prkd1*^{Em1(IMPC)Wtsi} and *Prkd1*^{Em2(IMPC)Wtsi} lines

Only two embryos were identified with cardiac defects (n=2/29). A single *Prkd1*^{Em1/+} E15.5 embryo out of the 9 examined, had an interrupted aortic arch (IAA). A muscular VSD was seen in one of the *Prkd1*^{Em2/+} mice. The number of embryos analysed is small so results should be interpreted with caution. Further *Prkd1* embryos from both lines need to be examined to determine if these are true findings, and if they are associated with loss of *Prkd1*. It would also be beneficial to carry out whole embryo HREM as this prevents any disruption of the outflow tract and great vessels. HREM at earlier gestations should also be carried out. We may see CHD in mice at earlier gestations, which would have died by E15.5 and be unsuitable for analysis at this later stage.

In the developing heart, there are a number of structures that come together to form the AV valves, the semilunar valves and the septae. Correct formation of all of these prevents the defects seen in the *Prkd1* mice and individuals with CHD and Ectodermal Defects. Defective endocardial cushion formation could result in an AVSD and potentially abnormalities of the pulmonary valve. The muscular VSD implies a problem in the developing ventricular septum. Malalignment of

this septum with the infundibular septum could lead to obstruction of the LVOT and subsequent IAA. Pulmonary stenosis could also result from altered flow. It is possible therefore that the abnormalities we see are linked to septation, and the endocardial cushions specifically.

A number of different cell populations contribute to the structures affected by the abnormalities we have seen. The muscular septum is formed from components of the primary and secondary heart fields, with some contribution from the proepicardial precursor cells [38, 39]. The aorticopulmonary septum derives from the second heart field also. All valves develop from endocardial cushions. Both endocardium and ectoderm contribute to valve formation and in addition, neural crest cells (from ectoderm) contribute to valve formation in the OFT. Therefor there is not one specific cell population that could be linked to all the defects exclusively.

There is also evidence from other mouse models to support *PRKD1* as a cause of the specific CHD defects we have seen. *Mef2c* knockout mice die around E10 due to cardiac abnormalities [739]. Abnormalities include failure of normal heart tube looping, a dilated dorsal aorta, disorganised endocardium, dilated ventricles and dilated cardiomyopathy [793]. Knockout of *Mef2c* in the second/anterior heart field results in nearly 60% of embryos affected with VSDs or outflow tract abnormalities in combination with a VSD [786]. All had abnormalities of the aortic or pulmonary valve. Whilst there are similarities in the phenotypes described between *Prkd1* and *Mef2c,* however the VSDs seen in the *Mef2c* mouse were membranous and not muscular.

Loss of *Prkd1* could result in a similar picture to loss of *Mef2c*, given Mef2c would be either absent or bound to HDACs and this might be one mechanism by which mutations in *Prkd1* could result in CHD. Alternatively, gain of function SNVs in *MEF2C* have also been shown to cause outflow tract abnormalities in humans [794]. p.A103V, was identified in an individual with pulmonary atresia and VSD. However, suggesting that loss of *Prkd1* is equivalent to loss of *Mef2c* is likely to

be too simplistic an explanation, given the number of other genes involved and up regulation of normal alleles for example.

Mice deficient for both HDAC5 and HDAC9 have lethal VSDs [550]. However, loss of HDACs would presumably result in less repression of the *MEF2* gene targets, whereas loss of *Prkd1* should result in more repression of *MEF2* genes and therefore produce a different effect. It is also likely that alteration of HDAC phosphorylation by PRKD1 can be at least partially compensated for given evidence that HDACs can be phosphorylated by more than one PRKD isoform [715, 716].

TBX5 is also required for normal septal development. *Tbx5* knockout mice develop VSDs, including muscular VSDs [103]. *PRKD1* relieves the repression of *TBX5,* mediated by *HDAC4/5* [785]. *TBX5* has a well-established role in heart development and might be another possible mechanism by which alterations in *PRKD1* lead to CHD.

IAA is a rare defect, but is seen in 22q11.2 deletion syndrome and 1q21 deletions (*Cx40)* [795]. Animal models suggest a link with endothelin signalling pathways [796] and genes including *Foxc1*, *Foxc2, Cited2* and *TBX1* [76, 797, 798]. Abnormal apoptosis has also been shown to contribute to cardiac defects including IAA and septal defects [799]. PRKD1 is cleaved by caspase-3 at Asp-378 between the CRD and PH domains under conditions of apoptosis. The result of this cleavage is not clear and both activation and loss of function have been demonstrated [672, 800, 801]. Although *Prkd1* is involved in apoptotic processes and could potentially contribute to the formation of IAA in this way, the exact mechanism by which IAA results remains to be determined.

Overall, the types of CHD seen in mice and humans could be explained by deficient septal formation primarily, and endocardial cushion malformations.

More detailed expression studies during the formation of the septae and the endocardial cushions are required.

4.10 Limitations and Future Work

The major limitation of the work in this chapter is small sample size. This relates to both the number of individuals reported with CHD and Ectodermal defects, and the number of mice studied. In addition to this functional evidence is currently lacking. Determination of PRKD1 protein levels in these mice is ongoing on the brain, liver and heart tissue I collected. Kinase assays are being carried out by other members of the Brook Laboratory. More detailed expression studies of *Prkd1* in embryogenesis, including the endocardial cushions are also required. It is unfortunate that the CRISPR work in the Zebrafish was unsuccessful in vivo as this might have added further evidence for mutational effect.

Given the mouse models have been generated and colonies successfully established it offers the opportunity to gather more comprehensive data. Sacrifice and collection of the heart for HREM limits the analysis of other parts of the mouse. Whole embryo HREM would allow for better visualisation of the great vessels, which is important given the finding of IAA in a single embryo. This would also allow us to detect limb and structural brain abnormalities for example. It may also show evidence of cardiac insufficiency if hepatomegaly and other sequelae were present. The same would apply for whole mouse MRI scans.

I had planned to measure heart weight to tibia length ratios. This would allow me to compare heart size between mutants and WT embryos. Phenotyping pipelines such as those used in the IMPC project weigh the heart immediately after it is removed. This was not possible due to the requirement for immediate transfer of the heart to warm PBS for lysis. There was also a variable amount of soft tissue attached to the superior aspect of the heart at explantation. In retrospect it might have been possible to weight the hearts after removal of all blood, fixation and subsequent fine dissection. I was easily able to lyse soft

tissue from the tibia using the same mix of lysate and proteinase K as I used for DNA extraction.

Other aspects of embryonic phenotyping that could be improved include more accurate measurement of CRL and weights in adult heterozygous and WT mice. Whilst I did review every placenta for gross abnormalities, additional collection of the placenta for histology would provide better evidence for placental function.

The mouse models did not have a high incidence of CHD. Larger cohorts are required to determine the exact incidence and whether IAA ($Prkd1^{Em1+/-}$) and VSDs ($Prkd1^{Em2+/-}$) are a true phenotypic associations. The same applies to the low incidence of other possibly associated abnormalities. Only one minor limb defect was noted ($Prkd1^{Em2+/-}$) and one mouse with absent whiskers ($Prkd1^{Em1/+}$).

4.11 Conclusions

PRKD1 is a novel CHD gene. Whilst clinical reports are few at this stage, we have laid the groundwork for reporting of additional affected individuals and further investigation of abnormal function of *PRKD1* in the heart. At this stage, the emphasis should be on identifying other affected individuals to define the clinical phenotype and guide clinical management and investigations. Paired functional studies are required to understand the mechanism which causes the clinical phenotypes in individuals.

Chapter 5 Cardiac Findings in the Adult *Prkd1* Mouse

5.1 Introduction

Breeding of *Prkd1* mice to maintain the colonies and produce females for timed matings, meant that excess male mice were generated. These male mice provided an opportunity to collect further phenotype data in post-natal life, when there is less chance of missing subtle heart defects that might not be seen in the embryo. In addition, a single *Prkd1*[Em2/+] E15.5 embryo was identified with a muscular VSD using HREM (described in Chapter 4). This defect would not necessarily cause foetal demise or symptoms in early life, and therefore it was possible that other apparently normal *Prkd1*[Em2/+]adult mice might have CHD. We hoped to identify further affected mice by imaging adult mice. Importantly, *Prkd1* also plays a role in cardiac hypertrophy in response to stress, and variations from the normal hypertrophic response in a heart with CHD might be apparent in these adult mice. This chapter described the development of an MRI imaging protocol to assess two month old *Prkd1*[Em2/+] mice.

5.1.1 Role of Prkd1 in Cardiac Hypertrophy

PRKD1 regulates cardiac hypertrophy through its action on *HDACs* and *MEF2* signalling [668, 715, 732]. Its actions have been studied in a number of mouse models. These and the proposed mechanisms are described below.

Class II *HDACS* are signal responsive repressors of cardiac hypertrophy, and act by supressing myocardial expression of pro-hypertrophic genes via *MEF2* [736, 802, 803]. As described previously in chapter 4, PRKD1 phosphorylates HDAC5, leading to its nuclear export via the actions of XPO1/CRM-1. This produces subsequent *MEF2A* transcriptional activation, which results in activation of gene targets that cause myocyte hypertrophy and cardiac remodelling [715, 742]. *Hdac5* knockout mice develop enlarged hearts in response to increased afterload or constitutive activation of calcineurin, which

forms part of the cardiac stress signalling pathways. *Hdac9* knockout mice are also more sensitive to signals of cardiac stress. Mice deficient for both *Hdac5* and *Hdac9* have a thin myocardium and lethal VSDs [550].

The transcription factor CREB is also phosphorylated at Ser133 by PRKD1 [681]. This results in expression of *CREB* related genes. This includes atrial natriuretic factor and brain natriuretic factor [681, 804, 805]. These two molecules are secreted by the atria and ventricles respectively, in response to increased transmural pressure [806]. *C-fos* is also regulated by *CREB* and has been associated with cardiac hypertrophy [807, 808]. This suggests that there might be more than one possible mechanism by which *PRKD1* regulates cardiac hypertrophy.

PRKD1 might also influence contractility, as it mediates phosphorylation of TNNI3 (cardiac troponin I). This leads to reduced calcium sensitivity of the myofilament, and accelerated cross bridge cycling in rat ventricular cardiomyocytes, and confirms that *PRKD1* plays a role in regulation of the function of heart muscle [689, 774]. There is evidence to suggest that stimulation of *PRKD1* may be important in acute regulation of contractility of the heart in cardiac failure, when other proteins such as protein kinase A (a known regulator of contractility) are downregulated [741, 809].

5.1.2 Previous relevant *Prkd1* mouse models

Feilitz et al. generated a mouse expressing cre-recombinase in cardiomyocytes to produce a *Prkd1*conditional model [732]. The mice were viable and appeared entirely normal. They then performed arterial banding to increase the afterload, and found that the homozygous conditional knockout mice showed minimal fibrosis and hypertrophy, no reduction in contractility, increase in LV end diastolic diameter or bradycardia in response. There was no activation of foetal or fibrosis genes. The WT mice however showed poor cardiac function, reduced

contractility and heart rate, and up regulation of foetal, hypertrophic and fibrosis genes. In comparison, cardiac specific expression of activated *Prkd1* in mice leads to cardiomyopathy and death [742]. Initial cardiac hypertrophy was followed by ventricular dilatation and thinning of the ventricular walls and reduced contractility. This confirms the role of *Prkd1* in controlling hypertrophy, contractility and remodelling in stressed hearts.

Further evidence of its role in hypertrophy was shown by Harrison et al. who performed *Prkd1* knockdown with siRNA and found that this reduces subsequent hypertrophy seen in response to stress in both mice and in vitro cardiomyocytes [742].

The effect of β adrenergic stimulation on *Prkd1* is unclear. In comparison to these in vitro studies where *Prkd1* was not affected by β adrenergic stimulation [742], *Prkd1* was required to promote hypertrophy in response to β adrenergic stimulation in a conditional mouse model Feilitz et al. Other studies have shown that pressure overload and angiotensin II stimulation both require *PRKD1* for pathological remodelling [732].

5.1.3 Imaging Modalities

Given that HREM was used to analyse the embryonic hearts, this was considered as a method for analysing the adult hearts too. Personal communication with Tim Mohun revealed that he had had variable success with HREM in adult mouse hearts. Limitations included the heart arresting in systole and difficulty removing blood from the heart, which obscures morphology and impairs phenotyping. Dr Mohun had not found using antiarrhythmics or anticoagulants had helped. I did try a modified protocol with extended time for infiltration, but this was only partially successful (see chapter 4, section 4.67, figure 4.22)

A number of alternative imaging options were considered and are discussed below. The major limitations were the size of the adult mouse heart and the inability to carry out any intervention on mice prior to death, due to lack of an appropriate Home Office licence. Post mortem MRI presented its self as a feasible option, with the potential additional benefit of being able to carry out downstream work as the tissue would not require fixation. There is an MRI facility within the University which allowed easy transport of mice. Ideally I would have validated the findings using histology afterwards, but unfortunately this was not feasible due to time restrictions.

Traditional methods used to phenotype the mouse heart include paraffin embedding and sectioning, along with newer methods such as ultrasound, computed tomography (CT), optical projection tomography (OPT), magnetic resonance imaging (MRI), episcopic fluorescence image capture (EFIC) and high resolution episcopic microscopy (HREM). Some of these methods are non-invasive and some allow image reconstruction in multiple planes. At the time within the lab, we were exploring if there was a method of imaging we could use that would also allow us to carry out RNA studies afterwards. Many of these techniques rely on the use of fixative, dehydration and embedding, which makes downstream work more difficult. Another significant limiting factor was the absence of a specific project licence allowing scans under sedation, use of contrast or any other intervention on live animals. Possible false positive and false negative results led Lui et al. to recommend histological confirmation of phenotypes after all other imaging techniques [810].

MRI is one of the more popular techniques for imaging the mouse. However most of the published methods either rely on use of contrast and fixation in deceased animals with the associated risks of shrinkage, or require specific animal licencing to allow injection of contrast and imaging under anaesthesia in a live mouse. Given it is extensively used in both animal and human cardiac studies, and the University has a well-established MRI facility, I explored this option further and ultimately used MRI imaging in post mortem mice.

Advantages

- Non-invasive.
- Option for image manipulation and 3D constructions using the data collected.
- No radiation.
- Good resolution in soft tissues without contrast.
- Potential for imaging multiple samples at once [811].
- A wide variety of scanning options to answer the clinical question posed.

Disadvantages

- Longer scan times to achieve high resolutions.
- Technical difficulties such as wrap around artefact.
- Can be difficult to visualise smaller structures such as valves, papillary muscles and small vessels [810].
- Many protocols rely on either imaging with contrast in a live animal, or fixation and contrast in ex-vivo embryos.

MRI was pursued given the resources at the University Of Nottingham and the excellent soft tissue resolution it provides, even without contrast. As the Home Office Licence we were working under was for breeding only, any imaging of the adult mice would have to be carried out after death. Post mortem imaging does have some advantages. Image acquisition time is not limited by anaesthesia, there is no motion artefact, and longer scan times allow higher resolution images.

5.1.4 The Basis of MRI

MRI relies on the spin of protons contained within water molecules in cells. The direction of spin of protons varies throughout the body, unless it is aligned with

a magnetic field. MRI exploits the ability to manipulate proton alignment using a magnetic field produced by a superconducting magnet. The aligned protons have a net magnetisation that rotates around the direction of the main magnetic field. Radiofrequency (RF) pulses (low energy electromagnetic waves) cause the net magnetisation to rotate orthogonally to the main magnetic field. This is the resonance part of MRI. The rotating net magnetisation gives rise to a detectable signal. There are two processes that cause the decay of this detectable signal. These processes are known as relaxation and have the time constants T1 and T2.

Different tissues exhibit different relaxation times, which are known as T1 and T2. T1 is the longitudinal relaxation time and T2 is the transverse relaxation time. T1 is the time constant for the protons to regain their original pattern of spin within the magnetic field. The spin relaxes from the transverse plane to the longitudinal magnetic vector. Two important settings in MRI are the repetition time (TR^{MRI}) and time to echo (TE). The TR^{MRI} is the time between pulses in a specific slice, and TE is the time it takes between sending the RF pulse and receiving the signal. T1-weighted imaging relies on a short TE and TR^{MRI}. T2 is the time constant for the loss of alignment of proton spin in the transverse plane. It represents the time taken for the excited protons to go out of phase with each other along the line of nuclei spinning perpendicular to the magnetic field. T2-weighted imaging relies on a longer TE and TR^{MRI}. In MRI, it is common to acquire repeated measurements of the data, which are then added together to ensure that enough signal can be generated to be detected and used to form an image. This frequency information is translated into intensity information at each specific location and used to produce pixels of varying grey scale values.

5.1.5 Previous relevant MRI studies

Whilst planning MRI imaging in the adult mouse, I considered a number of published protocols for mouse embryo MRI imaging to see if any could be adapted for a post-mortem adult mouse. Embryos are obviously a lot smaller, and also show a comparative lack of contrast between tissues making it more

difficult to delineate structures adequately in early to mid-gestation [811]. MRI in live and dead embryos is possible. In live animals prospective gating can be used to combat movement artefact from respiratory and cardiac cycles. Gadolinium contrast is used to help delineate the anatomical structures. This can be introduced by injection, or by soaking the embryo in contrast solution after fixation.

Parasoglou et al. scanned in utero mouse embryos at E10.5 and E12.5 with manganese enhancement [812, 813]. Whilst this worked well for the cerebral vasculature, limbs, spine and CNS, it worked less well for the heart. Longer echo times were required for good contrast between the vasculature and other soft tissues in the earlier embryos (E10.5 for example). Use of contrast in this situation was not be possible under our current licence.

Schneider et al. have published a number of MRI protocols to image single mouse embryos with a resolution of 25 x 25 x 26µm voxels, validated by histology. These were fixed embryos and relied on a T1 weighted gradient echo sequence, rather than spin echo [814, 815] and fast spoiled 3D gradient echo sequence with T1 weighting [816]. They were able to identify ASD, VSD, DORV and aortic arch abnormalities. Subsequent to this they developed a protocol that allowed imaging of up to 32 embryos in a single scan, reaching a resolution of 43 x 43 x 36µm [817]. The embryos were fixed with PFA and embedded in 1% agarose gel containing Magnevist (gadolinium-DTPA). Similar resolutions were obtained by Zouagui et al. [818] using the same methods [817]. I discussed their methods with S. Bamforth who is one of the authors involved in pioneering the multi embryo MRI imaging using PFA and Magnevist. They had not tried using Magnevist in the adult mouse, but felt it would not be suitable given the requirement for adequate infiltration by the contrast agent.

Gabbay-Benziz et al. used MRI to characterise the E17.5 mouse embryo heart [819]. They removed the embryos and fixed them in 4% paraformaldehyde and held them in place using Fomblin (a perfluoropolyether oil). With an eight hour

acquisition time in a 17.6T MRI coil they achieved a voxel resolution of 51 x 53 x 78μm. This method allowed the identification of TGA and VSDs, and did not require any other contrast agent. The higher power coil used in this study meant that the resolution was adequate for analysis without complex imaging protocols or contrast required for the previous MRI studies of the mouse embryo. Unfortunately I did not have access to such a powerful magnet.

Post-mortem cardiac MRI also has a role in detecting cardiac lesions in humans and has been proposed as a first line assessment for all deceased foetuses and neonates. It could replace conventional autopsy and picks up the vast majority of CHD [820]. Unfortunately there is a relative lack of post mortem mouse MRI studies that concentrate on the heart, and much of the research focusses on the brain instead. Von Bohlen Und Halbach et al. used post mortem MRI studies to compare scans of anaesthetised live mice, unfixed post mortem in situ and fixed ex-situ brains [821]. Post mortem imaging of the unfixed brain provided much more detailed images than the scans on the live mice. This was due to the longer T2 weighted sequences that could be used, rather than the faster T1 weighted scans used in live mice. Scans of the fixed brains showed that the structures in the brain were much less well preserved and had decomposed significantly, compared to the immediate post mortem scans. Long term, but not short term fixation was associated with significant shrinkage. Other studies have confirmed adverse effects of fixation in the brain and artefacts have been reported in almost one third of fixed mouse brains. This was thought to be related to flow of perfusates through capillaries and inadequate fixation causing unusual appearances on scan [822]. This suggests that post mortem imaging is fruitful in the brain, and short term fixation can be considered if immediate scanning is not possible. Brain tissue is more fragile and unlikely to be directly comparable to cardiac tissue. Immediate post mortem scanning might be a viable option to consider.

Crowe et al. performed post mortem imaging in mice who had undergone ischaemia reperfusion injuries [823]. They were first scanned under anaesthetic

after injection with gadolinium intraperitoneally, and scanned using a 3D flash GRE MRI protocol (16 minutes, isotropic resolution of 130µm) post-mortem. The post mortem images were higher resolution with no movement artefact, although gadolinium would have also been present at post mortem scanning. The mice were killed with injection of potassium chloride which causes cardiac arrest. Interestingly, in 50% of cases, the heart had contracted between the live animal MRI and post-mortem scan. Presumably this would limit what could be detected from the images. The authors did not comment on specific structural defects, but it suggests that potassium chloride should be avoided if investigators wanted to detect structural lesions.

5.2 Aims

Given there are a number of *Prkd1* mouse models already reported, I chose to concentrate on the *Prkd1*em2(IMPC)Wtsi humanised mutation mouse. The aims were as follows:

- Devise an MRI protocol with adequate resolution to allow detection of CHD in E15.5 mouse embryos.
- Arrange MRI scans of *Prkd1+/+, Prkd1Em2/+ and if possible Prkd1Em2/Em2* adult mice.
- Determine if there is any evidence of CHD or cardiac hypertrophy in any of these mice.

5.3 Results and Discussion

The age of the mice at scan was decided by balancing the cost of keeping mice for prolonged periods of time, with the benefit of older mice with a larger heart and a longer period of life that might allow decompensation or heart remodelling if there was a structural cardiac abnormality. Based on previous studies we would not expect any features of hypertrophy without additional workload on the

heart [732, 742]. C57BL6 mice do become obese over time, although availability of fat suppression meant that body composition of the mice should not matter. Considering all these factors, a decision was made to image two month old male mice.

Ideally I would have liked to scan three groups of mice; *Prkd1*[+/+], *Prkd1*[Em2/+] and *Prkd1*[Em2/Em2] mice, but lethality in the *Prkd1*[Em2/Em2] mice was almost complete, therefore it was not possible to include this group. A single surviving female *Prkd1*[Em2/Em2] had to be killed by schedule one methods after she scratched her cheek, and this heart was used to trial HREM in the adult mouse heart, rather than MRI. The results have already been described in Chapter 4 (section 4.6.7, Figure 4.22). The surviving male homozygote was used for breeding and was therefore not available for study. This was intended to be a preliminary study to see if the technique is feasible, so power calculations are not appropriate.

5.3.1 Determining the Required Resolution for MRI Scanning

Review of the previous MRI studies reported in section 5.1.5, helped define the requirements for scanning the adult mice. A resolution of 25-50µm is sufficient to detect CHD in a mid to late gestation mouse embryo. For an adult mouse, equal or lower resolution should be sufficient to detect similar types of CHD. Lack of a licence for interventions on live mice, fixation processes that would prevent downstream work or not work in larger adult mice and lack of very high power magnets meant that the same protocols could not be used. Immediate post mortem imaging would have been ideal, but did not prove possible due to ongoing requirements of the MRI scanner, and ultimately the length of time required for the cardiac scans.

Turnbull et al. recommended a spatial resolution of 50µm isotropic (same resolution in all axes) for imaging of mouse embryo between E9.5-12.5, and 100µm for E13.5-15.5 to assess cardiovascular development [824]. Reaching resolutions less than 50 µm proved difficult, but the mice to be scanned would

be much bigger. I searched the available literature for the heart dimensions of the adult C57BL6 mouse to ensure adequate resolution for phenotyping. The results are shown in table 5.1 below.

Reference	Age and Mode	Landmark	Measurement
Doevendans et al. [825]	Adult mice fixed in systole Histology	LV diameter (mm)	2
		LV wall (mm)	1.1
		IVS (mm)	1mm
Baumann et al. [826]	12-14 week mice Echo	LVEDD (mm)	3.79 (F) 4.05 (M)
		LVESD (mm)	2.59 (F) 2.83 (M)
		DWTA (mm)	0.6 (F) 0.7 (M)
		SWTA (mm)	0.92 (F) 1.04 (M)
		DWTP (mm)	0.56 (F) 0.64 (M)
		SWTP (mm)	0.9 (F) 0.96 (M)
		Body weight (g)	20.4g (F) 26.4g (M)
Hinton et al. [827]	10 day – 2 month Echo	Aortic valve cusps length (µm)	342-474
		Aortic valve cusps width (µm)	Proximal 45-49 Middle 21-34 Distal 53-82 Largest at 10 days old
		Mitral valve cusps length (µm)	332-931
		Mitral valve cusps width (µm)	Proximal 60-78 Middle 17-41 Distal 35-57

Table 5.1 C57BL6 cardiac measurements found in the literature.

LVEDD = left ventricular end diastolic dimension, LVESD = left ventricular end systolic internal dimension, DWTA = left ventricle anterior end diastolic wall thickness, SWTA =left anterior wall end systolic thickness, DWTP = left ventricle posterior end diastolic wall thickness, SWTP = left ventricular posterior end systolic wall thickness

Review of the known dimensions for structures in the C57BL6 mouse suggests that most structures will be visible, although the valves may not be seen along their entire length. It is likely though that in the event of a significant valve lesion, other abnormalities would be apparent such as post stenotic dilatation.

5.3.2 Preservation of Mice by freezing

To reach the required image resolution without contrast, long acquisition times were required. This meant that initially overnight, and then weekend scanning was arranged. Due to ongoing commitments within the MRI department and availability of suitable mice, the scan could not always be carried out on the same day that the mice were culled. To get round this issue we tried freezing the mice after culling, and then thawing them for scans.

I considered whether the freeze thaw process could damage cardiac structures at the structural level that I was considering. Studies in canine tissue have shown that freezing can cause a range of abnormalities. This includes loss of staining, extracellular fluid accumulation and cell shrinkage, fractures, haemolysis and haematin formation, prominence of collagen in alveolar septae and the meninges, and intracellular vacuoles in epithelial cells [828]. On an architectural level bronchial cilia were lost. However the authors felt that the tissue could still be adequately examined. In human liver and heart tissue, freezing at -12 to -80C and subsequent thawing resulted in larger extracellular spaces and cell shrinkage [829]. Longer storage time led to more pronounced effects. Repeated freeze thaw cycles (-10, -18C) in buffalo muscle did result in some more significant changes including shrinking and tearing of the muscle fibres [830].

Freezing and then thawing of the mice could lead to abnormalities at the cellular architecture level. I reduced the chances of this occurring by freezing for the shortest time possible, and only using a single freeze thaw cycle. Given I was examining the hearts for much larger supracellular abnormalities, the evidence suggests that freezing the mice should not induce artefacts that would alter my results. In addition, alternative methods such as formalin fixation and paraffin embedding also lead to cell shrinkage but are still used to detect structural lesions in embryonic hearts.

A number of possible methods are available to reduce the chances of damage induced by the freeze thaw process. Snap freezing to -80c and cryosectioning tend to be used. Reduction of ice crystals formation by dehydration and reduced freezing time has proved useful in rat hearts and skeletal muscle [831]. Skeletal muscle samples can be immersed in isopentane at -150C [832] and thawed and re-frozen if there is artefact. These samples were much smaller though (1.5cm in maximum diameter). In our protocol we did not immerse the tissue or snap freeze the mouse as they were large samples. We could try faster freezing protocols with the whole mouse to see if there was any difference in architecture in future studies, or if there was any evidence of tissue damage visible on the scans.

5.3.3 Test MRI Scans

A protocol to image the whole mouse thorax was produced and initial studies were carried out on a deceased mouse to ascertain the image quality and contrast between tissue and blood. A standard 50μm scanning procedure was used initially, then with zero-filled image and then a shortened TR with twice as many iterations. A number of test scans were carried out with different resolution, acquisition time and echo chain. It is necessary to use spin-echo for scans at higher fields such as this, as there is a high incidence of susceptibility

artefacts (artefacts and distortions as a result of local changes in the magnetic field). A bright blood protocol was thought to be the best starting point to act as a contrast in the heart and great vessels.

Scans were carried out overnight initially to allow plenty of time for image acquisition. We then carried out further test scans over the weekend period at 50um, 60um and 70um resolution isotropic (Figure 5.1). Scanning at a lower resolution such as 70um allows for more signal averaging which reduced the "noise" in the images. After comparison of these three resolution scans, it was felt that it would be best to opt for the highest resolution possible so as to not miss any lesions, because the amount of noise was fairly minimal at 50um resolution. Multiple scans were then run over a whole weekend period (>60 hours) which showed that the samples were stable enough to collect data for several signal averages over a longer period than just the overnight scan.
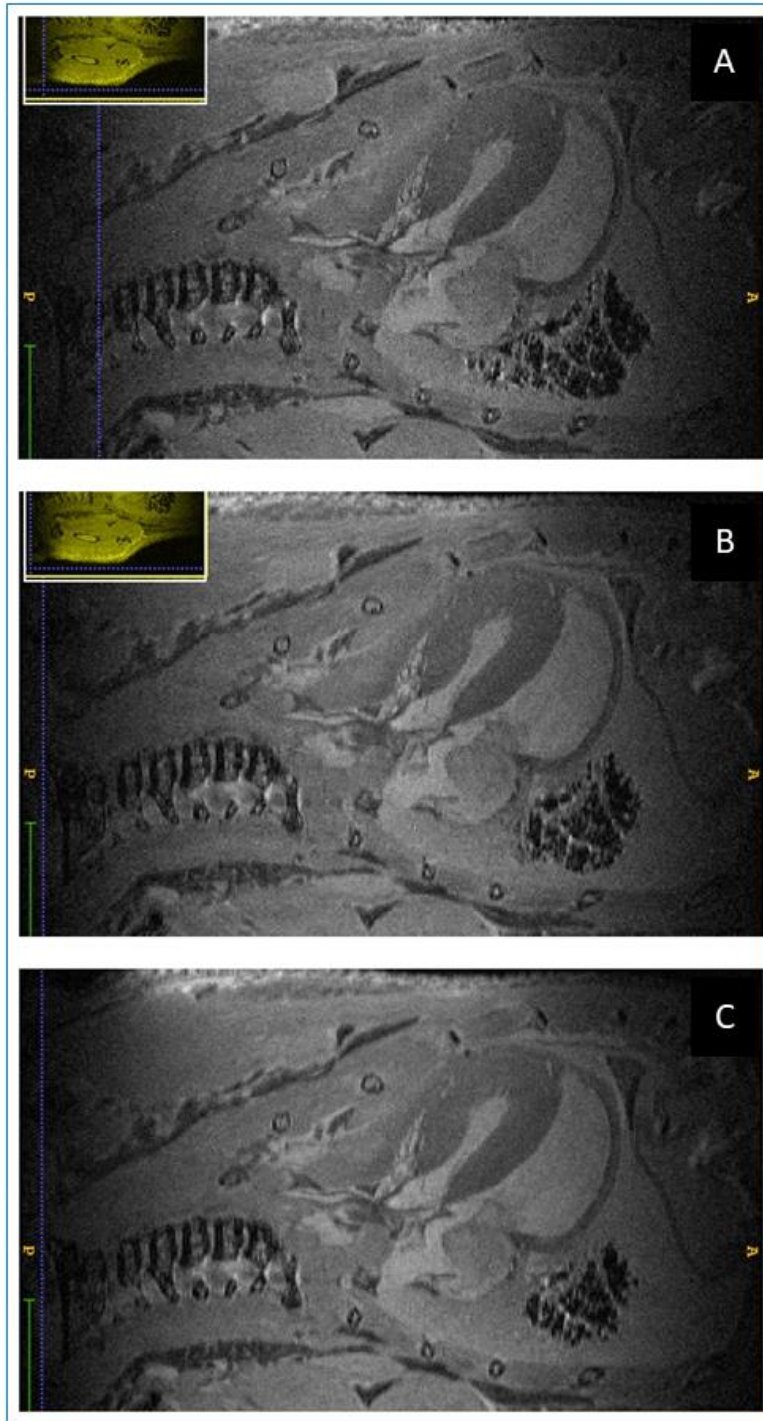
Figure 5.1 MRI images at differing resolutions.

Comparison of the same mouse showing the LVOT at 50 (A), 60 (B) and 70µm (C) resolution. The signal to noise ratio is improved at 70 µm. The right and left ventricles are shown, and the right atrium. The structural details appear the same across all images and there does not appear to be any significant difference across the heart at different resolutions.

Wrap around artefact was present on some of the images, but was sufficiently distant from the heart to allow a full analysis of cardiac morphology (figure 5.2 and 5.3). Triple images occasionally occurred on test scans as a result of an extra-long echo chain (number of echoes used in fast spin echo/RARE imaging, 16 resulted in the triple image artefact) (figure 5.3). Some of the initial images were darker towards the posterior parts of the chest, which sometimes impeded the view of the heart. This was due to using a coil which allowed a very small field of view and was resolved by positioning the embryo so that the heart was at the 6 o'clock position.
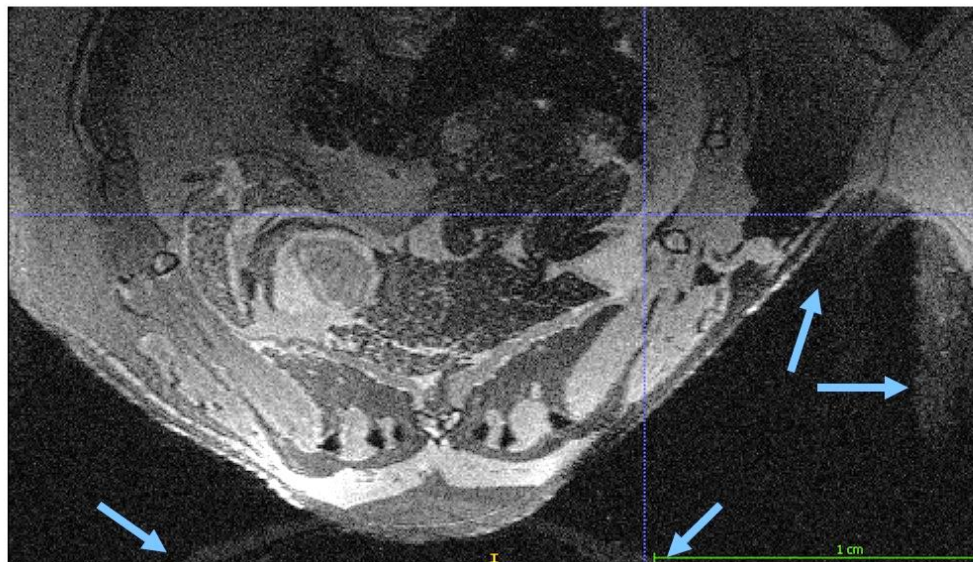
Figure 5.2 Wrap around artefact.

Artefact indicated by blue arrows. The thorax can be seen reflected at the bottom of the image (anterior of the mouse) and the limb on the right side of the picture.

Figure 5.3 Wrap around artefact and image shadow.

MRI test image showing wrap around artefact from the anterior of the thorax (bottom blue arrows) and the limb (blue arrow on right). In addition there is an appearance of blurred boundaries of the tissue, as if multiple images are overlaid, thought to be a result of using an extra-long echo chain (16 echoes). Cross hairs intersect in the left ventricle.

5.3.4 Structural Cardiac findings in 2 month old *Prkd1*Em2(IMPC)Wtsi mice

A total of 23 *Prkd1*Em2 mice were scanned. This included 13 male *Prkd1*Em2/+ mice aged 2 months, and some additional mice that were used for test scans. Test scans were carried out on mice which included some at 1 month of age and a female mouse. All 23 mice had detailed external phenotyping and all were normal. No internal phenotyping by dissection was carried out as they were required intact for the MRI studies. An example of the transverse, coronal and sagittal images with 50 x 50 x 50µm resolution are shown below (Figure 5.4). The ability to reference a structure in three planes simultaneously is helpful when identifying cardiovascular anatomy. Normal structures identified on the MRI scans are shown in figure 5.5 - 5.9 below.

Only two mice of the 23 scanned had structural defects. This included a single two month old male *Prkd1*[+/+] mouse with a large right atrial appendage (RAA) (figure 5.10 A). The significance of an enlarged RAA is unclear. It is not characteristic of, or pathognomonic for anything specific. The right atrium and rest of the heart appeared normal. The importance of this in a WT mouse is not certain.

The second mouse with a structural abnormality was a *Prkd1*[Em2/+] male mouse. A dark mass was identified apparently attached to a leaflet of the tricuspid valve. This measured 0.77x0.8487x 0.8003mm (Figure 5.10 B). It could it be fibroelastoma, as it looks to be associated with the valve, rather than the papillary muscles. The significance of this is uncertain.
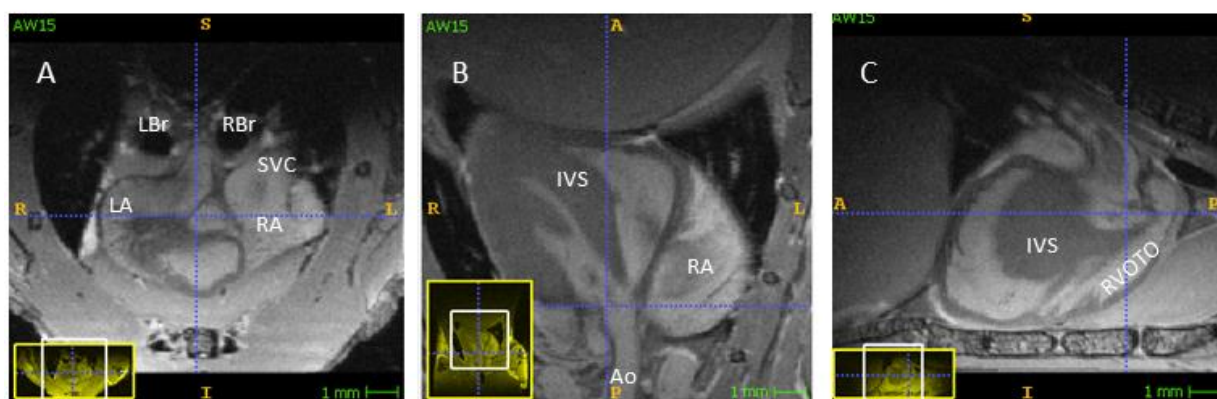
Figure 5.4 Typical images shown in three planes generated in ITK-Snap.

A: Transverse, B: Coronal, C: Sagittal sections of a normal *Prkd1*[Em2/+] mouse. Crosshairs pinpoint the aortic valve. Anterior is the bottom of the image and the right side of the heart is on the right side of the image in the transverse cut.

Ao: Aorta, IVS: Interventricular septum, LA: left atrium, LBr: Left main bronchus, RA: Right atrium, RBr: Right main bronchus, RVOTO: Right ventricular outflow tract, SVC: Superior vena cava.

Normal structures identified on the MRI scans are shown in figure 5.5 - 5.9 below.
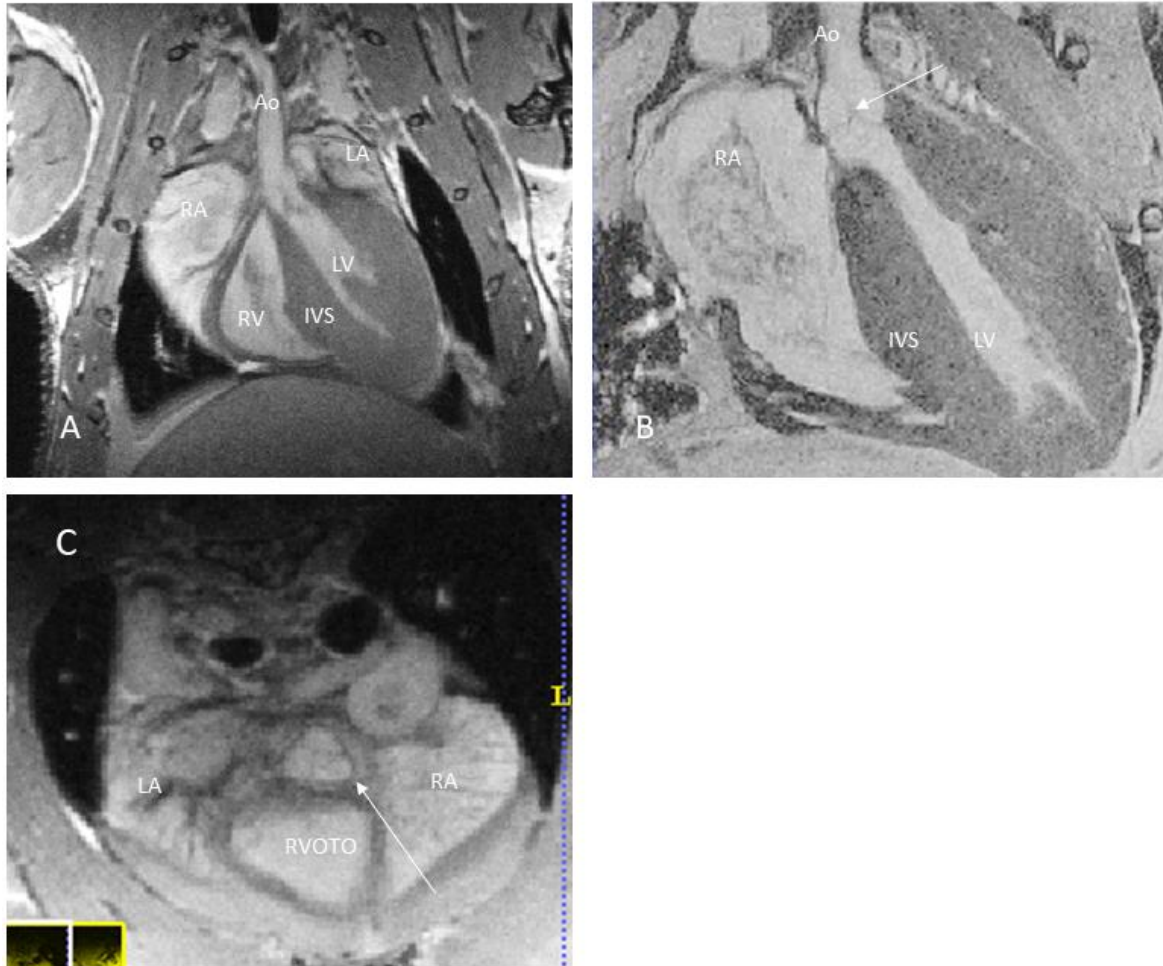
Figure 5.5 Left ventricular outflow tract and aortic valve.

Images of a normal *Prkd1*Em2/+ mouse.

A: Coronal slice through the mouse thorax through the heart showing the LVOTO. B: Further coronal slice showing the LVOTO and AV in more detail. The valve leaflets (arrow) in cross section, bulbous shape of the sinus of Valsalva and the sinotubular junction can be seen. C: Transverse slice. The trileaflet structure is just discernible (arrow). Anterior is inferior on this image.

Ao: Aorta, IVS: Interventricular septum, LA: left atrium, LV: Left ventricle RA: Right atrium, RV: Right ventricle, RVOTO: Right ventricular outflow tract.
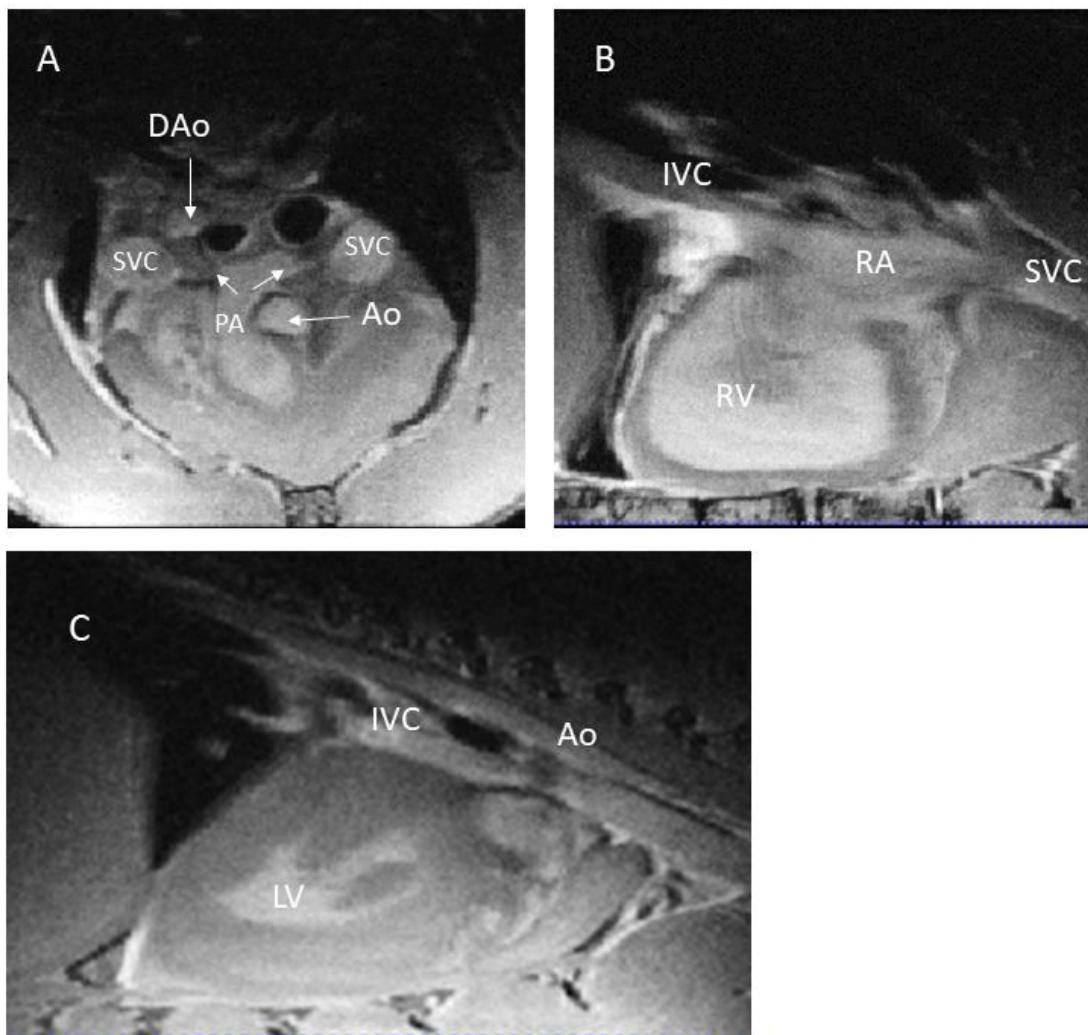
Figure 5.6 Great Vessels on MRI sections of a *Prkd1*^{Em2/+} mouse.

A: Transverse slice across the top of the heart at the level of the pulmonary artery bifurcation (PA and arrows). Both the ascending and descending aorta can be seen. Bilateral SVC are present. B: Sagittal slice through the right ventricle showing the SVC and IVC. C: Sagittal slice through the left ventricle showing both the IVC and descending aorta.

Ao: Aorta, Dao: Descending aorta, IVC: Inferior vena cava, IVS: Interventricular septum, LV: Left ventricle, PA: pulmonary artery, RA: Right atrium, RV: Right ventricle, SVC: Superior vena cava
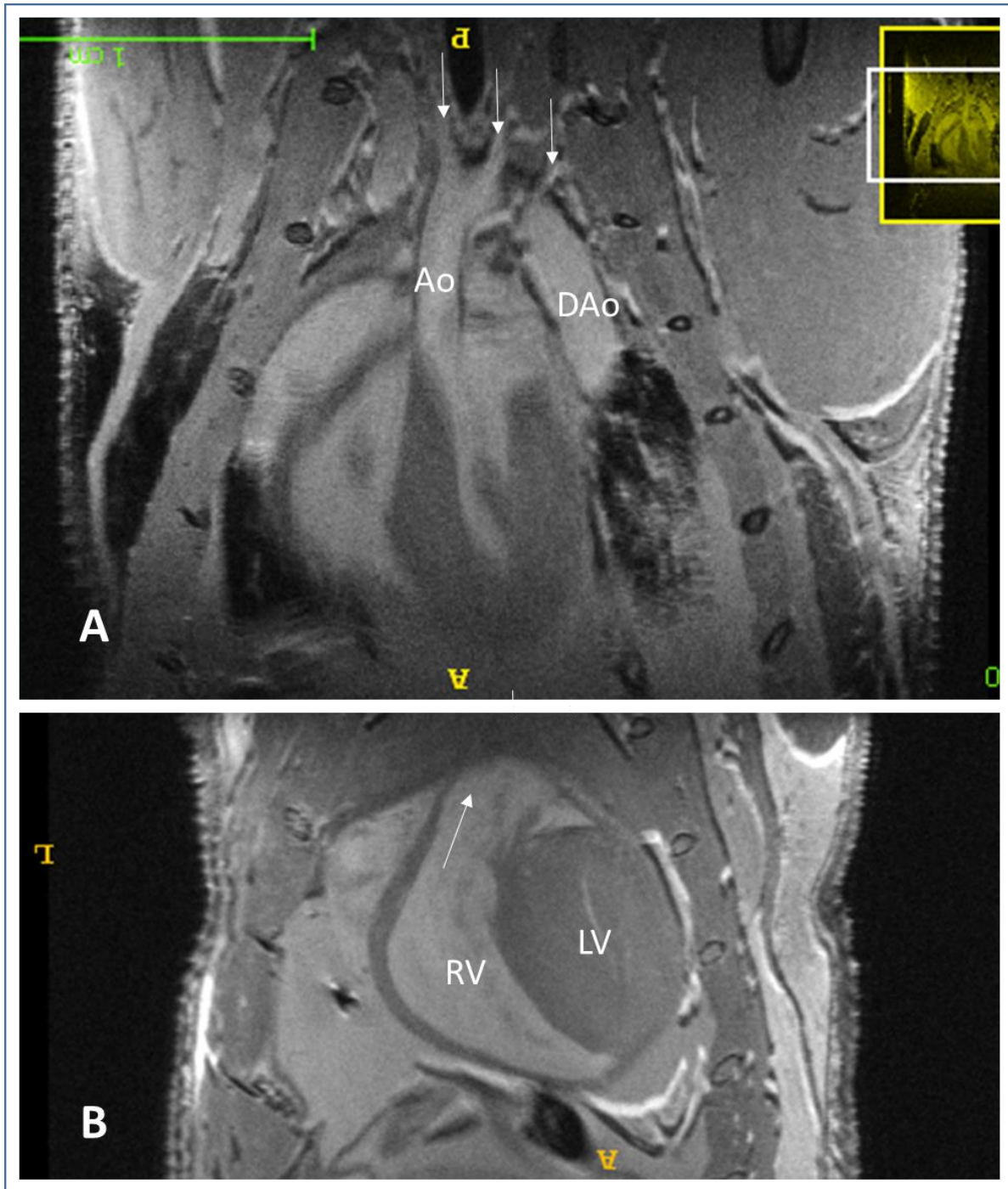
Figure 5.7 Right and left outflow tracts in a *Prkd1*^{+/+} mouse.

A. Aortic arch. The ascending, descending and arch portions of the aorta can all be seen. The three main branches are also shown (arrows).

B. B. Coronal sections of a normal *Prkd1+/+* mouse displaying the right ventricular outflow tract (arrow).

Ao: Aorta, Dao: Descending aorta, LV: left ventricle, RV: Right ventricle
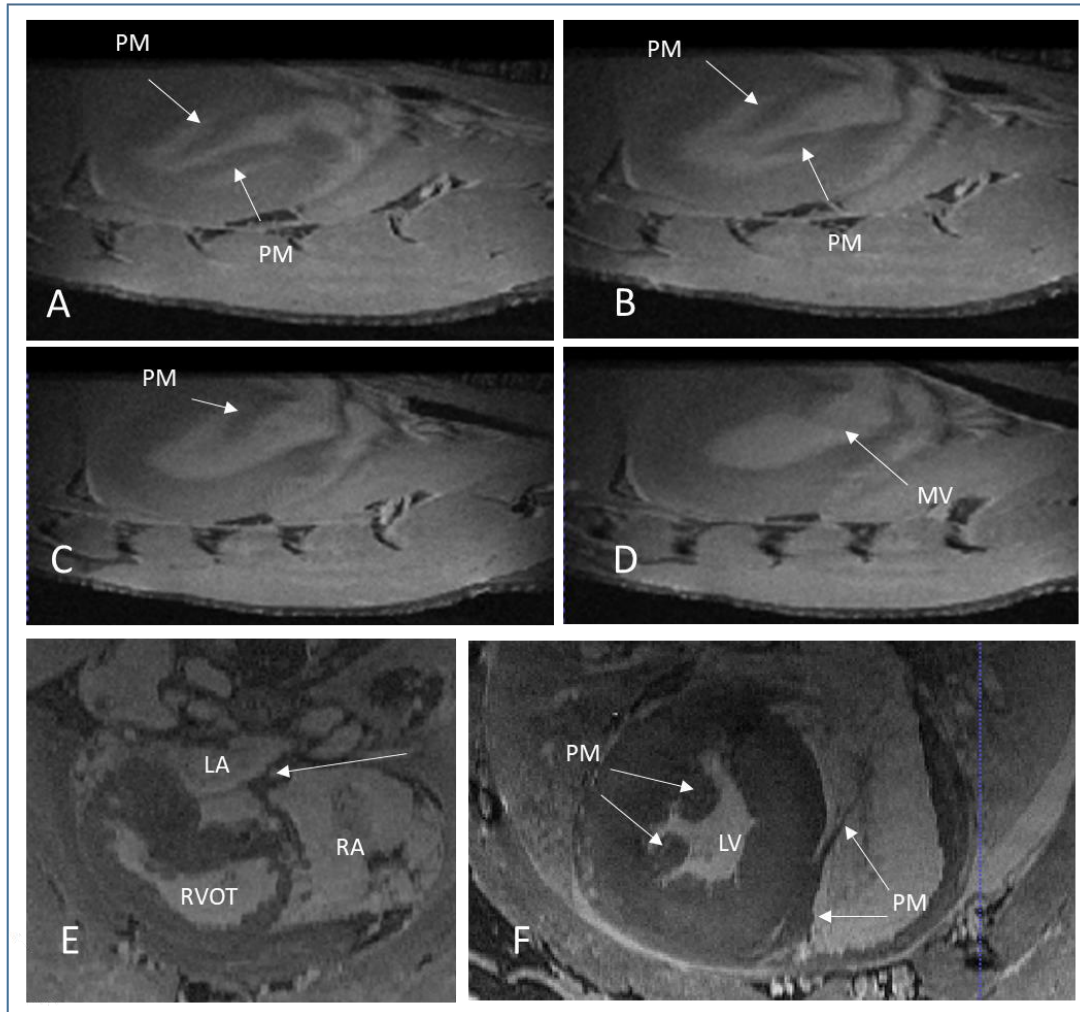
Figure 5.8 Valves and atrial septum in a normal *Prkd1*[+/+] mouse.

A-D. Mitral valve and papillary muscles. 4 images moving up from the base of the papillary muscles (A), following their course up to the mitral valve where the leaflets are just visible (image D). E-F: Atrial septum and components of atrioventricular valves E: Transverse slice showing the intact atrial septum (arrow). Part of the right ventricular outflow tract is seen. F: Transverse image through the ventricles to demonstrate the papillary muscles. In the right ventricle the papillary muscle can be seen reaching up towards the tricuspid valve. Comparison between the thick walled left ventricle and thinner walled right ventricle can be seen.

LA: Left atrium, LV: Left ventricle, MV: Mitral valve, PM: Papillary muscle, RA: Right atrium, RV: Right ventricle, RVOT: Right ventricular outflow tract
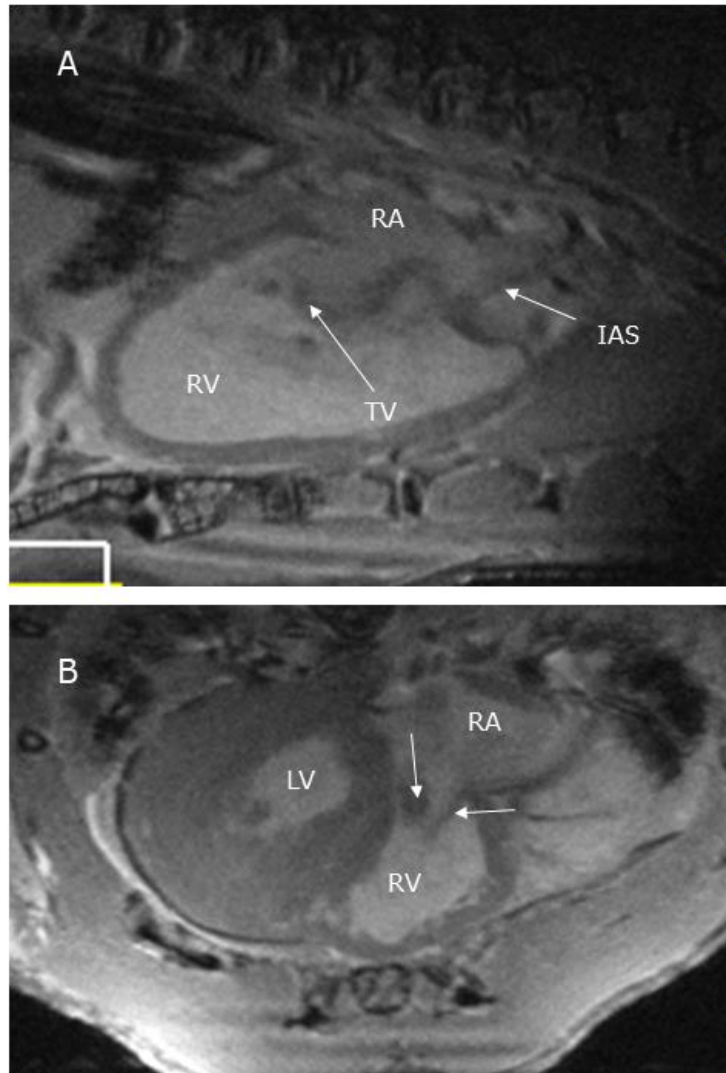
Figure 5.9 Atrial septum and components of atrioventricular valves in a normal _Prkd1_$^{+/+}$ mouse.

A: Sagittal section through the right atrium and ventricle showing the tricuspid valve between them. Part of the atrial septum is also visible

B: Transverse slice through the heart to illustrate the tricuspid valve. The leaflets are marked with arrows.

IAS: Interatrial septum, LV: Left ventricle, RA: Right atrium, RV: Right ventricle, TV: Tricuspid valve
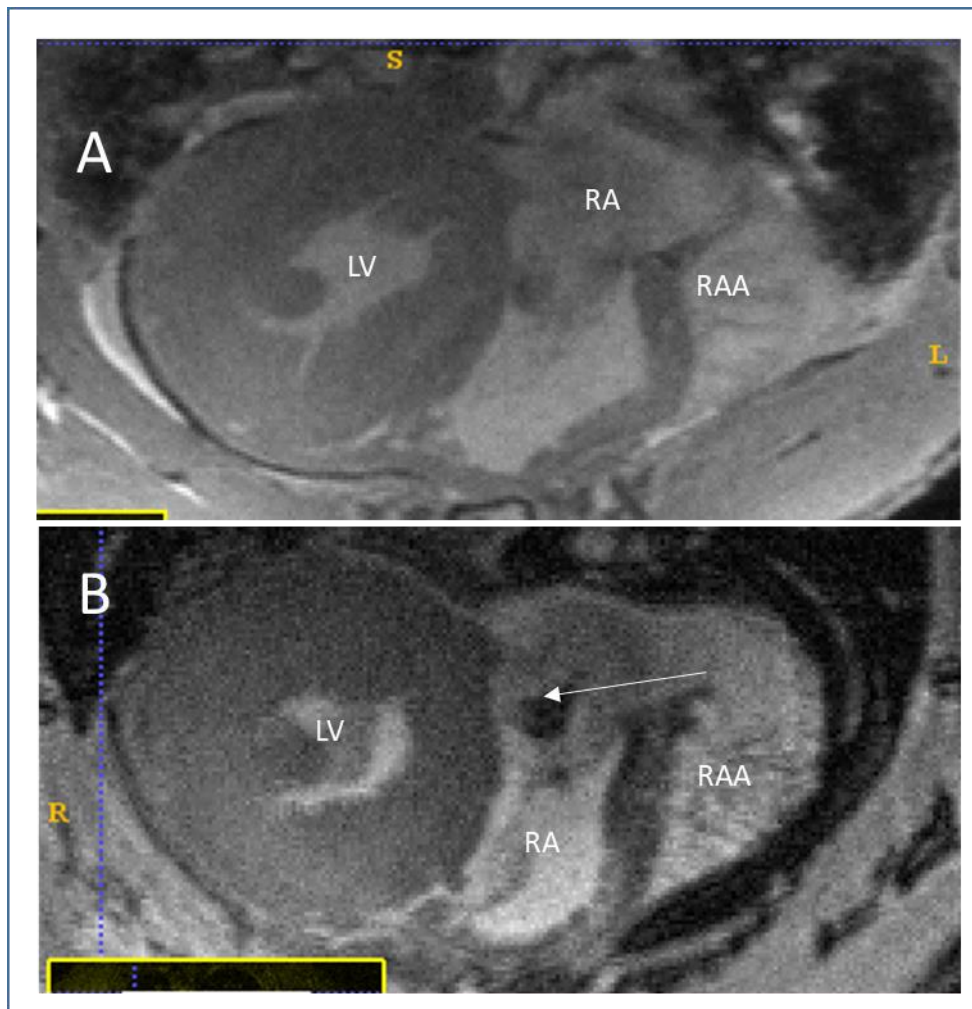
Figure 5.10 Abnormal cardiac phenotypes detected.

A. Transverse section of a *Prkd1*+/+ mouse. The right atrial appendage appeared much larger than in other mice. Although difficult to appreciate in a single section, it took up a significant portion of the right side of the thorax. There were no other abnormalities and the significance of this is uncertain. B. Dark mass on TV leaflet in a *Prkd1*Em2/+ mouse. A dark mass is apparent (arrow) and is associated with the TV. It measures 0.77x0.8487x 0.8003mm. This appearance and its apparent association with the valve rather than the papillary muscles, mean that it could it be fibroelastoma. It does not appear to be clot and there are no other real differentials.

LV: left ventricle, RA: Right atrium, RAA: Right atrial appendage

Overall, no structural abnormalities were identified with the exception of a possible fibroelastoma associated with the TV, and an enlarged right atrial appendage (RAA) in a wild-type mouse. The significance of these two findings is unclear. The fibroelastoma was identified in a *Prkd1*^{Em2/+} mouse. Fibroelastomas are rare tumours that occur on the heart valves. Whilst histology would have been helpful to confirm if this was indeed a fibroelastoma, the MRI appearance is typical and there are few other differentials. It is unclear if this has any relevance to the genotype of the mouse. The enlarged RAA was identified in a *Prkd1*^{+/+} mouse. An enlarged RAA is not indicative of any particular condition. Ideally real time assessment of the TV is required to rule out severe tricuspid regurgitation as a cause, as well as an assessment for pulmonary disease. However, I would expect right atrial enlargement too if these pathologies were present.

5.3.5 Ventricular Wall and Septal Thickness in 2 month old *Prkd1*^{Em2(IMPC)Wtsi} mice

Measurements of the thickness of the ventricular walls and septum were made in a total of 13 mice, 8 male *Prkd1*^{Em2/+} and 5 male *Prkd1*^{+/+} mice, who were exactly 2 months old. Measurements were taken in the short axis, at the mid-ventricular level where the papillary muscles are present. The average wall thickness at each site is shown in table 5.2 below. The site of the measurement and corresponding p values are shown in figure 5.11. Box and whisker plots for each set of measurements are shown in figures 5.12 and 5.13.

Site	*Prkd1*+/+ (mm)	*Prkd1*Em2/+(mm)	T test
RV free wall	0.408	0.483	0.012
IVS	1.203	1.348	0.062
LV anterior wall	1.116	1.233	1.110
LV posterior wall	1.106	1.261	0.118
LV lateral wall	1.162	1.158	0.947
LV internal diameter (AP)	2.696	2.680	0.929
LV internal diameter (transverse)	1.997	1.974	0.865

Table 5.2 Ventricular wall thickness in Prkd1+/+ and Prkd1Em2/+ mice.

Average measurements of wall thickness at mid ventricular level in two month old male mice measured on MRI. IVS: Interventricular Septum, RV: Right ventricle.
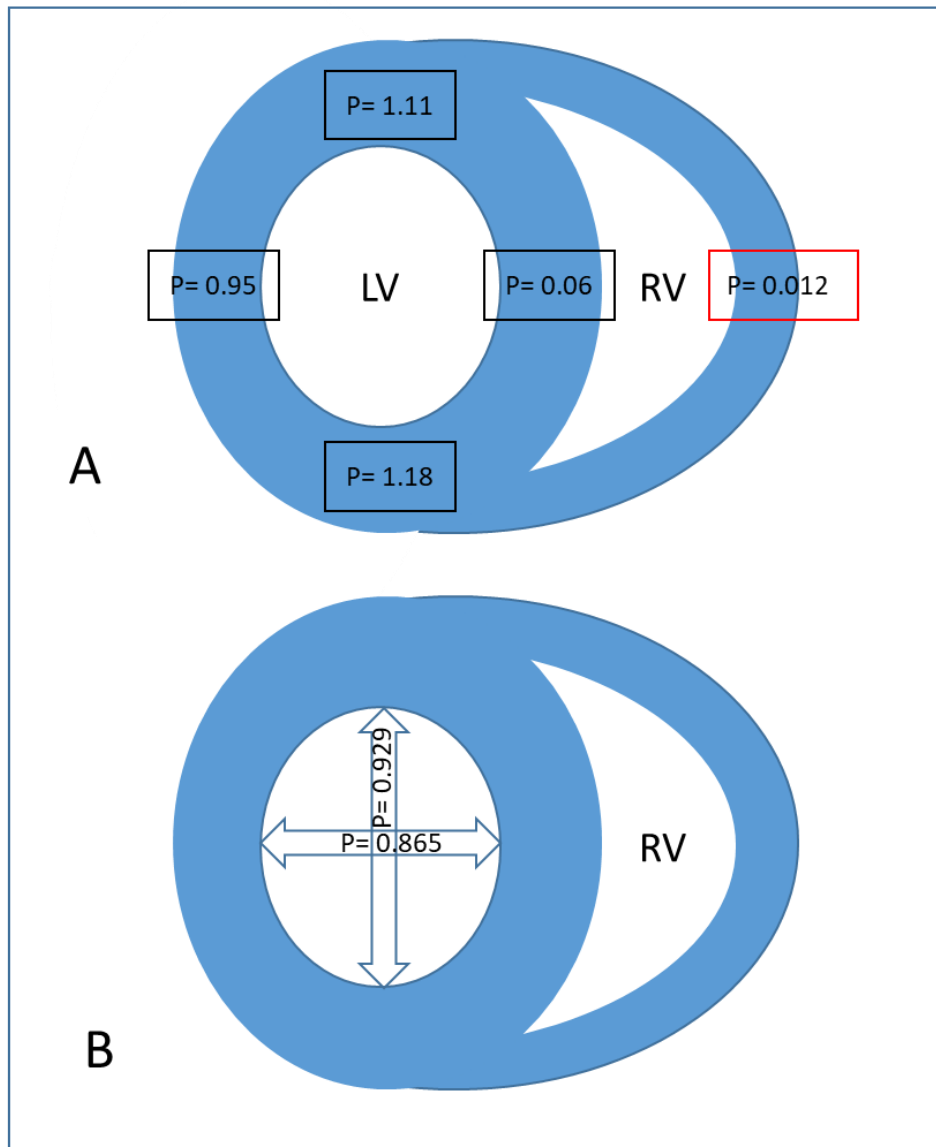
Figure 5.11 Representation of the measurements of ventricular wall and septal thickness in *Prkd1+/+* and *Prkd1Em2/+* mice.

The only significant value calculated by two tailed T test with unknown variance was the measurements of the right atrial free wall, highlighted in red.

A: Measurements of the width of the ventricular walls and septum were measured at the level of the papillary muscles (mid ventricle). P values at each location we measured are shown. B: Internal diameters of the left ventricle were taken in both anteroposterior and transverse directions. There was no significant difference between the two groups of mice.

LV: left ventricle, RA: Right atrium, RAA: Right atrial appendage

Figure 5.12. Box and whisker plots for all measurements of the ventricular walls and interventricular septum. Two tailed T test with unequal variances was insignificant in all measurements with the exception of the RV free wall.

LV: left ventricle RV: Right ventricle
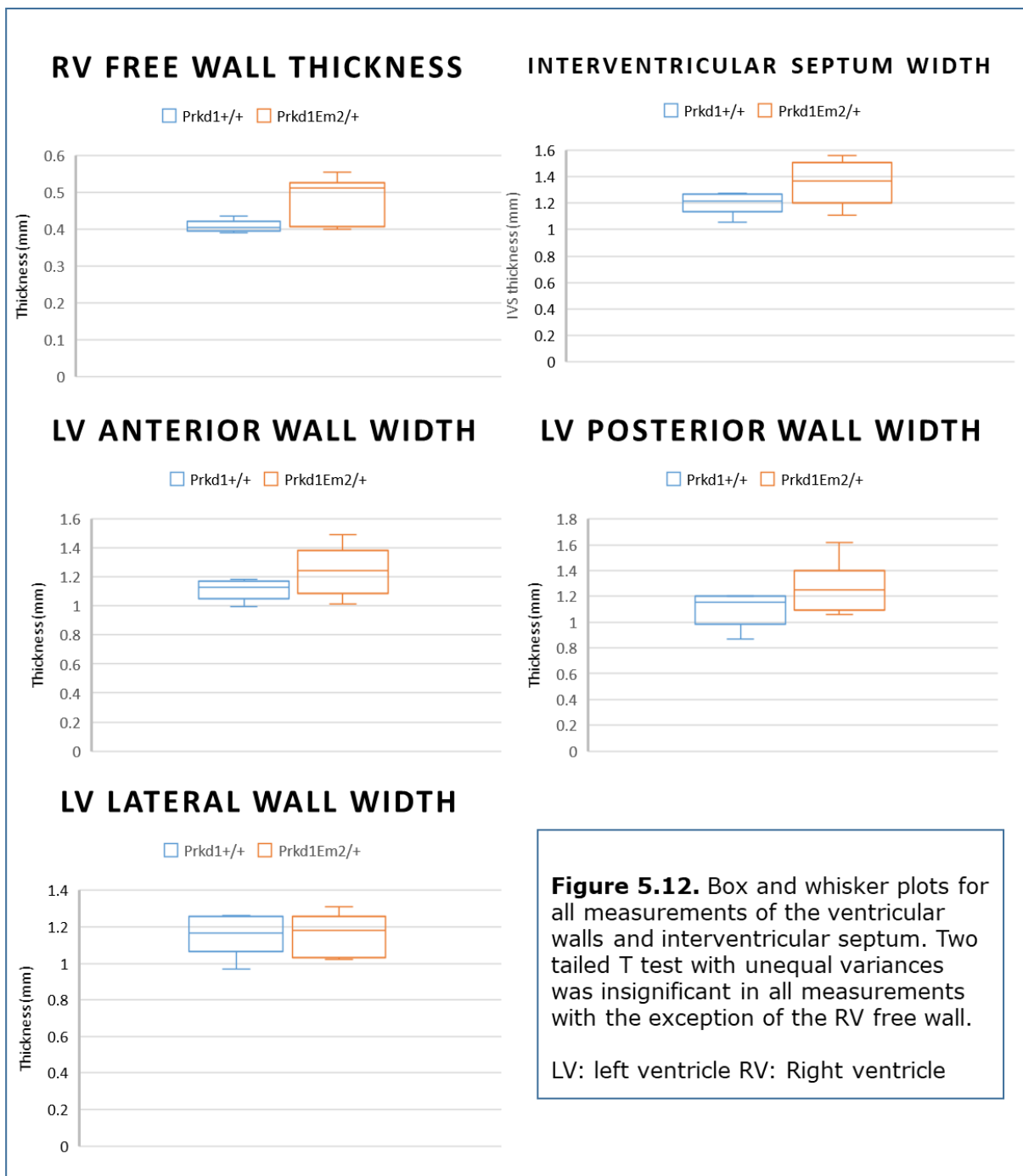
Although p values were insignificant in all except right ventricular free wall, it is evident that there is a trend towards increased wall thickness in the *Prkd1*[Em2/+] mice in all measurements except the lateral LV wall, when compared with the wild type mice.
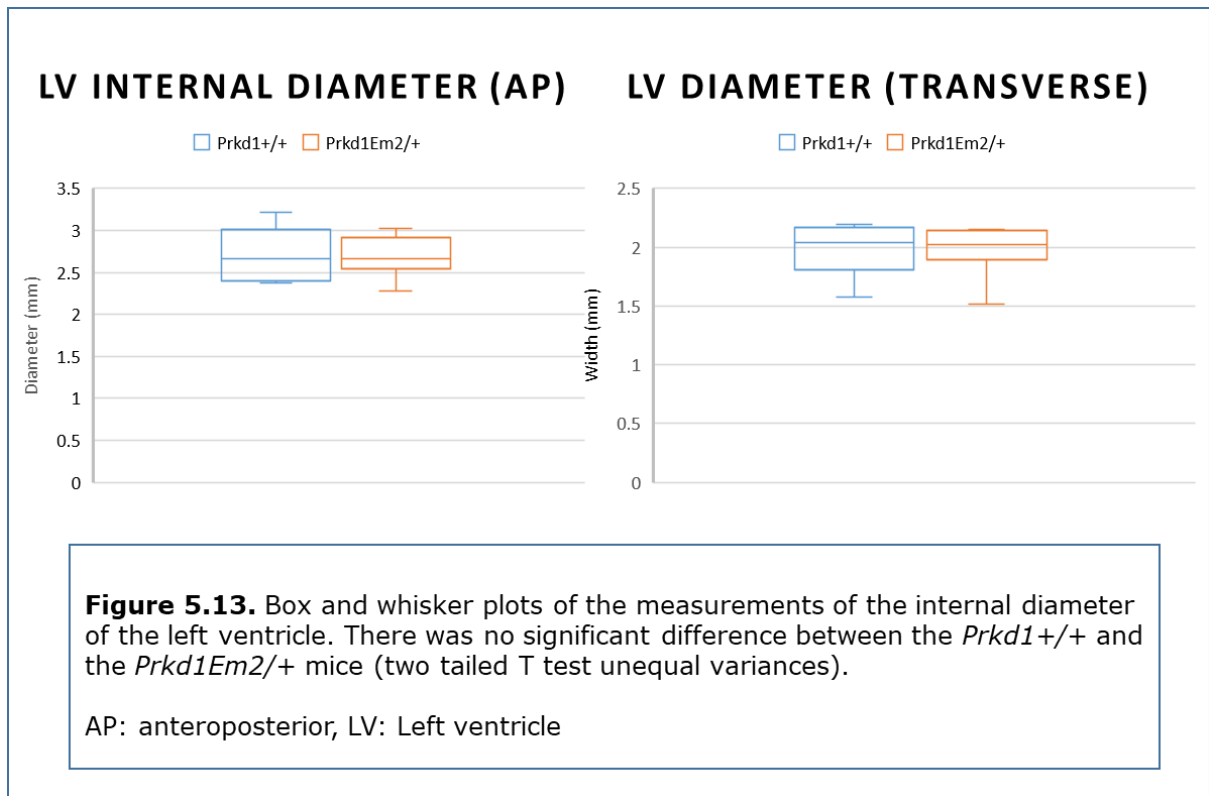
Figure 5.13. Box and whisker plots of the measurements of the internal diameter of the left ventricle. There was no significant difference between the *Prkd1+/+* and the *Prkd1Em2/+* mice (two tailed T test unequal variances).

AP: anteroposterior, LV: Left ventricle

Measurements of the aortic root were taken at the level of the sinus of Valsalva. Three measurements were taken across the valve cusps in the short axis and averaged. Two further measurements taken in the long axis (coronal and sagittal planes) and averaged. The results are shown in table 5.3. Figure 5.14 illustrates the sites of these measurements and the associated P values based on a two tail T test assuming unequal variances. None reached significance. Box and whisker plots in figure 5.15.

Site	*Prkd1*+/+ (mm)	*Prkd1*Em2/+(mm)	T test
SoV (across the cusps)	1.031	1.042	0.840
SoV (coronal and sagittal planes)	0.976	0.998	0.638
Ascending Aorta	0.872	0.780	0.111

Table 5.3 Average measurements of the aortic root at the sinus of Valsalva (SoV) and ascending aorta.
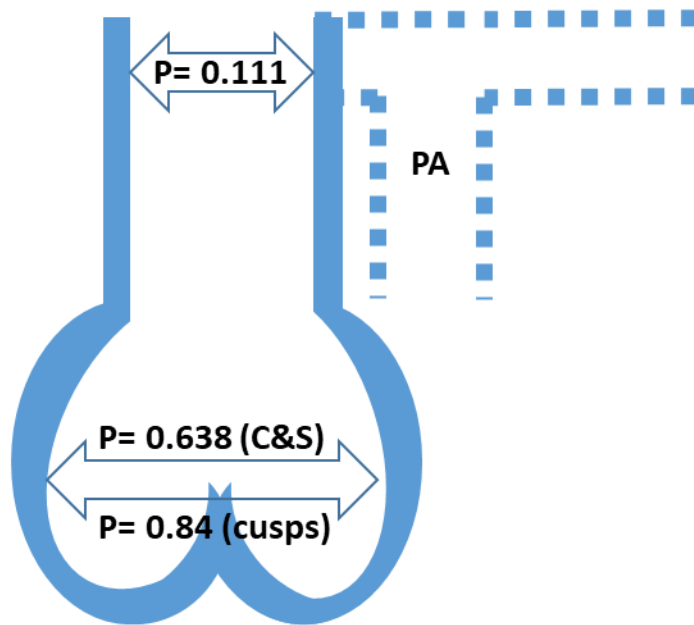
Figure 5.14. Summary of aortic measurements in the *Prkd1* mice.

Measurements of the aortic root were taken at the Sinus of Valsalva. Three measurements were taken across the cusps in the short axis and averaged. Two measurements were taken in the long axis (sagittal and coronal planes) and averaged. P values suggest no significant difference between the *Prkd1+/+* and *Prkd1*Em2/+ mice (two tailed T test, unequal variances).

The diameter of the ascending aorta was measured at the level of the pulmonary artery. There was no significant difference between the *Prkd1+/+* and *Prkd1*Em2/+ mice.

C&S: Coronal and Sagittal views, PA: Pulmonary artery.

Figure 5.15. Box and whisker plots of aortic measurements in the *Prkd1* mice.

Measurements of the aortic root were taken at the Sinus of Valsalva on a transverse image. Two measurements were taken in the long axis (sagittal and coronal planes) and averaged. The diameter of the ascending aorta was measured at the level of the pulmonary artery. P values suggest no significant difference between the *Prkd1*+/+ and *Prkd1*Em2/+ mice.

Whilst only the right ventricular free wall reached significance, it is intriguing that all ventricular walls with the exception of the LV lateral wall, trended towards being thicker in the *Prkd1*[Em2/+] mouse. Isolated RV hypertrophy is usually seen in response to pressure overload as a result of pulmonary hypertension or TV disease. To investigate this further, would require either echocardiographic or real time MRI imaging, to assess the RV and TV function, and to look for evidence of pathology such as emboli in the pulmonary circulation for example. Without a larger sample size and possibly additional dynamic imaging, we cannot confirm that RV hypertrophy is a feature of *Prkd1*[Em2/+] mice. Alternatively, it may be that the whole heart was showing a hypertrophic response. and failure to reach significance in the other ventricular wall measurements was a result of the small sample size, or potentially short

lifespan of the mice not allowing adequate time for a hypertrophic process to occur for long enough to detect a difference.

Previous mouse models have shown that loss of *Prkd1* in the heart results in reduced hypertrophy in response to stress [732]. Therefore, if the mutation in the mouse results in loss of function, we would not expect to see any hypertrophy under normal conditions, or even under conditions of increased work. Given there was no clear evidence of CHD in the mouse hearts, I would not expect the heart to be under any increased workload. Therefore there should not be any difference in ventricular wall thickness between the groups. We could consider if increased hypertrophy might be a result of this mutation causing a gain of function, as demonstrated in mice with activated *Prkd1* [732]. These mice showed dilation of the ventricles, wall thinning and enlarged atria. Histological assessment showed myocyte disarray, myocyte hypertrophy, increased interstitial but only minimal fibrosis [742]. These effects were seen from four weeks of age. Therefore it is likely that if the *Prkd1*$^{Em2/+}$ mice had a gain of function mutation, we might actually see reduced wall thickness given that they were two months old. The consequences of a germline non-conditional gain of function mutation is also unknown. If the mutation in the mouse does not cause a gain of function, another explanation for hypertrophy occurring would be upregulation of the normal *Prkd1* allele.

Simple up or down regulation of *Prkd1* does not easily explain why we might see increased wall thickness in the mouse and the occurrence of hypertrophy might instead indicate other compensating mechanisms are at work. There is evidence that other genes have a similar functional role to *PRKD1*. Prkd1, 2 and 3 can all phosphorylate class II HDACs and facilitate their nuclear export. Loss of *Prkd1* alone does not result in complete loss of HDAC5 nuclear export, and in B cells both *Prkd1* and *Prkd*3 had to be knocked out to prevent phosphorylation and nuclear export of HDAC5 [715, 716]. These other mechanisms may maintain normal hypertrophic response. It is not clear if this would result in hypertrophy under no increased cardiac workload though.

To move forward with these results, a number of questions need answering. The first is whether this increased wall thickness remains significant in a larger population. If increased wall thickness is confirmed to be significant we would then need to consider why it is occurring. We would need to exclude other causes of hypertrophy. This should include some kind of functional assessment and subsequent histology. We should also determine what effect this mutation has by performing protein and functional studies.

Aortic measurements were also taken to determine if the root and ascending aorta were normal, given that dilated aortae have been seen in *Mef2c* knockout mice [793]. There was no significant difference between the *Prkd1*$^{Em2/+}$ and WT mice with respect to aortic measurements.

Whilst there is room for improvement and further studies, these results do provide some useful information about the incidence of CHD in the *Prkd1*Em2 mouse model of the commonest p.G592R mutation. It also suggests that this might not be a gain of function mutation based on the appearance of the heart alone. It would be interesting to perform scans at later ages and with additional afterload on the heart to assess hypertrophic response. These data provides a baseline for such further studies.

5.4 Challenges, Limitations and Future Work

The main limitations of MRI work are the small cohort analysed and a lack of control method of analysis. Lui et al recommend histological confirmation of phenotypes after all other imaging techniques [810]. However, not all agree that histology is the best method for cardiac phenotyping [819] and MRI has been shown to be superior [816].

Whilst every effort was made to ensure measurements were as accurate as possible there will be some error. The precision of measurements was partly limited by the resolution of the scan, but also the nature of some of the heart structures. Because the aortic root can often have a triangular shape, multiple measurements were taken and averaged. The presence of the papillary muscles also makes some measurements more difficult such as the LV internal diameter (transverse) and the lateral LV wall. Real time imaging would also allow us to confirm that all measurements are taken in end diastole.

Although the atrial and ventricular septae were relatively well demarcated on final scans, ASDs and VSDs are often only diagnosed when flow can be demonstrated across the septum, so a static image alone is not sufficient. Real time imaging was not possible due to license restrictions, but the scan resolution should mean that any significant defects would be apparent. Valve defects were the other phenotype we were concerned about visualising, which again ideally require dynamic imaging with Doppler flow.

The MRI protocol we developed is convenient because it allows use of the MRI scanner overnight and at weekends, making good use of potentially redundant scanning time. This is also more cost efficient. It also allows for batch processing and storage of the mice until you reach a point where it is convenient to scan them. However, the question of whether the freeze thaw process we used would distort tissue architecture at a supracellular level to the extent we might produce false positives and negatives remains unanswered. The literature suggests cellular level changes mainly, and I haven't been able to find any specific literature describing the effect we might see on the structure heart at the level of detail I was looking [828-832]. It is likely that changes that do occur, are below the 50μm resolution we used. Ideally, we would have imaged each mouse immediately after death, but this was not possible. Histological analysis afterwards and a control group who were scanned without prior freezing could help answer this question.

To gain high resolution, the image acquisition periods were very long. Given the mice were not preserved in any way, there is likely to be some decomposition occurring during this time. However, multiple scans of the same mouse over a whole weekend did not result in any differences in cardiac structures on sequential scans. This suggests that any decomposition or degradation occurring it not sufficient to be seen on the MRI scans.

Injection of contrast via the umbilical vessels is something that could be investigated further as a post mortem intervention to potentially improve images. I did manage to cannulate the umbilical vessels with the same glass capillary needles that I used for zebrafish injections, so potentially this could be easily done. Whether this would allow adequate perfusion of the heart and great vessels is uncertain. Cannulation of other vessels would be preferable, but we would still need to determine whether contrast could adequately reach the heart and expel blood, and that the pressures required to do this would not result in any damage to cardiac structures.

A number of other interventions would also provide useful information. Given *Prkd1* conditional knockout mice appear to have normal hearts until stressed [732], it would also be interesting to stress the *Prkd1*Em2 mice to see what effect this has and whether this might help inform current clinical recommendations. Given *Prkd1*[Em2/+] and live born *Prkd1*[Em2/Em2] mice are apparently normal, it would not be necessary to generate an additional conditional model. ECG analysis of these mice would also be useful given the role of *Prkd1* in contractility of the heart [689, 741, 774].

5.5 Conclusions

In addition to its role in CHD and ectodermal dysplasia, *PRKD1* is likely to be an important gene target in managing pathological hypertrophy in the stressed heart, angiogenesis and in preventing ischaemic damage. This is important given the prevalence of heart failure within the population and its associated morbidity

and mortality, especially given the ageing population. Potential treatment strategies may also be informed by monitoring cardiac function in individuals with CHD and Ectodermal Defects. These results suggest that there is a low incidence of CHD in the *Prkd1*^{Em2(IMPC)Wtsi} heterozygous mouse, and that it does not demonstrate hypertrophy under normal conditions. This may provide a useful platform for further research into the mechanisms of this mutation and methods for treating cardiac hypertrophy.

Chapter 6 The Role of *CDK13* in Congenital Heart Disease

6.1 Introduction

We identified *CDK13* as a novel gene causing S-CHD, and since then, further publications have confirmed that individuals with mutations in this gene have a recognisable phenotype [12, 489, 833-836]. The syndrome is referred to as CHD, dysmorphic facial features and intellectual developmental disorder (CHDFIDD), or *CDK13*-related disorder. Just under 40% of individuals exhibit CHD, and exactly how variants in this gene result in aberrant cardiac development remains to be understood. Over 80% of the pathogenic mutations lie within the kinase domain. Clinical recommendations for care of these individuals is now available, and there is an active social media group for the families who are affected.

This chapter focuses on the genotypes and phenotypes of affected individuals, and phenotyping of a *Cdk13* knockout mouse [12].

6.1.1 The Structure and Function of *CDK13*

CDK13 is a member of the cyclin-dependant kinase (CDK) family, which comprises of 21 ATP-dependant serine-threonine kinases. The proteins encoded by these genes are involved in transcription and cell cycle control and are required for the normal function of RNA Polymerase II.

Five human CDKs have been described with a role in control of transcription and are listed below. Of these five, only *CDK8* and now *CDK13* have been shown to cause syndromic neurodevelopmental disability in humans. Several have a role in cancer, and *CDK12* also contributes to the pathogenesis of Myotonic Dystrophy [837-839].

- *CDK7* and its subunit Cyclin H are important in cell cycle control and transcription. It forms part of the general transcription factor TFIIH [840]. *CDK7* is involved in *initiation* of transcription and phosphorylates Serine 7 in RNA polymerase II, and acts as a CDK activating kinase [841].
- *CDK8* and its subunit Cyclin C form part of the mediator kinase module, which regulates RNA polymerase II [842]. De novo missense mutations in CDK8 have recently been shown to cause syndromic neurodevelopmental disability [843].
- CDK9 and its subunits Cyclin T1/T2 form part of the positive transcription elongation factor P-TEFb [844]. CDK9 is also needed to prevent promotor proximal arrest and is located near the transcription start site [845].
- *CDK12* and its subunit Cyclin K (CCNK) [846, 847].
- *CDK13* and its subunit Cyclin K (CCNK) [846].

Although *CDK13* is the subject of this chapter, it is important to note that it is structurally similar to *CDK12* and shares the same subunit, Cyclin K (CCNK). Despite their similarities, they do not appear to be redundant and instead regulate different sets of genes [848]. Both *CDK12* and *CDK13* are thought to exhibit some influence over RNA polymerase II, required for transcription of protein coding genes in eukaryotes [849-852].

To understand the effect of *CDK13* mutations, we need to understand how the protein interacts with CCNK and RNA polymerase II in more detail.

6.1.1.1 RNA Polymerase II and the transcription process

The RNA polymerase II complex is required to transcribe DNA into RNA. This process is controlled by modification of the amino acid residues in the heptad repeats ($Y_1S_2P_3T_4S_5P_6S_7$) found within the largest subunit of the polymerase, Rpb1. A number of the CDKs, including *CDK13*, are required for phosphorylation at these residues. This process is required to facilitate normal elongation of

transcripts and production of mature mRNA. An illustration of the CDK13-Cyclin K complex is shown in figure 6.1, and RNA polymerase II in figure 6.2.
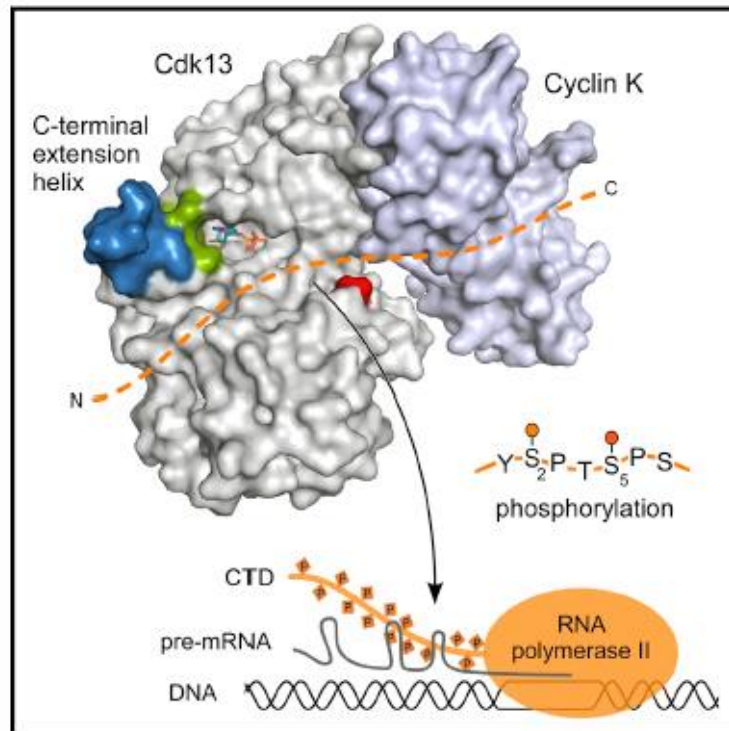
Figure 6.1 CDK13 Cyclin K complex.

Reproduced with permission from Elsevir [848]. Graphical abstract showing the Cdk13 (grey) Cyclin K (lilac) complex and its interaction with RNA polymerase II**.** CDK13 may interact with RNA polymerase II, due to the attraction of the negatively charged polymerase C-terminal domain, to the positively charged closely associated basic patch and ATP binding site.

Blue section: Polybasic cluster, Green: DCHEL motif. CTD: C terminal domain.

Orange and Blue molecular interactions: Interaction between the C-terminal extension helix and the kinase domain and the bound nucleotide. ADP binds here with the DCHEL motif.

Red: Phosphorylated T871 residue in the T-loop of the kinase.

The hepta-sequence in the CTD of RNA polymerase II subunit Rpb1 is shown, and phosphorylation sites are denoted by orange markers. CDK proteins including CDK13 are responsible for this phosphorylation process.

The RNA Polymerase II core is made up of ten subunits, with some mobile elements (Figure 6.2). It is joined by the sub-complex Rbp4/7, to form the complete enzymatic complex and initiate transcription [853]. RPB1 and RPB2 are the largest subunits of the complex and the active site of RNA polymerase II lies between them in the large central cleft.
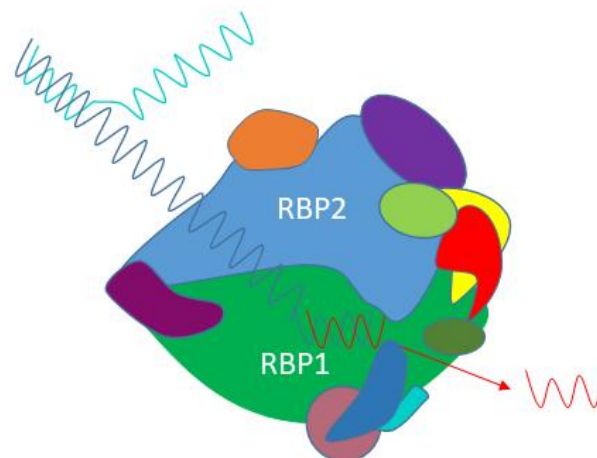
Figure 6.2 RNA polymerase II subunit composition and assembly.

Transcription process is shown. The turquoise and blue lines represent the two strands of DNA which are unwound. Transcription occurs after the template strand is taken into the active site cleft, between the Rpb1 and Rpb2 subunits. The RNA strand produced is shown as a red line.

Red: Rpb3, Mauve: Rpb4, Purple: Rpb5, Turquoise: Rpb6, Dark blue: Rpb7 Dark green: Rpb8, Orange:Rpb9, Purple: Rpb10, Yellow: Rpb11, Light green:Rpb12

Open chromatin structures mean that the pre-initiation complex is able to bind and allow transcription to take place. This includes unphosphorylated RNA polymerase II [848]. The DNA helix is unwound to produce a single strand of DNA which acts as the template (Figure 6.2). It is brought into the active site cleft, which can be opened and closed. Transcription can then begin at the transcription start site [854, 855]. An elongation complex forms after a number

of false starts, and a DNA-RNA hybrid is produced which is about eight to nine base pairs long. This elongation complex is responsible for the addition of the next nucleotide to the RNA strand as the DNA template is ratcheted through the machinery [856]. A complimentary RNA transcript is produced, attached to the DNA strand. The RNA strand is then stripped away and leaves the complex through an exit groove.

A pore that runs from the protein surface to the active site, allows entry of the nucleotide triphosphates (NTPs). The initial binding of NTPs requires association of a magnesium ion with three conserved aspartate residues in the RNA polymerase subunits. The NTP is rotated into position at the catalytic A site, to be paired with the strand of DNA. This requires a second magnesium ion, associating with three aspartate residues [853].

6.1.1.2 Structure of the RPB1 subunit and the Heptad Repeats

RPB1 is the largest subunit of RNA polymerase II (Figure 6.2). It acts as a main site for binding of a number of factors required during transcription and mRNA processing. The ratchet mechanism that moves the DNA-RNA through RNA polymerase II is provided by a bridge which originates from RPB1 and crosses the cleft near the catalytic site, and bends and straightens with each addition of a nucleotide [853].

Elongation of the RNA molecule is dependent on phosphorylation of the C-terminal domain (CTD) of RPB1 (Figure 6.1), which consists of the heptad repeats. This series of repeats is thought to provide a platform for binding of transcription factors [857]. Importantly, phosphorylation of serine (Ser) 2, 5 and 7 is an essential part of the transcription process [848, 858-861]. CDK proteins, including CDK13, are required to phosphorylate these residues in the heptad repeats.

Transcription starts when Ser 7 is phosphorylated. This phosphorylation also means that the CTD can be modified further and levels of phosphorylated Ser 7 are high throughout transcription [862]. The transcription process is paused when the polymerase has reached 50bp downstream from the transcription start site, but continues when Ser 5 (and possibly Ser 7) are phosphorylated. This allows elongation of the transcripts [858]. Levels of phosphorylated Ser 5 and 2 increase in opposite directions, with phosphorylation of Ser 2 increasing as RNA polymerase II progresses towards the transcription termination site as a result of the action of RNA processing factors working at the 3' end [848, 863]. Phosphorylation status is reset at the end of the process by phosphatases, which remove phosphorylation at these sites. CDK13 plays an important role in phosphorylation of these serine residues, allowing transcription to continue.

6.1.1.3 Structure of CDK13 and binding with Cyclin K

CDK13 is a long protein (1512 a.a.) and although it is larger than other CDKs, it demonstrates the usual kinase fold confirmation. It is made up of a kinase domain, with an N-terminal arm on one side, and a C-terminal arm on the other. It is activated when it binds with Cyclin K (CCNK) [848, 864, 865]. Figure 6.3 shows a number of important domains within the protein. Understanding the structure and important sites helps us interpret the effects of mutations found within these domains.
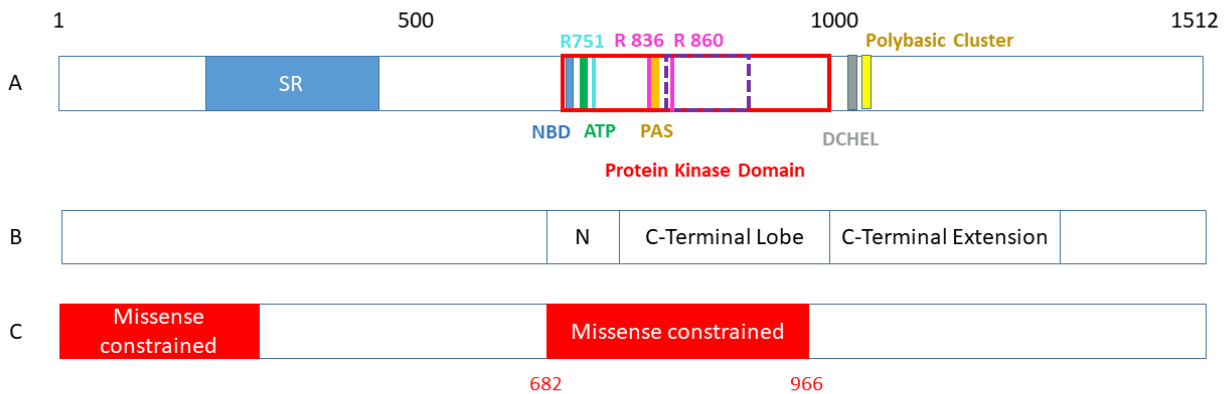
Figure 6.3 A Schematic of CDK13 structure.

A. Some important domains and residues in CDK13:

- Blue box SR: Serine Rich domain a.a. 200-435.
- Red box: protein kinase domain a.a. 705-998.
- Purple dashed box: Activation loop a.a. 854-878.
- Blue line NBD: Nucleotide Binding Domain 711-719. Valine at position 719 is conserved in human CDKs.
- Green line ATP: ATP binding site Lysine at residue 734 (conserved in human CDKs).
- Turquoise line R751: Part of the PITAIRE motif, this arginine forms electrostatic connections with 105KVEE of Cyclin K.
- Orange line PAS: Proton acceptor site, Aspartate at residue 837 (conserved in human CDKs).
- Pink Line: Phospho-threonine co-ordination. R836 and R860 are canonical arginines that form ionic interactions with the phosphate group.
- Grey box: DCHEL sequence including HE motif (H1018, E1019). H1018 and E1019 form water mediated contacts with ADP, D1016 forms interactions with residues in the N and C terminal kinase lobes. The HE region forms water mediated bonds that allow interaction of the complex with the ribose of the bound ATP, as seen in the transcription elongation kinases [866, 867].
- Yellow box: Polybasic Cluster 1023-1029, forms a helix with the ^{1016}DCHEL sequence, which lies adjacent to the active kinase centre.

B. Corresponding main domains of the protein. N. N-terminal lobe a.a. 695-794, C-terminal lobe 795-998, C-terminal extension 999-1302

C. Areas of missense constraint [768]. Areas identified as missense constrained were between a.a. 1-275 (exhibited 8 missense mutations, compared with the expected 109 by chance), and between a.a. 682-966 (exhibited 27 missense mutations, compared with the expected 96.5).

In addition to the important domains outlined in figure 6.1 and 6.3, other important features of CKD13 are outlined below [848].

- The [746]PITAIRE helix (or αC helix) is involved in binding with Cyclin K.
- [855]DFG motif is at the start of the activation segment. Substrates can access the active site of CDK13 when this motif and the activation segment of CDK13 assume the relevant conformation.
- L858. This follows the DGF motif. This suggests that CDK13 is likely to use a serine phosphoacceptor motif (SP), rather than a threonine phosphoacceptor motif for phosphorylation. Many serine-threonine kinases show a bias towards either serine or threonine as their phosphoacceptor of choice (SP or TP) [868].
- C-terminal extension unit (aa 999-1032). This associates with the kinase domain and follows the canonical cyclin box.
- The [1016]DCHEL sequence. This is found within the C terminal extension. There is a histidine (H) at position 1018 and a glutamate (E) at 1019. These two amino acids bind CDK13 to the ADP/ribose of the bound nucleotide using water mediated contacts. Additional support for this bond between the HE motif and the kinase domain are provided by a bond between D1016 and L1020, which interact with the N and C-terminal lobes of the kinase domain.
- The [1023]KKRRRQK sequence /polybasic cluster. This forms a helix with the [1016]DCHEL sequence, which lies adjacent to the active kinase centre.

For CDK13 to function, it must bind with Cyclin K (CCNK). Cyclin K is 580 amino acids long. Figure 6.4 displays some of the important domains in Cyclin K.

Figure 6.4 Representation of Cyclin K.

Missense constrained areas are found between residues 1-282 (observed 17 mutations, expected 79) [768]. The KVEE at residue 105-108 is marked by a red line, it is important for binding with CDK13.

Greifenberg et al. characterised the binding sites between CDK13 and CCNK [848]. All the residues that are involved in this interaction are in the N-terminal lobe of the kinase domain (Table 6.1).

CDK13 interaction site	Cyclin K interaction site
αC helix	αH3 helix and adjacent loop formed from F101 to K112
αC helix	α5 helix and adjacent M141-Q156 loop
D969 and K699	α5 helix and N-terminus of Cyclin K
I770 and T772 in the β4 helix	α5 helix

Table 6.1 CDK13 Cyclin K complex.

The paired interaction sites between CDK13 and Cyclin K [848]. αC helix of CDK13 is amino acids 747-758 and includes part of the PITAIRE motif.

Syndromic intellectual disability has been reported in individuals with deletions and a mutation in CCNK. This is thought to be due to haploinsufficiency [869].

6.1.2 Similarities and differences between members of the CDK family

CDK12 and 13 are much larger proteins than CDK 7, 8 and 9, but there are many conserved residues shared between them. These include the PITAIRE helix, HE motifs and polybasic regions. Table 6.2 displays some of the differences and similarities between these proteins.

Flavopiridol inhibits transcription elongation kinases [867], but interestingly it inhibits CDK7 to a higher extent than CDK13. Greifenberg et al. proposed that this was due to competition at the HE site between flavopiridol, and this site already being occupied as part of the normal binding processes that occur at this motif (steric hindrance) [848] . The C terminal extension domain is much larger in CDK12 and 13 compared to CDK7. This suggests that there are likely to be differences in function, despite structural similar within the CDK family.

The most closely related of the CDKs are *CDK12* and *CDK13*, which share 43% sequence identity. They are thought to have arisen from a gene duplication. They both bind Cyclin K, but form separate complexes [846]. They both have a central kinase domain, but differ from CDK 7, 8 and 9 as they contain expanded serine-arginine rich motifs in the N-terminal regions [848]. These serine-arginine rich motifs are found in the SR protein kinase family, which are usually involved in protein interactions, RNA processing and in regulation of splicing [865, 870]. This has been confirmed as CDK13 has been shown to interact with SRSF1, a splicing factor in this family [871]. SR proteins may attach to pre-mRNA and prevent it from attaching to the coding DNA strand [872]. This is thought to have a stabilising effect on the genome. CDK13 tends to be found in areas where splicing factors are stored, called speckles, in the nucleus [865]. It interacts with another splicing factor SRSF1 in HIV [871] and protein p32 which is required in splicing regulation [873].

CDK12 and 13 form fewer salt bridges than the other CDKs. These salt bridges are a marker of activation and stabilise loops within the protein [874]. The bridges form between the phosphorylated threonine (pT) of the T loop, and arginine residues in other loop sections of the kinase (Figure 6.1). In addition to stabilising loops, the bridges produce a twist in the DFGxxR motif, re-orientate the αc helix that contains the PITAIRE motif, and allow correct positioning of the HRD motif, which contains a catalytic aspartate. CDK12 and 13 form two of the three bridges, but lack the bridge between the PITAIRE motif and the phosphorylated threonine [848]. Instead it forms between the KVEE motif of Cyclin K and the arginine (Figure 6.4).

Similarities between CDK12 and CDK13	Similarities with other CDKs	Differences from other CDKs
CDK12 and CDK13 have 43% shared sequence identity.	CDK13 exhibits the typical folding pattern of CDK proteins. CDK13 N terminal lobe shares 46.9% sequence identity with CDK9, CDK13 C terminal lobe shares 41.4% sequence identity with CDK2.	Expanded serine-arginine (SR) repeats in the N terminal domain sets CDK12 and 13 apart from CDK7, 8 and 9.
Similar position of the cyclin subunit and kinase domain relative to each other. In CDK12 and 13, only 2 of the 3 canonical arginines are involved in maintaining position, the third forms a salt bridge with the second glutamate of ^{105}KVEE in Cyclin K.	Position of the cyclin subunit with respect to the kinase domain is similar to CDK9/cyclin T1 and CDK12/Cyclin K. The two subunits are "twisted" in comparison to CDK2/cyclin A. This might relate to the absence of the third salt bridge seen in other CDKs.	Position of the cyclin subunit with respect to the kinase domain is rotated further towards an open conformation in CDK2/cyclin A. All three canonical arginines are involved in the link between the cyclin subunit and the kinase domain. These three salt bridges stabilise the loop portions of the kinase, which is indicative of CDK activation [874].
The arginine residue of the PITAIRE motif is the only intermolecular link with CCNK **[866]**		In Cdk9/CyclinT1, the arginine of the PITAIRE motif forms intermolecular links with

		both part of the T-loop and the KVEE motif [875]. In Cdk2/Cyclin A, all 3 canonical arginines form intermolecular links with sites in Cyclin A [876].

Table 6.2 Structure of the CDK family proteins.

Differences and similarities between the CDKs [848].

6.1.3 Function of CDK13-Cyclin K Complex in transcription

CDK13 plays an important part in transcription and splicing processes. CDK13 binds with CCNK and is then able to phosphorylate Ser 2 and Ser 5 in the heptad repeats of RPB1, allowing transcription to continue [848, 864, 865]. Greifenberg et al. demonstrated this by examining the activity of the CDK13-CCNK complex on a synthetic RNA polymerase II CTD, which included three heptad repeats [848]. The complex displayed the highest activity with pre-phosphorylated Ser7, and much lower activity on pre-phosphorylated Thr4, Ser2 and Tyr1. There was no activity with phosphorylated Serine 5 or with a Lysine substitution at position 7. The same synthetic peptide produced minimal activity if there was no pre-phosphorylation within the heptad repeats. These results suggest that CDK13-CCNK phosphorylates Ser 2 and 5, and does so most efficiently when Ser 7 is pre-phosphorylated. Phosphorylation mark was preferentially attached to the N-terminal end of the CTD, to a pre-phosphorylated Ser7 site by CDK13-CCNK (14).

Kinase assays revealed that both native and recombinant CDK13-CCNK and CDK12-CCNK complexes set Ser2 and Ser5 as the site of phosphorylation, showing an overlap of function [848]. No Ser7 phosphorylation marks were set. In contrast, Cdk9 strongly phosphorylated Ser5, minimally phosphorylated Ser7

and did not phosphorylate Ser2 sites indicating the different CDKs also act on different residues. This overlap in function of CDK12 and CDK13 is important when considering how variants in one of these two genes might cause abnormalities.

6.1.4 Differences in Gene Regulation by *CDK12* and *CDK13*

Given the structural similarities between CDK12 and CDK13 and that they share the same cyclin partner, it is important to consider whether there is any further evidence of redundancy. Research has shown that CDK12 and CDK13-CCNK complexes regulate expression of around 1000 genes, but mainly have an effect on distinct gene sets [848]. Comparison between RNA expression profiles in *Cdk12* and *Cdk13* null HCT116 cells showed that *CDK12* regulates genes involved in DNA repair and *CDK13* is involved with processes including growth signalling, extracellular pathways, including protein kinase signalling and cell growth regulation [848]. These results are supported by Liang et al., who showed *CDK13* knockdown (also in HCT116 cells) affected expression of snoRNA genes, and *CDK12* knockdown affected genes linked with DNA damage [877]. Given *CDK12* regulates genes such as *BRCA1, FANC1* and *FANC2*, it is not surprising that much of the research around this gene focuses on its role in cancer.

There is some cross over between genes regulated by *CDK12* and *CDK13* and 62 genes were downregulated by both. They both appear to play a part in neuronal development (Table 6.3) [848, 868, 878].

Changes in gene regulation	Genes
Downregulated genes by depletion of both Cdk12 and Cdk13	62 genes. Enriched for basic biological processes
Downregulated genes with depletion of Cdk12 only	Enriched for DNA replication and repair processes
Downregulated genes with depletion of Cdk13 only	Enriched for extracellular and growth signalling pathways
Upregulated genes by depletion of both Cdk12 and Cdk13	49 genes. Enriched for different biological processes
Upregulated genes with depletion of Cdk12 only	Enriched for neuron development and other biological processes (different to Cdk13)
Upregulated genes with depletion of Cdk13 only	Enriched for neuron development and other biological processes (different to Cdk12)

Table 6.3 Effect of Cdk12 and Cdk13 on RNA expression.

Summary of changes in RNA expression in HCT116 cells when *Cdk12* and *Cdk13* were downregulated [848]. It appears that *Cdk12* is required for DNA damage response genes including the PARP1/BRACA pathway [846, 848, 877, 879-883] and both are important in neuron development.

The regulation of different genes by *Cdk12* and *Cdk13* is consistent with knockout studies in mice. The *Cdk12* null mouse succumbs to early embryonic lethality just after implantation [884], whereas demise of *Cdk13* null mice is much later [885]. This is perhaps not surprising given that there is reduced expression of genes involved in DNA replication and repair when *CDK12* is lost.

6.1.5 CDK13 Expression

CDK13 expression is fairly ubiquitous in the human and includes adipose tissue, brain, thymus and uterus. There is low expression in the adult human heart including the LV, LA, MV, PV and TV [886]. Moderate expression was seen in the coronary arteries of human post-mortem samples [887]. Specifically in the human embryo, low expression was seen in the liver, with moderate expression in the fore/hindbrain, testes, ovary, kidney and heart between 4 and 20 weeks post conception (Carnegie stage 10 – fetal stage, Theiler stage 10 – fetal period in mouse) [888]. Low to medium expression is seen in the embryonic mouse heart E11-18 (Theiler stage 18 to 26, equivalent is Carnegie stage 13 - fetal period) [886]. Expression in the developing human heart adds credibility to it being a CHD causing gene.

6.1.6 Animal Models

Cdk13 deletion mutant mouse lines are available and there is already some information about the cardiac phenotype. A modified Cdk13 mouse line was established using an insertion cassette, which can be converted to several different alleles as shown in fig 6.5.
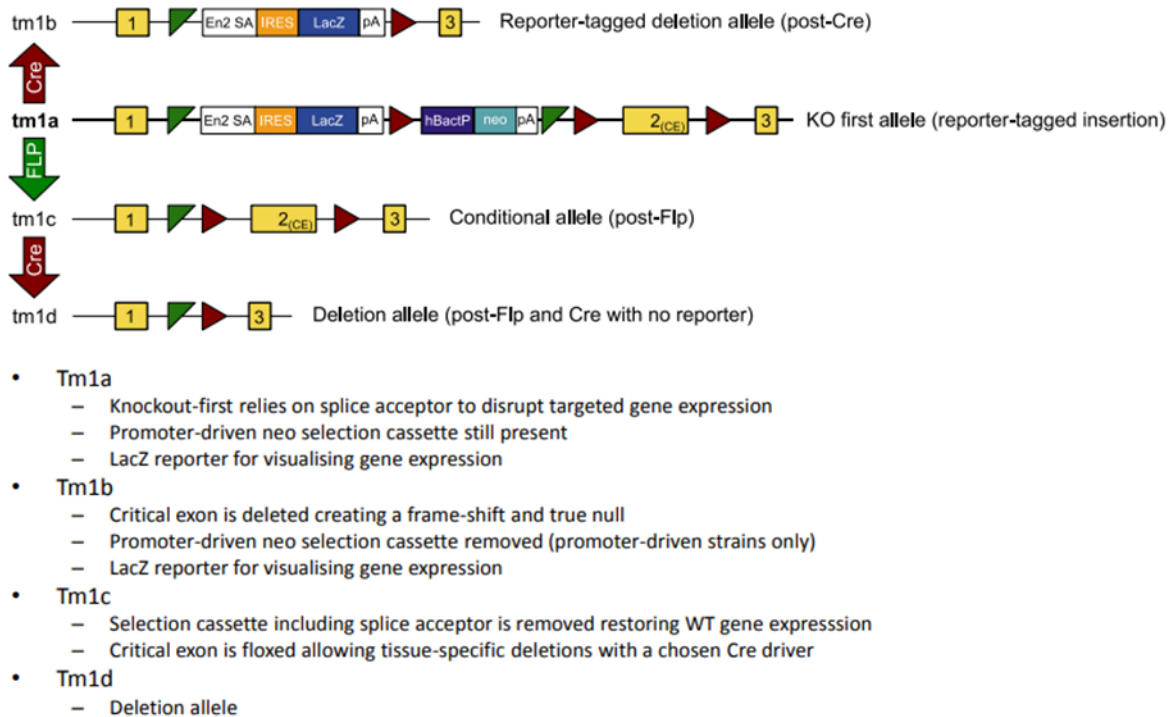
Figure 6.5 Mouse allele models.

Generation of different mouse models used in Cdk13 animal studies. Reproduced with permission from the allele conversion guide, Ed Ryder, https://www.infrafrontier.eu/resources-and-services/access-emma-mouse-resources/eucommkomp-csd-allele-conversion-service

https://www.infrafrontier.eu/sites/infrafrontier.eu/files/upload/public/pdf/Resources%20and%20Services/eucomm_komp-csd_allele_conversion_guide_v3a_2016.pdf.

Yellow: Exons

CRE: cre recombinase, FLP: Flippase recombinase, En2SA: engrailed 2 gene splice acceptor, hBactP: human beta-actin promoter, IRES: IRES-lacZ sequence

LacZ: LacZ cassette, Neo: neomycin resistance gene, pA: polyadenylation sequence.

Novakova et al. examined the *CDK13*tm1a and tm1d alleles [885]. This is the same mouse line I have used, but I have studied the tm1b allele. The tm1a allele is the first step in producing the knockout mouse, and the promotor driven neo selection cassette is present. The tm1b allele no longer contains the selection cassette and the critical exon has been removed. The tm1d allele is the deletion allele (post Flp and Cre with no reporter).

Novakova et al. demonstrated that the *Cdk13*tm1a allele produces a CDK13 protein that is truncated around exon 2, with no kinase or C-terminal domain. Immunoblot detected a minor WT CDK13 protein as well as the truncated protein in *Cdk13*[tm1a/tm1a]. Despite reduced expression of *Cdk13*, there was no discernible effect on phosphorylation of Ser2 or Ser5 in the CTD of RNA polymerase II. They proposed that this represents a hypomorphic allele and suggested that the protein may have some residual function. *Cdk13*tm1d mice were then produced, in which exons 3 and 4 were deleted. CDK13 protein levels in *Cdk13*[tm1d/+] embryonic brain tissue was greatly reduced, and absent in *Cdk13*[tm1d/tm1d] mice. There was also no clear downregulation of Ser2 phosphorylation in brain tissue in this mouse line.

We might expect to see reduced phosphorylation of Ser 2 and 5 with loss of functioning CDK13. However previous experiments had shown that downregulation of CDK13 in cell lines also had minimal to no effect on levels of Ser 2 phosphorylation [846, 848]. Novakova et al. suggested that *CDK13* might phosphorylate RNA polymerase II for specific sets of genes, in specific organs at specific times in development, so that there was no overall difference in phosphorylation levels at Ser2 or Ser5 in the mouse [885]. Serine 5 is known to be phosphorylated by other CDKS in addition to CDK13 [846].

Expression studies confirmed CDK13 expression in the mouse heart. At E14.5, CDK13 protein expression was highest in the brain, followed by the lung, kidney and heart in the tm1a line. In the adult, CDK13 protein was expressed in the retina, testes, ovaries and uterus, the gall bladder, bladder, renal pelvis and

thyroid. In the adult heart, expression was greater in the ventricle than the atria in tm1a mice [885].

Embryonic lethality in *Cdk13*^{tm1a/tm1a} embryos was observed by E16.5, with live embryos found at E15.5 [885]. Novakova et al. reported an increasing number of absorbed embryos from E13.5-16.5, although the results seem to suggest that the numbers are relatively stable and then decrease (13% at E12.5, 14% at E13.5, 5% at E14.5 and 7% at E15.5). In contrast, the *Cdk13*^{tm1d/tm1d} null mice appeared to show earlier mortality, with a higher proportion of empty decidua seen at E12.5 [885].

Cdk13^{tm1a/tm1a} mice were on average a day behind with their development than the rest of the litter [885]. All *Cdk13*^{tm1d/tm1d} mice exhibited growth retardation, which was more severe than the tm1a line. There is always some variability of developmental stage within a litter [889], however the growth retarded *Cdk13*^{tm1a/tm1a} mice also had nuchal oedema. Nuchal oedema can be seen in association with a number of abnormalities, including reduced cardiac function, suggesting that there may be an organic cause for the delayed development [885].

Multiple developmental defects were seen in both the *Cdk13*^{tm1a/tm1a} and *Cdk13*^{tm1d/tm1d}, including heart, renal and cleft abnormalities. *Cdk13*^{tm1a/tm1a} mice had smaller less well developed lungs and livers, and brain development lagged behind the WT embryos. Cleft palate was also seen at E15.5 and renal tissue exhibited a reduced number of tubules and glomeruli. Orofacial clefting rates were higher in the *Cdk13*^{tm1d/tm1d} mice, and they exhibited midface hypoplasia. *Cdk13*^{tm1d/tm1d} embryos also had smaller livers and displayed "abrogated liver parts arrangement". Hypervascularisation and pericardial oedema were also noted in the null mice. Interestingly the same abnormalities were seen in a few *Cdk13*^{tm1a/+} mice too.

Micro CT was used to examine the mice for cardiovascular abnormalities. Ventricular wall thickness was thinner in *Cdk13*^{tm1a/tm1a} mice compared to WT, and the heart wall was less compact with reduction of myocardium. Histological examination revealed fewer layers of cardiomyocytes and reduced myosin expression in the ventricular myocardium in E14.5 *Cdk13*^{tm1a/tm1a} mice compared to WT.

Haemodynamic studies were performed at E14.5 and E15.5 in WT, heterozygous and *Cdk13*^{tm1a/tm1a} embryos. Doppler imaging of embryos revealed normal flow velocity in the dorsal aorta at E14.5. At E15.5 normal flow was seen in only 30% of the null embryos. In the other 70%, cardiac function was reported to be declined, although the authors comment that "standard measurements were not possible due to irregular or spare heart beating (11/16)" and that those with preserved blood flow were the result of variability in embryo size. Presumably by this, the authors are suggesting that these embryos were larger and therefore had higher velocities, but did not necessarily have normal cardiac function.

The *Cdk13*^{tm1d/tm1d} embryos also exhibited smaller hearts and had what the authors described as "severe ventricular deficiency". It is not clear from the images exactly what the defect is. In the *Cdk13*^{tm1d/tm1d} embryos, pericardial oedema was seen in 82% of E13.5 embryos, and 44% of E14.5 embryos. Only 19% of *Cdk13*^{tm1a/tm1a} embryos had pericardial oedema at E14.5. 2% of heterozygous embryos, but no WT embryos displayed pericardial oedema. They did not report any gross structural cardiac defects as part of their study.

These mouse models provide evidence that *Cdk13* is required for normal embryonic development, including the heart. Defects were more severe in the *Cdk13*tm1d mouse than the hypomorphic *Cdk13*tm1a mouse, and interestingly some abnormalities were seen in a proportion of heterozygous mice. Novakova et al. suggested that the primary abnormality may have been failure of normal perfusion as a result of heart defects, which then led to the abnormalities seen in the other organs, especially those with higher levels of *Cdk13* expression

[885]. The authors have based their conclusion that the null mice died of heart failure based on the thinner myocardium, oedema and reduced velocities. It is not clear if the abnormal heart rhythms they describe referred to as "irregular or spare heart beating" are the last stages of a failing heart or a primary conduction problem. The absence of reports on structural defects in the hearts means that we can add more information here.

6.2 Aims

The aims of this chapter are:

- To describe the phenotype and genotype of individuals with mutations in *CDK13* reported by Sifrim et al. [12], and those reported in the literature subsequently.

- To assess *Cdk13*$^{tm1b(EUCOMM)Hmgu}$ mice to determine the effects of loss of *Cdk13* on embryonic survival and formation of the heart.

- To consider the mechanism by which mutations in *CDK13* cause abnormalities in humans.

6.3 Results and Discussion

6.3.1 Individuals reported with mutations in *CDK13*

Six individuals with S-CHD were found to have de novo missense mutations in *CDK13* using WES [12]. The mutations clustered together in the serine/threonine protein kinase domain. Three individuals had an identical mutation; p.Asn842Ser. A seventh individual has this same mutation, but corresponding parental samples were not available to determine if this was de novo or inherited.

These 7 patients have recognisable facial gestalt. This includes a broad forehead, arched eyebrows, ptosis, a broad nasal root and bridge (Figure 6.6). All have septal defects (VSD n=2, ASD n=5) and two have abnormalities of the pulmonary valve. Additional shared clinical features include significant developmental delay in all, and intellectual disability. Mild-moderate microcephaly and structural brain abnormalities (n=4) including agenesis of the corpus callosum (n=2) were seen. Clinodactyly was a common feature (n=6/7). Joint hypermobility (n=3) and hypotonia (n=2) were also present in a few individuals.

Figure 6.6 Clinical features seen with missense mutations in *CDK13*.

Reproduced from with permission from Nature Publishing Group [12]. Photographs of affected children and phenotype summary of probands carrying missense mutations in *CDK13*. Colours indicate the number of times a certain phenotype was observed in individuals carrying a *de novo* mutation in *CDK13*. We were not able to obtain consent to publish photographs for two probands (258830 and 270818).

Based on homology modelling using the CDK12 structure which was available at the time, the mutations were predicted to impair ATP and magnesium binding which is required for normal enzyme activity, or interaction between CDK13 and CCNK. Given the clustering of missense mutations, gain of function is a

possibility, but would require functional assessment to confirm this. The crystal structure produced by Greifenberg et al., shows that asparagine at 842 is part of the β7 helix and just adjacent to the HRD motif, which contains the catalytic aspartate [848]. This is the most commonly mutated site.

6.3.2 The Phenotypic Spectrum of CHDFIDD/*CDK13-related disorder*

Since the initial publication of the first seven individuals with CHD and mutations in *CDK13*, a total of 44 individuals have been reported [12, 489, 833-836, 890]. The syndrome is referred to as CHDFIDD (CHD, dysmorphic facial features, and intellectual developmental disorder), and *CDK13*-related disorder. All were de novo mutations, with the exception of one mosaic (ID: DDD259007) and five individuals where inheritance was uncertain. 82% of individuals have missense mutations within the protein kinase domain.

The main features are characteristic facies, developmental delay, structural brain abnormalities and CHD. Figure 6.7 gives more detail about the clinical phenotypes and the frequencies of abnormalities. Structural abnormalities are reported in a number of systems including skeletal, palate, eye, ear and renal.

CDK13 Clinical Phenotype

CNS
Structural brain abnormalities (absent/hypoplastic corpus callosum, periventricular leukomalacia/gliosis, Chiari malformation
Micro (and macro)cephaly

Spinal cord syrinx
Seizures

Neurodevelopmental
Developmental Delay (motor, speech)
Intellectual disability (mild-moderate)

Autism/ADHD

Facial Features
Hypertelorism, telecanthus, USPFs, blepharophimosis, flat midface, broad nasal bridge and tip, small mouth, thin upper lip, low set posteriorly rotated ears

Cleft, Peg teeth

Renal
Duplicated/dilated collecting systems
Fused renal ectopia

Heart
ASD, VSD, PV and PA abnormalities

TOF, EA, LVNC, BAV AS

Feeding difficulties
Gastro-oesophageal reflux
Velopharyngeal insufficiency

Prenatal
Mild IUGR, low birth weight

More severe IUGR

Musculoskeletal
Clinodactyly

Mild short stature
Scoliosis
Sacral abnormalities

Sensory
Squint
SNHL

 Figure 6.7 Phenotypes reported in individuals with mutations in *CDK13*.

More common features are contained in the red sections, and less common in the yellow. All individuals have developmental delay (motor and speech) and all have characteristic facial features. 96% have intellectual disability and 48% have a structural brain abnormality. Feeding difficulties are seen in 70% [891]. Typically there is mild to moderate intellectual disability and characteristic facies. Septal defects predominate in the heart.

ADHD: Attention Deficit Hyperactivity Disorder, IUGR: Intrauterine growth retardation, SNHL: Sensorineural hearing loss.

The DDD project information about individuals with *CDK13* mutations in their cohort reveals that average gestation at delivery was 37.7 weeks and the birth weights tended to be lower than average (z score -0.48)[522]. The individuals remained small with height, weight and head circumference z scores -2.01, -1.05 and -1.67 respectively [496].

Cdk13 is expressed in the lungs, kidney, brain and heart of the E14.5 mouse. This and the known role of *CDK13* in regulating genes in neurone development is supportive of *CDK13* mutations resulting in NDD, CHD, renal and structural brain malformations.

Now that a clinical phenotype and facial gestalt have been defined, we might be able to identify further individuals for testing. Genotype rather than phenotype driven testing might also identify further individuals who do not fit the current phenotypic model and help expand the phenotype further.

6.3.3 CHD in CHDFIDD/*CDK13*-related Disorder

38.6% of the reported individuals have CHD [12, 489, 833-836]. Septal defects predominate, along with abnormalities of the PV and PA. The specific CHD phenotypes are recorded in table 6.4. The small number of individuals described means that it is difficult to be sure of the full cardiac phenotype at this time. 10 of the 16 individuals with CHD have an ASD. Abnormalities of the PV and PA are seen in four individuals, and two have VSDs. There is a single individual each with EA, TOF and BAV with LVNC. TOF is driven by abnormal septation, and includes abnormalities of the pulmonary valve and a VSD, and so shares some common features with the more commonly reported septal and pulmonary abnormalities.

Individual ID	Sex	Mutation	Inheritance and consequence	Domain	CHD	Original publication
265645[1], DDDP101673[2] 7[3]	F	p.N842S	de novo missense variant	PK	+VSD	Sifrim et al
265813[1], DDDP104368[2] 8[3]	F	p.N842S c.2525A>G	de novo missense variant	PK	+ ASD and PV abnormality	Sifrim et al
262889[1], DDDP111394[2] 2[3]	M	p.G717R c.2149G>A	de novo missense variant	PK	+ VSD and PV abnormality	Sifrim et al
271894[1], DDDP112485[2] 1[3]	F	p.G714R c.2140G>C	de novo missense variant	PK	+ ASD	Sifrim et al
259460[1], DDDP111320[2] 9[3]	F	p.N842S c.2525A>G	de novo missense variant	PK	+ Secundum ASD	Sifrim et al
258830[1], DDDP102232[2] 6[3]	F	p.R751Q c.2252G>A Also has de novo missense variant in TCF12, likely pathogenic exome MAF:4.06x10[-6]	de novo missense variant	PK	+ ASD, VSD	Sifrim et al
270818[1], DDDP112274 10[3]	F	p.N842S c.2525A>G	de novo missense variant	PK	+ Secundum ASD	Sifrim et al
DDD4K.04200[2], 261411[3]	F	p.V719G c.2156T>G	de novo missense variant	PK	-	DDD
DDD4K.00381[2], 264613[3]	F	p.R860Q c.2579G>A	de novo missense variant	PK	-	DDD

DDD4K.02553[2], 259077[3]	F	p.G717R c.2149G>A (Gly717Arg)	de novo missense variant. Possibly mosaic	PK	-	DDD
DDD4K.00408[2], 264961[3]	F	p.K734R c.2210A>G Lys734Arg	de novo missense variant	PK	-	DDD
DDD4K.02000[2], 271710[3]	F	c.2898-1G>A c. 40118318G/A	de novo splice acceptor variant	Splice site	-	DDD
DDD4K.02539[2], 270857[3]	F	p.V874L c.2620G>T Val874Leu	de novo missense variant	PK	-	DDD
11[3]	F	p.N842S Asn842Ser	de novo missense variant	PK	Secundum ASD	Hamilton et al
331720[3]	F	p. N842D c.2524A>G Asn842Asp	de novo missense variant	PK	ASD	Hamilton et al
301509[3]	F	p.D896N, c.2686G>A Asp896Asn	de novo missense variant	PK	-	Hamilton et al
1001	F	p.N842D c.2524A > G	de novo missense variant	PK	+ ASD and hypoplastic LPA	Bostwick et al
1002	F	p.N842S c.2525A > G	de novo missense variant	PK	+ ASD	Bostwick et al
1003	M	p.N842S c.2525A > G	de novo missense variant	PK	-	Bostwick et al
1004	M	p.N842S c.2525A > G	de novo missense variant	PK	+ Hypoplastic LPA, dilated PA	Bostwick et al
1005	F	p.N842S c.2525A > G	de novo missense variant	PK	-	Bostwick et al

1006	F	p.N842S c.2525A > G	de novo missense variant	PK	+ TOF (VSD)	Bostwick et al
1007	M	p.N842S c.2525A > G	de novo missense variant	PK	-	Bostwick et al
1008	M	p.K734E c.2200A > G	de novo missense variant	PK	+ EA	Bostwick et al
1009	M	p.N842S	Missense variant, inheritance unknown	PK	+ BAV, AS, LVNC	Bostwick et al
1	M	p.G717R c.2525A > G	de novo missense variant	PK	-	Uehara et al
2	F	p.N842S asn842ser	de novo missense variant	PK	-	Uehara et al
3	M	p.G717R Gly717arg	de novo missense variant	PK	-	Uehara et al
1	M	p. N842S asn842ser	de novo missense variant	PK	-	Van den Akker et al
2	F	p. N842S asn842ser	de novo missense variant	PK	-	Van den Akker et al
3	F	p. N842S asn842ser	de novo missense variant	PK	-	Van den Akker et al
4	F	p. N842S asn842ser	de novo missense variant	PK	-	Van den Akker et al
5	M	p. N842S asn842ser	de novo missense variant	PK	-	Van den Akker et al
6	F	p.G714D, Gly714asp	de novo missense variant	PK	+ ASD	Van den Akker et al
7	F	p.R737C Arg737cys	de novo missense variant	PK	-	Van den Akker et al
8	M	p.R880C Arg880cys	de novo missense variant	PK	-	Van den Akker et al
9 (monozygotic twin of 10)	F	c.2601 -2 A>G	de novo splice site variant	Splice site	+ Open ductus of Bottali	Van den Akker et al

10 (monozygotic twin of 9)	F	c.2601 -2 A>G	de novo splice site variant	Splice site	-	Van den Akker et al
11	F	A162fs	de novo frameshift variant	Near site of canonical arginine salt bridges (pT160)	-	Van den Akker et al
12	M	A162fs	de novo frameshift variant	Near site of canonical arginine salt bridges (pT160)	-	Van den Akker et al
13	M	A162fs	Frameshift variant inheritance unknown	Near site of canonical arginine salt bridges (pT160)	-	Van den Akker et al
14	M	p.Arg999*	de novo nonsense variant	After PK domain (a.a.705-998)	-	Van den Akker et al
15	F	p.Arg1025*	de novo nonsense variant	Polybasic Cluster	-	Van den Akker et al
8	F	p.D855H c.2563G>C	De novo missense variant	PK	-	Carneiro et al

Table 6.4 All reported mutations and CHD phenotype reported up to November 2019.

The majority are missense variants which fall within the protein kinase domain (PK). Loss of function mutations are highlighted.

Sifrim et al. [5], Bostwick et al. [1], DDD [6], Van den Akker et al. [4] , Uehara et al. [3], Hamilton et al. [2], Carneiro et al. [890]

Bostwick et al. described an individual with cardiomyopathy and electrical disturbance and suggested that this may be an age related consequence of the BAV as the individual was 38 years old (individual 1009) [833]. Given BAV is a more common type of CHD in the general population with a reported prevalence of 0.5-2% [892, 893], we should consider the possibility that this is an incidental finding. No other individuals with variants in *CDK13* have been reported with BAV. However, a single individual with a 407 Kb duplication, with one breakpoint in *CDK13,* has BAV. They also have another 180kb duplication on chromosome 18 of uncertain significance. Inheritance of these CNVs is unknown so the contribution of both of these CNVs to the individual's phenotype is unclear. Irregular heart rhythms and reduced ventricular myocardium were noted in Cdk13-/- mice, although it is not clear if this represents a primary conduction problem or could be linked with LVNC [885].

More detailed expression studies will be required to explore where exactly and when in heart development *CDK13* is expressed. This may help explain why certain defects are more common than others.

6.3.4 Mutational Mechanism causing Clinical Abnormalities

Given *CDK13* has relatively ubiquitous expression in development, it is feasible that its absence or reduction causes the widespread phenotypes reported. The mechanism is currently under discussion and there has been limited functional work so far [836]. The mutational spectrum seen in individuals with CHDFIDD/*CDK13*-related Disorder includes both missense and loss of function (LoF) mutations. The missense mutations are mainly clustered within the kinase domain. Other pathogenic mutations are reported in other domains of the protein. Figure 6.8 shows the mutations reported, with and without CHD, and table 6.5 lists some of the mutations and the important residues they affect. Characteristics of the protein structure and genotype phenotype correlations can provide clues as to the mechanism of disease.
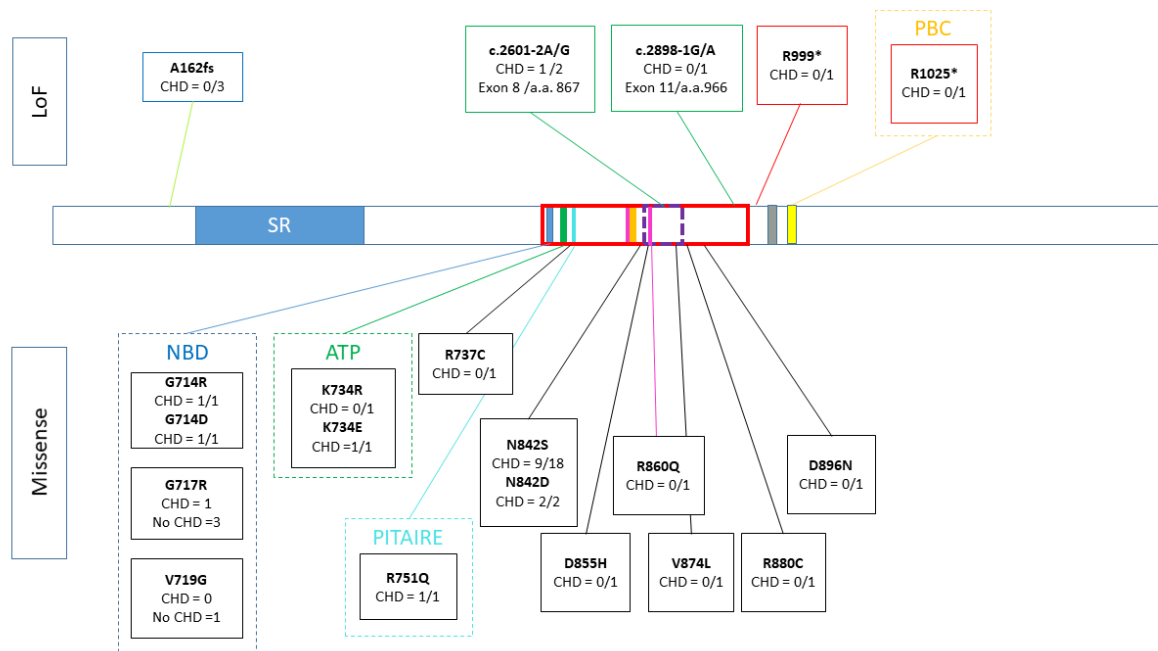
Figure 6.8 Representation of the CDK13 protein and reported mutations.

The number of individuals with and without CHD at each locus are recorded. [12, 489, 833-836]. The majority fall within the kinase domain. Loss of function mutations are shown above the protein, and missense mutations are shown below. Loss of function frame shift mutations are in blue boxes, splice site in green and nonsense in red. Some important domains and residues in CDK13: SR: Serine Rich domain, Red box: protein kinase domain, Purple dashed box: Activation loop, Blue line NBD: Nucleotide Binding Domain, Green line ATP: ATP binding site Lysine, Turquoise line: arginine which binds to Cyclin K, orange line PAS: Proton acceptor site, Pink Line: Phospho-threonine co-ordination, Grey box: DCHEL sequence including HE motif, yellow box: Polybasic Cluster.

A number of these mutations affect important domains and residues within the protein, including clusters in the protein kinase domain, nucleotide binding domain and magnesium and ATP binding sites. It is likely that mutations in the protein kinase domain disturb the phosphorylation activity of CDK13. CDK13 contributes to phosphorylation of Ser 2 and Ser 5 in the heptad repeats of RNA polymerase II. CDK13 also binds splicing factors and proteins, and associates with parts of the splisosome complex (ASF/SF2) and is located in areas of the

cell where proteins involved in pre-mRNA processing are found [873, 877, 894]. Loss of normal CDK13 function is likely to result in inability of RNA polymerase II to form and splice RNA effectively. This would result in abnormalities in the cell transcriptome and is likely to involve pathways involved in cell signalling and growth [848].

Mutation	Consequence
p.K734R p.K734E	These two variants affect the residue involved in ATP binding, lysine 734, which is conserved in human CDKs. This may well alter function of *CDK13,* given that ATP hydrolysis is required for CDKs to phosphorylate their targets.
p.R751Q	This variant affects the conserved arginine in the PITAIRE motif, R751. This arginine is required for binding Cyclin K. It forms a salt bridge with the second glutamate in the 105KVEE motif in CCNK. This is likely to be significant given this motif is conserved across cyclins K, C, H and T which interact with Cdk 7, 8, 9, 12 and 13. CDK13 function relies on binding CCNK.
p.N842S p.N842D	The commonest site for mutations is at residue 842, in the kinase domain. This is close to proton acceptor site (aspartate at 837) and a canonical arginine at 836. A change from asparagine to aspartic acid introduces a negative charge in the active site of the protein kinase domain [836]. It is also thought to impair magnesium binding, which is required for normal function [12].
p.R860Q	Another of the canonical arginines at residue 860 is affected by a further missense mutation. These arginines are required to form salt bridges with the phosphate group of T871, which falls within the activation segment. Salt bridge formation is one of the hall marks of CDK activation [848].
p.A162fs	The frameshift mutations p.A162fs fall near the site of one of the salt bridges. The canonical arginines form salt bridges with the

	phosphate group of T160. This might produce a protein that is degraded by nonsense mediated decay or an alternative translation start site could be used and this is unknown.
p.G714D p.G717R p.V719G	Seven individuals have missense mutations within the nucleotide binding domain (residues 711-719). The effects are unknown but tyrosine at residue 716 in CDK13 is known to be phosphorylation target [895].

Table 6.5 mutations reported in CDK13, with information about their possible functional consequences.

Initially gain of function was suggested as the mechanism given that there was more clustering of variants than expected by chance, and no loss of function mutations [12, 833]. Since then, loss of function mutations have been reported, so loss of function should also be considered as a disease mechanism [836]. A single splice site mutation was reported initially [489] and then further splice site, frame shift and nonsense mutations [836]. This suggests that the initial proposed gain of function mechanism might not hold true for all individuals.

The missense constraint score for *CDK13* is 3.77, indicating that it is intolerant of missense mutations [494]. The pLI value is 0.75 and it has a haploinsufficiency index of 11.15%, suggesting that it is not highly intolerant of haploinsufficiency [772]. GnomAD records 9 loss of function mutations (4 splice site, 2 frame shift and 9 stop gained, accessed Dec19). As suggested by Van den Akker et al., it would be interesting to find out more about these individuals included in gnomAD to see if they might have a mild phenotype.

The CNV z score for *CDK13* is -2.08 which suggests CNVs are not a major cause of problems when *CDK13* is included [896]. However, the Database of Genomic

Variants shows a relative lack of deletions involving *CDK13*, with the exception of a few small deletions, most of which appear to be intronic [897]. There are a large number of CNVs overlapping *CDK13* on the Decipher database (30 CNVs overlapping *CDK13,* accessed September 2019), which may or may not contribute to the individual's phenotype [522]. Given some are inherited from normal parents, it is unlikely to explain the affected individual's phenotype without other contributing factors. Two thirds are losses. There are 10 CNVs that are between 100kb and 1 Mb, only one is a deletion, of uncertain inheritance. Intellectual disability or neurodevelopmental delay are reported in 4 individuals. Together this suggests that while deletions including *CDK13* are not seen very often in a normal population, there is no conclusive evidence that deletions including *CDK13* are damaging.

The frameshift and nonsense variants reported could result in haploinsufficiency or altered function of the CDK13 protein. Van den Akker et al. reported three individuals with the frame-shift variant p.Ala162fs [836]. It is not known if nonsense-mediated decay of this transcript occurs, or if there is an alternative translational start site. This mutation falls close to the threonine at position 160 which is involved in forming salt bridges with the canonical arginines in CDK13 [848]. Two nonsense mutations are reported (c.2995C > T, p.Arg999* and c.3073C > T, p.Arg1025*), but both truncate the protein after the protein kinase domain, so the effect on the protein function is uncertain. Additionally, one variant falls within the polybasic cluster which is important for helix formation within CDK13 and might be important for attractions to RNA polymerase II [848]. This could result in altered function. Haploinsufficiency might result for all of these products, if they are removed by nonsense mediated decay.

The two splice site mutations affect introns 7 and 10, which fall in the middle and towards the end of the protein kinase domain. Van den Akker et al. carried out some functional work to determine the functional consequence of c.2601-2A>G [836]. They detected three transcripts; one normal transcript from the WT allele, one with an out-of-frame coding sequence from a cryptic splice site in

exon 8, and a third transcript with absent exons 8 and 9, but exon 10 was present with the correct reading frame. There was no evidence of nonsense-mediated decay. Interestingly, the other splice site mutation occurs before exon 11, equivalent to amino acid 966. This is not associated with CHD. Given the protein kinase domain ends at 998, it may be that loss or abnormal protein at the end of the kinase domain is less significant and does not have such a detrimental effect, resulting in a normal heart and milder phenotype.

All of this evidence taken together suggests that haploinsufficiency and loss of function is a possible mechanism in some individuals. However, as discussed already, the clustering of missense mutations in the protein kinase domain could suggest a possible gain of function mechanism. Individuals with frame shift and nonsense variants exhibit a milder phenotype, perhaps because they result in haploinsufficiency [836, 891].

The commonest missense mutation reported affects residue 842, and are associated with more structural abnormalities. Hamilton et al. suggested the more severe phenotype may be due to sequestration of Cyclin K by the mutant protein resulting in a dominant negative effect, as opposed to just loss of kinase action [891].

A syndromic neurodevelopmental disability syndrome has been reported with abnormalities in *CCNK*, which *CDK13* and CDK12 form complexes with. Fan et al. reported three individuals with intellectual disability and de novo deletions including *CCNK,* and one with a rare de novo missense variant in *CCNK* [869]. The mutation p.Lys111Glu falls within the highly conserved cyclin box and is not reported in gnomAD or ExAC [570, 896]. This is also sited close to 105KVEE, which is required for binding with CDK13. The authors report that they share some features with individuals who have mutations in *CDK13* including midface defects, hypertelorism, broad nasal tip, thin upper lip and clinodactyly. The proposed mechanism was haploinsufficiency based on three individuals having deletions of *CCNK* and structural modelling of the missense mutation leading to

partial loss of function. This individual harbouring the mutation had a milder phenotype. Rescue experiments in Zebrafish also supported haploinsufficiency as a mechanism.

Although there are some similarities between these individuals and those with *CDK13* mutations, there are also phenotypic distinctions between the two. Individuals with mutations in *CDK13* show structural abnormalities in the brain, heart, palate and kidneys. No such structural defects are reported in individuals with deletions or mutations in *CCNK*. Some of the facial features are different and those with abnormalities affecting *CCNK* exhibit macrocephaly and normal or higher than average height and weight. CHDFIDD is more commonly associated with microcephaly and lower growth parameters. Posteriorly rotated low set ears and clinodactyly were reported in both groups, but these are soft signs and clinodactyly was only seen in one individual.

The developmental disability seen with *CCNK* haploinsufficiency is far more severe. All had severe or profound disability, whereas the individuals with *CDK13* mutations had mild-moderate disability, and some had low normal IQ [869]. The presence of NDD in both cohorts fits with the known regulation of the genes important in neurone development by *CDK12* and *CDK13* and the resulting lack of both CDK12-CCNK and CDK13-CCNK complexes. The zebrafish model produced to show the effects of haploinsufficiency in *Ccnk* produced deformed heads, reduced eye size, abnormal spinal cord and features suggestive of neural cell death consistent with the role of *Cdk12* and *Cdk13* in the central nervous system. They also showed evidence of DNA strand breaks consistent with the role of Cdk12-Ccnk in controlling genes involved in DNA repair [869].

Given that the Ccnk Zebrafish model relied on morpholinos, there is a potential for off-target effects. A Ccnk null mouse was embryonic lethal, but heterozygous mice were apparently normal [846]. Fan et al. commented that features such as facial dysmorphism and neurodevelopmental disability might not be detected in

these heterozygous mice without specific testing and therefore could potentially replicate the human phenotype [869].

Haploinsufficiency of *CCNK* would result in reduced amounts of both *CDK13-Cyclin K* and *CDK12-Cyclin K* complexes, but we do not know what the consequences would be for other CDK13-partner complexes including Cyclin L. Given that both genes are important in the brain, we would expect developmental disability of a greater severity than if *CDK13* or *CDK12* exhibited reduced function alone. The lack of structural abnormalities seen in individuals with haploinsufficiency of *CCNK* correlates with the lack of structural malformations seen in individuals with loss of function mutations in *CDK13*. This also suggests that in those individuals with structural malformations and mutations in *CDK13*, the mechanism is something other than loss-of-function or haploinsufficiency, and potentially could be gain of function as we suggested, or dominant negative effects as proposed by Hamilton et al. Loss of *CDK12* is likely to be embryonic lethal, even in the heterozygous state given that no individuals have been reported with mutations in this gene. Supporting this is the known role of *CDK12* in regulating genes required for DNA repair. It is likely that damage goes unchecked and results in cell death early on in gestation.

Whilst the CCNK reports add weight to the pathogenic nature of CDK13 and the genotype-phenotype correlations suggested, only a small number of individuals have been reported. It is likely that the full phenotypic spectrum of *CDK13* and *CCNK* has not yet been described and there may be more variability in neurodevelopmental disability and structural defects. Functional work to clarify the mechanism of action should be carried out.

De novo missense mutations in the ATP-binding pocket of the kinase domain of *CDK8* are also associated with S-CHD [843]. Like CDK13, CDK8 regulates transcription as it forms part of the mediator complex [842]. The mediator complex is made up of CDK8 or CDK19, together with cyclin C, MED12 or MED12L, and MED13 or MED13L. The mediator complex joins gene specific

activators and regulatory elements, to RNA polymerase II and general transcription factors [898].

Common phenotypes reported in association with mutations in *CDK8* include hypotonia, mild to moderate intellectual disability, behavioural problems and facial dysmorphism. Additional features included agenesis of the corpus callosum, ano-rectal malformations, seizures, and hearing or visual impairments. There is significant overlap with the syndromic developmental disorders that result from mutations in *MED12, MED13*, and *MED13L*. Relevant to this thesis is the report that 6 of the 12 individuals with mutations in *CDK8* had CHD. This included VSDs, double orifice mitral valve, HLHS, CoA with mitral stenosis, TOF, and ASD and BAV in combination with a hypoplastic aortic arch [843]. Mutations in MED13L and MED13 are also associated with CHD [899, 900]. This further supports the role of genes critical for normal transcription processes, in cardiogenesis.

6.3.4 Which Individuals develop CHD

It is not clear why some individuals with *CDK13* mutations have developed CHD and others have not. Overall, just under 40% of individuals with *CDK13* mutations have CHD. This is predominantly an ASD. Only mutations within the protein kinase domain between residues 714 and 842 are associated with CHD. However even within this smaller section of the protein, only 16/31 (52%) have CHD. The only addition to this is the splice site mutation that occurs just before exon 8. However, this is a PDA in a preterm infant so we should be cautious in applying blame to *CDK13*, even if it accounts for the rest of the phenotype.

Within specific domains there is reduced penetrance for CHD. There are 7 variants in the nucleotide binding domain (a.a711-719), but only 3 are associated with CHD. There are two mutations at residue 714, p.G714R and p.G714D. Despite the different amino acid substitution, both are associated with CHD and the same cardiac phenotype, an ASD. Glycine is hydrophobic and often

found in the protein core, whereas both aspartic acid and arginine are charged and have side chains that often form salt bridges. The similarity between these two substituted amino acids might result in a similar effect on the protein and subsequent CHD. This does not hold true for all mutations though, and even substitutions with similar amino acids do not result in CHD. There are two variants affecting the ATP binding site at residue 734, p.K734R without CHD and p.K734E with CHD. Lysine and the substituted amino acids arginine and glutamic acid are all charged amino acids.

The penetrance of CHD in identical mutations is not complete either. 25% (1/4) of those with p.G717R have CHD, and 50% (9/18) of those with p.N842S have CHD. The majority of individuals with p.N842S have an ASD, but one has an additional PV abnormality, one has TOF, one has BAV with LVNC, and one has a hypoplastic left PA only.

The chance of CHD is highest with mutations falling within the protein kinase domain between residues 714 and 842, but we cannot predict from the mutation alone who will develop CHD. Within the same domains and even at the same residue, and with the same mutation there is no consistent pattern to determine who will develop CHD. A single explanation as to why CHD is found in those mutations in the protein kinase domain, before the active site, is not clear. Perhaps there is a more significant effect on the kinase domain structure and function with mutations in the first half, and less damaging effects with mutations in the latter half of the kinase domain where perhaps there is some residual function. However fewer mutations are reported in the tail half of the protein kinase domain, which could mean they are either selected against or just less common.

Most individuals are affected with ASD. Some have additional VSD, PV or PA abnormalities. The only unusual occurrences are TOF at residue 842, but this shares VSD and PV abnormalities and so could have a similar aetiology. Another individual with a mutation at this locus has BAV and LVNC. Given BAV is a more

common congenital abnormality there is always the possibility that this is a coincidental lesion, although an aortic valve abnormality was seen in one of the heterozygous mice I examined. There is a single individual with EA. This is a less common form of CHD and is usually accompanied by an ASD. It stands out as this individual has a mutation at the ATP binding site (p.K734E). The other individual with a different mutation at the same residue (p.K734R) does not have CHD. More detailed analysis of possible effects on protein structure and ATP binding is required along with functional work to determine the effect of these mutations.

The splice site mutation c.2601-2 A/G which is reported to cause CHD is unusual as most other mutations are clustered in the kinase domain and are missense mutations. This individual is reported as having an open ductus of Bottali (PDA) [836]. It should be noted that the individuals with these mutations are monozygotic twins. Supplementary data confirms that they were affected by twin twin transfusion problems and were born at 35 and 36 weeks. It does not clarify the nature of the PDA in any more detail. As described previously, the effect of this mutation was production of WT and abnormal transcripts so it is unclear what the overall effect would have been. Given that only one of the two twins had CHD, and the PDA may well have been associated with prematurity, it is difficult to say that it is related to *CDK13* function.

One possible explanation for the reduced penetrance and variable expressivity for CHD is that CDK13 variants increase your chance of developing CHD, but that additional factors may push certain individuals over the threshold for developing CHD. This could include placental, other genetic, maternal or environmental influences. This is in line with the idea that CHD is an oligogenic disease with influences from other environmental and epigenetic factors, which are discussed more in the introductory chapter.

Aside from the influence of the CDK13-Cyclin K complex on RNA polymerase II, there may be other undiscovered functions and regulators of CDK13 that

contribute to CHD occurring. CCNK expression levels are reasonably stable, as seen in other cyclins that interact with transcriptional CDKs, so this does not seem to be a major controlling factor of CDK13 function [848]. CDK13 is a very large protein and there are numerous domains that allow for multiple intermolecular interactions that might contribute to controlling its activity.

Another cyclin that interacts with CDK13 is Cyclin L1. Cyclin L1 interacts with CDK11A, CDK12 and CDK13 [901, 902] and may play a role in regulation of RNA polymerase II. Expression of Cyclin L1 in the human heart is high [544] and it is expressed in the human foetal heart at three and five months post conception [903]. There is currently no known clinical syndrome associated with this gene, although it may play a role in head and neck cancers. There is no mouse model as yet. There are likely to be additional proteins that influence CDK13 that are currently unknown and may help explain reduced penetrance and variable expressivity of all the clinical features of CHDFIDD/CDK13-related disorder.

6.3.5 Clinical Recommendations

Current clinical recommendations have been provided based on the phenotypes seen. As more individuals are reported this may well change.

Core suggestions [833, 836, 891]:

- ECHO at diagnosis
- Developmental assessment and support for development and educational needs
- Support with feeding if required

Consider these further investigations based on presentation and balance of risks [833, 891]:

- CNS imaging

- Renal ultrasound
- Ophthalmology review
- Sacral imaging

Specific factors that might contribute to developmental delay including hypotonia and hearing loss. Speech and language therapy should be instigated if required. Additional support for neonatal feeding problems may be required. Van den Akker et al. suggested that close attention should be paid to risk of seizures during neurological examination [836]. Clefts should be considered at all newborn checks and those with feeding difficulties. Eye and hearing checks will also be part of routine checks/questions for new parents.

Given the relatively high incidence of CHD in this cohort I would agree that a baseline echo is wise, even if asymptomatic. Small ASDs can easily go unrecognised if the individual is asymptomatic. At this stage I think it is impossible to decide whether or not to check an individual's heart based on genotype, and it would seem sensible to scan all individuals given the low risk nature of an echocardiogram. The significance of LVNC in the older adult is unclear. Bostwick et al. also proposed that the arrhythmias seen in this individual may be related to the underlying BAV and not a primary process [833]. Only continuing follow up of individuals will give us a clearer insight into the risks of complications in later life. Many individuals with CHD including septal defects will have continuing follow up including monitoring for rhythm disturbances.

Life expectancy is unknown but two individuals have been reported who are 38y and 54y [836, 904]. The role of *CDK13* (and *CDK7*, 8, 9 and 12) in tumour oncogenesis is currently being explored and the implications of this for individuals with CDK13 mutations is currently unclear [865]. Any concerning signs or symptoms should be followed up promptly.

Although all individuals have de novo mutations, parents may wish to consider prenatal testing if they are planning further children in case of gonadal mosaicism. The incidence of this is unknown.

Whilst the phenotypic spectrum is unknown, it makes it difficult to define a population who should be offered a genetic test for *CDK13*. We should employ a genotype first approach in the majority, unless *CDK13* is strongly suspected. I would propose offering testing in individuals with septal defects and NDD currently as a minimum. Bostwick et al. suggested testing those with mutation negative Kabuki syndrome due to phenotypic overlap between the two conditions [833].

6.3.6 Results from the *Cdk13*tm1b(EUCOMM)Hmgu Mouse

This tm1b allele mouse was provided by the WTSI and employs a reporter tagged deletion allele (post-cre). Exons 3 and 4 are deleted. There is 95% sequence homology between the mouse and human CDK13 (Figure 6.9). The protein kinase domain sequence is identical.

```
Score              Expect  Method                        Identities        Positives        Gaps
1800 bits(4661)    0.0     Compositional matrix adjust.  1165/1232(95%)    1188/1232(96%)   5/1232(0%)

Query  283   TKSSKEPPSAYKEPPKAYREDKTEPKAYrrrrslsp-lggrDDSPVSHRASQSLrsrksp  341
             TKSSKEPPSAYKEPPKAYREDK+EPKAYRRR+      LGGRD+SPVSHRASQSLRSRKSP
Sbjct  283   TKSSKEPPSAYKEPPKAYREDKSEPKAYRRRQRSLSPLGGRDESPVSHRASQSLRSRKSP  342

Query  342   spagggsspysrrlprspspsysrrrspsysrhssyERGGDVspspsysssswrrsrspysp  401
             SPAGGGSSPYSRRLPRSPSPYSRRRSPSYSRHSSYERGGDVSPSPYSSSSWRRSRSPYSP
Sbjct  343   SPAGGGSSPYSRRLPRSPSPYSRRRSPSYSRHSSYERGGDVSPSPYSSSSWRRSRSPYSP  402

Query  402   VLrrsgksrsrspyssrhsrsrsrhrlsrsrsrhssispsTLTLKSSlaaelnknkkara  461
             VLRRS KSRSRSPYSSRHSRSRSRHRLSRSRSRHSSISPSTLTLKSSLAAELNKNKKARA
Sbjct  403   VLRRSAKSRSRSPYSSRHSRSRSRHRLSRSRSRHSSISPSTLTLKSSLAAELNKNKKARA  462

Query  462   aeaaraaeaakaaeatkaaeaaakaakaSNTSTPTKGNTETSASASQTNHVKDVKKIKIE  521
             AEAARAAEAAKAAEA KAAEAAAKAAKASN STPTKGNTET AS SQTNHVK+VKK+K E
Sbjct  463   AEAARAAEAAKAAEAAKAAEAAAKAAKASNASTPTKGNTETGASVSQTNHVKEVKKLKTE  522

Query  522   HAPSPSSGGTLKNDKAKTKPPLQVTKVENNLIVDKATKKAVIVGKESKSAATKEESVSLK  581
             HAPSPSSGGT+K+DKAKTKPPLQVTKV+NNL V+KATKK V VGKESK AATKEE VS K
Sbjct  523   HAPSPSSGGTVKSDKAKTKPPLQVTKVDNNLTVEKATKKTV-VGKESKPAATKEEPVSTK  581

Query  582   EKTKPLTPSIGAKEKEQHVALVTSTlpplplppmlpEDKEADSLRGNISvkavkkevekk  641
             EK+KPLTPS GAKEKEQHVALVTSTLPPLPLPPMLPEDK+ADSLRGNISVKAVKKEVEKK
Sbjct  582   EKSKPLTPSTGAKEKEQHVALVTSTLPPLPLPPMLPEDKDADSLRGNISVKAVKKEVEKK  641

Query  642   LRClladlplppelpggddlSKSPEEKKTATQLHSKRRPKICGPRYGETKEKDIDWGKRC  701
             LRCLLADLPLPPELPGGDDLSKSPEEKKTA QLHSKRRPKICGPRYGE KEKDIDWGKRC
Sbjct  642   LRCLLADLPLPPELPGGDDLSKSPEEKKTAAQLHSKRRPKICGPRYGEIKEKDIDWGKRC  701

Query  702   VDKFDIIGIIGEGTYGQVYKARDKDTGEMVALKKVRLDNEKEGFPITAIREIKILRQLTH  761
             VDKFDIIGIIGEGTYGQVYKARDKDTGEMVALKKVRLDNEKEGFPITAIREIKILRQLTH
Sbjct  702   VDKFDIIGIIGEGTYGQVYKARDKDTGEMVALKKVRLDNEKEGFPITAIREIKILRQLTH  761

Query  762   QSIINMKEIVTDKEDALDFKKDKGAFYLVFEYMDHDLMGLLESGLVHFNENHIKSFMRQL  821
             QSIINMKEIVTDKEDALDFKKDKGAFYLVFEYMDHDLMGLLESGLVHFNENHIKSFMRQL
Sbjct  762   QSIINMKEIVTDKEDALDFKKDKGAFYLVFEYMDHDLMGLLESGLVHFNENHIKSFMRQL  821

Query  822   MEGLDYCHKKNFLHRDIKCSNILLNNRGQIKLADFGLARLYSSEESRPYTNKVITLWYRP  881
             MEGLDYCHKKNFLHRDIKCSNILLNNRGQIKLADFGLARLYSSEESRPYTNKVITLWYRP
Sbjct  822   MEGLDYCHKKNFLHRDIKCSNILLNNRGQIKLADFGLARLYSSEESRPYTNKVITLWYRP  881

Query  882   PELLLGEERYTPAIDVWSCGCILGELFTKKPIFQANQELAQLELISRICGSPCPAVWPDV  941
             PELLLGEERYTPAIDVWSCGCILGELFTKKPIFQANQELAQLELISRICGSPCPAVWPDV
Sbjct  882   PELLLGEERYTPAIDVWSCGCILGELFTKKPIFQANQELAQLELISRICGSPCPAVWPDV  941

Query  942   IKLPYFNTMKPKKQYRRKLREEFVFIPAAALDLFDYMLALDPSKRCTAEQALQCEFLRDV  1001
             IKLPYFNTMKPKKQYRRKLREEFVFIPAAALDLFDYMLALDPSKRCTAEQALQCEFLRDV
Sbjct  942   IKLPYFNTMKPKKQYRRKLREEFVFIPAAALDLFDYMLALDPSKRCTAEQALQCEFLRDV  1001

Query  1002  EPSKMPPPDLPLWQDCHELWSKKRRRQKQMGMTDDVSTIKAPRKDLSLGLDDSRTNTPQG  1061
             EPSKMPPPDLPLWQDCHELWSKKRRRQKQMGMTDD+STIKAPRKDLSLGLDDSRTNTPQG
Sbjct  1002  EPSKMPPPDLPLWQDCHELWSKKRRRQKQMGMTDDLSTIKAPRKDLSLGLDDSRTNTPQG  1061

Query  1062  VLPSSQLKSQGSSNVAPVKTGPGQHLNHSELAILLNLLQSKTSVNMADFVQVLNIKVNSE  1121
             VLP +QLKSQ +SNVAPV TGPGQ LNHSELAILLNLLQSK+SVNMADFVQVLNIKVNSE
Sbjct  1062  VLPPAQLKSQSNSNVAPVITGPGQPLNHSELAILLNLLQSKSSVNMADFVQVLNIKVNSE  1121

Query  1122  TQQQLNKINLPAGILATGEKQTDPSTPQQESSKPLGGIQPSSQTIQPKVETDaaqaavqs  1181
             TQQQLNKINLPAGILATGEKQTDPSTPQQESSK LGG+QP SQTIQPKVETDAAQAAVQS
Sbjct  1122  TQQQLNKINLPAGILATGEKQTDPSTPQQESSKSLGGVQP-SQTIQPKVETDAAQAAVQS  1180

Query  1182  afavLLTQLIKAQQSKQKDVLLEERENGSGHEASLQLRPPPEPSTPVSGQDDLIQHQDMR  1241
             AFAVLLTQLIKAQQSKQKD +LEERENGSGHEA LQLRPP EPSTP SGQDDLIQHQD R
Sbjct  1181  AFAVLLTQLIKAQQSKQKDAMLEERENGSGHEAPLQLRPPLEPSTPGSGQDDLIQHQDRR  1240

Query  1242  ILELTPEpdrprilppdqrppeppepppvteedLDYRTENQHVPTTSSSLTDPHAGVKaa  1301
             ILELTPEPDRPRILPPDQRPPEPPEPPPVTEEDLDYRTENQHVPTTSSSLTDPHAGVKAA
Sbjct  1241  ILELTPEPDRPRILPPDQRPPEPPEPPPVTEEDLDYRTENQHVPTTSSSLTDPHAGVKAA  1300

Query  1302  llqllaqhqpqDDPKREGGIDYQAGDTYVSTSDYKDNFGSSSFSSAPYVSNDGLGSSS-A  1360
             LLQLLAQHQPQDDPKREGGIDY GDTYV +SDYKDNFG SSFS+APYVS+DGLGSSS A
Sbjct  1301  LLQLLAQHQPQDDPKREGGIDYPTGDTYVPSSDYKDNFG-SSFSAAPYVSSDGLGSSSAA  1359

Query  1361  PPLERRSFIGNSDIQSLDNYSTASSHSGGPPQPSAFSESFPSSVAGYGDIYLNAGPMLFS  1420
              PLE RSFIGNSDIQSLDNYSTASSH+GGPPQ SAF+ESF SSVAGYGDIYLNAGPMLFS
Sbjct  1360  APLEARSFIGNSDIQSLDNYSTASSHTGGPPQTSAFTESFASSVAGYGDIYLNAGPMLFS  1419

Query  1421  GDKDHRFEYSHGPIAVLANSSDPSTGPESTHPLPAKMHNYNYGGNLQENPSGPSLMHGQT  1480
             GDKDHRFEYSHGPI VL NS+DPSTGPESTHPLPAKMHNYNYGGNLQENP GPSLMHGQT
Sbjct  1420  GDKDHRFEYSHGPITVLTNSNDPSTGPESTHPLPAKMHNYNYGGNLQENPGGPSLMHGQT  1479

Query  1481  WTSPAQGPGYSQGYRGHISTSTgrgrgrgLPY  1512
             WTSPAQGPGYSQGYRGHISTS GRGRGRGLPY
Sbjct  1480  WTSPAQGPGYSQGYRGHISTSAGRGRGRGLPY  1511
```

Figure 6.9 Alignment of mouse and human CDK13 proteins.

Alignment from https://www.ncbi.nlm.nih.gov/homologene. Query is Homo Sapiens, Subject is Mus Muscularis. Grey lowercase letters represent areas of low complexity sequence where artefacts might occur.

6.3.6.1 Embryonic lethality in *Cdk13*^{tm1b(EUCOMM)Hmgu} mice

Cdk13^{tm1b(EUCOMM)Hmgu} null mice exhibit embryonic lethality. Lethality was not limited to early gestation, and homozygotes were identified at E15.5 (Tables 6.6 and 6.7). Stages later than this were not examined. There does not appear to be any survival disadvantage in *Cdk13*^{tm1b/+} mice.

Gestation	Cross	Expected	Observed
E9.5	*Cdk13*^{Tm1b/+} x *Cdk13*^{Tm1b/+}	Tm1b/tm1b: 25% Tm1b/+: 50% +/+: 25%	Tm1b/tm1b: 0 (0%) Tm1b/+: 9 (100%) +/+: 0 (0%) Unknown: 0 (0%)
E10.5	*Cdk13*^{Tm1b/+} x *Cdk13*^{Tm1b/+}	Tm1b/tm1b: 25% Tm1b/+: 50% +/+: 25%	Tm1b/tm1b: 0 (0%) Tm1b/+: 1 (17%) +/+: 2 (33%) Unknown: 3 (50%)
E11.5	*Cdk13*^{Tm1b/+} x *Cdk13*^{Tm1b/+}	Tm1b/tm1b: 25% Tm1b/+: 50% +/+: 25%	Tm1b/tm1b: 1 (11%) Tm1b/+: 7 (78%) +/+: 1 (11%) Unknown: 0 (0%)
E12.5	*Cdk13*^{Tm1b/+} x *Cdk13*^{Tm1b/+}	Tm1b/tm1b: 25% Tm1b/+: 50%	Tm1b/tm1b: 1 (14%) Tm1b/+: 4 (57%) +/+: 2 (29%) Unknown: 0 (0%)

		+/+: 25%	
E13.5	*Cdk13*^{Tm1b/+} x *Cdk13*^{Tm1b/+}	Tm1b/tm1b: 25% Tm1b/+: 50% +/+: 25%	Tm1b/tm1b: 1 (16.5%) Tm1b/+: 4 (67%) +/+: 1 (16.5%) Unknown: 0 (0%)
E15.5	*Cdk13*^{Tm1b/+} x *Cdk13*^{Tm1b/+}	Tm1b/tm1b: 25% Tm1b/+: 50% +/+: 25%	Tm1b/tm1b: 2 (5%) Tm1b/+: 21 (51%) +/+: 6 (15%) Unknown: 12 (29%)

Table 6.6 *Cdk13*^{tm1b(EUCOMM)Hmgu} embryo genotypes.

Genotypes for all embryos collected between E9.5 and E15.5 from timed matings between *Cdk13*^{tm1b/+} mouse pairs. Homozygous embryos were collected at most stages and were still alive at E15.5.

Tm1b/tm1b: *Cdk13*^{Tm1b/tm1b}, Tm1b/+: *Cdk13*^{Tm1b/+}, +/+: *Cdk13*^{+/+}

	*Cdk13*tm1b/tm1b	*Cdk13*tm1b/+	*Cdk13*+/+	Genotype uncertain
1. Tm1b/+ x Tm1b/+	6.4% (n=5) (expected 25%)	60% (n=46) (expected 50%)	16% (n= 12) (expected 25%)	19% (n=15)
2. Tm1b/+ x Tm1b/+	0% (n=0) (expected 25%)	63.6% (n=14) (expected 50%)	31.8% (n=7) (expected 25%)	4.5% (n=1)
3. Tm1b/+ x +/+	0% (n=0)	47% (n=47) (expected 50%)	35% (n=35) (expected 50%)	6% (n=6)

Table 6.7 *Cdk13*[tm1b(EUCOMM)Hmgu] embryo genotypes.

Row 1: Genotypes of all embryos harvested at E9.5-E15.5, produced from heterozygous crosses. Row 2: Live-born mice from *Cdk13*[tm1b/+] crosses. Row 3: Live-born mice from *Cdk13*[tm1b/+] x *Cdk13*[+/+] cross.

The small numbers of embryos genotyped (n=78) means it is difficult to be completely sure that the ratios of *Cdk13*[+/+]/ *Cdk13*[tm1b/+]/ *Cdk13*[tm1b/tm1b] are as expected. This is compounded by empty decidua, which often meant that the genotype could not be determined (n=15). It is also possible that the majority of the empty decidua represent homozygous embryos that were deceased at an early gestation. However, the largest number of embryos were harvested at E15.5 and the ratios are not too different to expected results (Table 6.6, Chi squared test 0.045 assuming embryonic lethality prior to E15.5). Considering adults as well, there is no clear indication of reduced survival in *Cdk13*[tm1b/+] (Table 6.7).

Cdk13[tm1b/tm1b] mice exhibit complete embryonic lethality over a wide range of embryonic stages. The finding of two *Cdk13*[tm1b/tm1b] mice at E15.5 suggests that lethality can occur after this, but I did not examine any later stages. This is in keeping with the results of Novakova et al., who found complete embryonic lethality by E16.5 for their hypomorphic tm1a allele, and earlier mortality in the

null tm1d allele [885]. A single *Cdk13*^{tm1b/tm1b} embryo appeared to have died just prior to E12.5, and another harvested at E13.5 was necrotic with no clear distinguishable features, suggesting early demise. In addition, many empty decidua were found that failed the genotyping process. Some of these may represent null embryos. Together these results suggest that embryonic lethality in the tm1b allele can occur from early stages to after E15.5, but that null mice always die prior to birth. This indicates that *Cdk13* has an important role in embryogenesis.

All nine embryos harvested at E9.5 had the same genotype, *Cdk13*^{Tm1b/+}. This was confirmed by re-genotyping. Unfortunately I did not collect a further litter at this stage, but it is unlikely that this is a representative result. This should be repeated to determine what the appearance of *Cdk13*^{tm1b/tm1b} embryos is at this stage of gestation.

Despite the structural similarities, CDK12 and CDK13 regulate different genes [848]. Further evidence supporting non-redundancy is provided by the *Cdk12* and *Cdk13* mouse models, which are both embryonic lethal. Both *Cdk12* and *Ccnk* (Cyclin-K) null mice demonstrate embryonic lethality much earlier, at the blastocyst stage [846, 884]. This supports non-redundancy between *Cdk12* and *Cdk13* during the early stages of embryogenesis. *Cdk12* null embryos showed increased apoptosis in the blastocysts and reduced expression of certain DNA damage response genes [884]. There is no evidence of mutations *CDK12* causing a disorder in humans even in heterozygous states, suggesting embryonic lethality in humans too.

Dia et al. examined murine pluripotent embryonic stem cells and found that levels of CDK12, CDK13 and CCNK are highly expressed in mouse embryonic pluripotent stem cells, but levels decrease once differentiation has occurred [905]. This pattern was mirrored by levels of *CCNK*. Knockdown of CDK12 or CDK13 caused differentiation of the stem cells. They showed that Cyclin K associated with CDK12 and CDK13 proteins and that knockdown of these two

kinases also resulted in differentiation but via different gene pathways. This was replicated by inducing differentiation with CCNK knockdown. Dia et al. suggest that these two kinases are important in maintaining the self-renewal properties of embryonic stem cells [905]. This supports a role for CDK13 in early embryogenesis. Loss of *Cdk13* in the null mice could result in early differentiation and abnormal embryogenesis. The different stages of embryonic lethality between *Cdk12* and *Cdk13* may be a consequence of the different pathways each regulate.

It is not clear why there is such variability in the gestation at which *Cdk13*^{tm1b/tm1b} embryos died. I did not specifically assess for a cause of death, however Novakova et al. stated heart failure was the cause in their mice [885]. Whilst I did not specifically look for features of cardiac failure, given some embryos died early on and others were alive at E15.5, it is unlikely that there is a single defect or mechanism to explain the cause of death in all embryos. However, variable cardiac malformations could result in a variable time point of death depending on the effect on the foetal circulation. A *Cdk13*^{tm1b/tm1b} embryo found deceased at E12.5 appeared to have a looped heart, but further development had not occurred. The heart must have been able to provide adequate oxygenation up to this point, despite its severe abnormality. This is in contrast to the two surviving null embryos harvested at E15.5 that exhibited CAVCD, which in comparison had progressed much further with their development and must have functioned adequately for longer. The additional finding of reduced myocardium reported previously, may have compounded the reduced function of the abnormal hearts [885]. This could support the idea that more severe cardiac defects cause death earlier in gestation. In addition, there may be other organs affected by loss of *Cdk13* function, or other modifiers such as maternal or placental factors that alter the time point of death.

6.3.6.2 Forelimb Staging of *Cdk13*^{tm1b(EUCOMM)Hmgu} E15.5 Embryos

All the embryos collected at E15.5, were at S23+, or more advanced, with the exception of 6 heterozygous and 2 homozygous embryos which were at earlier stages of development (Table 6.8).

Gestation	Genotype	Litter	S22 -	S22	S22 +	S23 -	S23	S23 +	Average Stage (Mode)
E15.5	*Cdk13*^{+/+}	1	-	-	-	-	-	2	23+
		2	-	-	-	-	-	2	23+
		3	-	-	-	-	-	1	23+
		4	-	-	-	-	-	-	NA
E15.5	*Cdk13*tm1b /+	1	-	-	-	-	-	3	23+
		2	-	-	-	1	-	5	23+
		3	-	5	-	-	-	-	S22
		4	-	-	-	-	-	3	S23+
E15.5	*Cdk13*^{tm1b/tm1b}	1	-	-	-	-	-	-	NA
		2	-	-	-	-	-	-	NA
		3	-	-	-	-	-	-	NA
		4	-	-	2	-	-	-	S22+

Table 6.8 Forelimb staging of all *Cdk13*^{tm1b(EUCOMM)Hmgu} embryos collected at E15.5, split by litter.

Litter 4 was made up of *Cdk13*^{tm1b/+} and *Cdk13*^{tm1b/tm1b} mice. The null mice were delayed in comparison to the *Cdk13*^{tm1b/+} mice. Novakova et al. described delayed development in homozygous embryos [885], but we know that there is often variation in development of embryos within a single litter, which could account for these findings. The numbers are too small to generate reliable conclusions.

In *Cdk13*^{tm1b/+} mice, slightly delayed development compared to *Cdk13*^{+/+} embryos was seen in litter 3 mainly (Table 6.8). As stated previously, variation of developmental stage within a litter is well established, and the numbers studied here are too small to suggest whether the delay is related to genotype or not. Novakova et al. demonstrated abnormalities in heterozygous embryos, which could be a result of abnormal *Cdk13* product [885]. The ratios of the different genotypes born suggest that there is no survival disadvantage in the heterozygote mice and that heterozygosity for a normal *Cdk13* product in mice does not hamper embryogenesis.

6.3.6.3 Crown Rump Length Measurements of *Cdk13*^{tm1b(EUCOMM)Hmgu} E15.5 Embryos

Crown rump length (CRL) was measured for 26 E15.5 embryos and is shown below in table 6.9 and figure 6.10.

Gestation	Genotype	Litter	CRL (mm)	Average CRL (mm) with stage
E15.5	*Cdk13*^{+/+}	1	14, 15	14.5 at S23+
		2	13, 14	13.5 at S23+
		3	13	13 at S23+
		4	-	-
E15.5	*Cdk13*^{tm1b/+}	1	14, 15, 15	14.6 at S23+
		2	10, 13,13, 13, 13, 13	10 at S23- 13 at S23+
		3	13, 13, 13, 14, 15, 15, 15	14 at S22 14 at S23+
		4	13, 14, 15	14 at S23+
E15.5	*Cdk13*^{tm1b/tm1b}	1	-	NA
		2	-	NA
		3	-	NA
		4	9, 11	10 at S22+

Table 6.9 Crown rump length (CRL) of all *Cdk13*^{tm1b(EUCOMM)Hmgu} embryos.

Results are split by litter and by genotype.
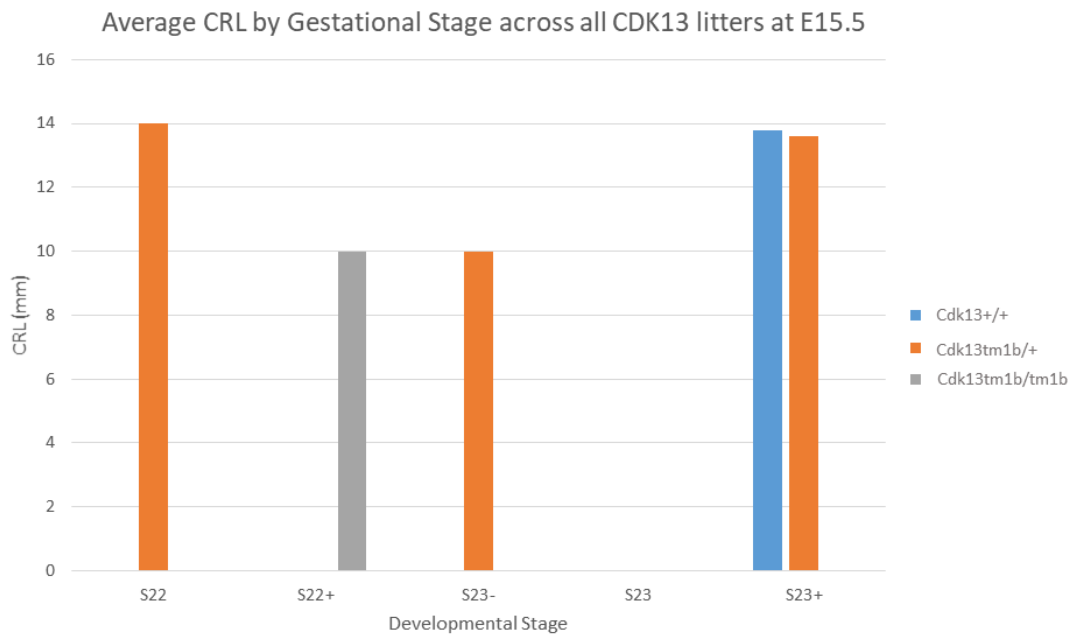
Figure 6.10 Average CRL across all _Cdk13_^{tm1b(EUCOMM)Hmgu} litters harvested at E15.5 by developmental stage.

We would expect CRL to increase with gestational age and stage of development. Reduced CRL can represent retarded growth. Data is only available for _Cdk13_^{tm1b/+} embryos at more than one stage, and does not follow this expected pattern (Figure 6.9). The lower CRL at S23- than S22 would not be expected, unless there were specific abnormalities in these heterozygotes not common to all _Cdk13_^{tm1b/+} mice. The numbers are too small to state anything conclusive from these results.

6.3.6.4 External Phenotype of _Cdk13_^{tm1b(EUCOMM)Hmgu} E15.5 embryos

I performed a gross assessment for general abnormalities at the time of embryo harvest (Table 6.10).

Gestation	Genotype	Alive/Deceased	Phenotype
E9.5	$Cdk13^{+/+}$ (n=0)	Alive: 0	NA
		Deceased: 0	NA
	$Cdk13^{tm1b/+}$ (n =9)	Alive: 9	Normal n=7 Abnormal n=2 slight developmental delay
		Deceased: 0	NA
	$Cdk13^{tm1b/tm1b}$ (n=0)	Alive: 0	NA
		Deceased: 0	NA
	Unknown (n=0)	Alive: 0	NA
		Deceased: 0	NA
E10.5	$Cdk13^{+/+}$ (n=2)	Alive: 2	Normal phenotypes
		Deceased: 0	NA
	$Cdk13^{tm1b/+}$ (n = 1)	Alive: 1	Normal phenotype
		Deceased: 0	NA
	$Cdk13^{tm1b/tm1b}$ (n=0)	Alive: 0	NA
		Deceased: 0	NA
	Unknown (n=3)	Alive: 0	NA
		Deceased: 3	Empty Decidua. Small uterus and sac, necrotic tissue only, oligohydramnios, discoloured amniotic fluid
E11.5	$Cdk13^{+/+}$ (n=1)	Alive: 1	Normal
		Deceased: 0	NA
	$Cdk13^{tm1b/+}$ (n=7)	Alive: 7	Normal
		Deceased: 0	NA
	$Cdk13^{tm1b/tm1b}$ (n=1)	Alive: 1	Normal
		Deceased: 0	NA
	Unknown (n=0)	Alive: 0	NA
		Deceased: 0	NA
E12.5	$Cdk13^{+/+}$ (n=2)	Alive: 2	Normal phenotype
		Deceased: 0	NA
	$Cdk13^{tm1b/+}$ (n =4)	Alive: 3	Normal
		Deceased: 1	Recent death, small and delayed development, smooth paddles and some retinal pigmentation
	$Cdk13^{tm1b/tm1b}$ (n=1)	Alive: 0	NA
		Deceased: 1	Recent death, small and delayed development, smooth paddles and some retinal pigmentation, heart looped but without any subsequent development

	Unknown (n=0)	Alive: 0	NA
		Deceased: 0	NA
E13.5	*Cdk13*[+/+] (n=1)	Alive: 1	Normal
		Deceased: 0	NA
	Cdk13[tm1b/+] (n=4)	Alive: 4	Normal
		Deceased: 0	NA
	Cdk13[tm1b/tm1b] (n=1)	Alive: 0	NA
		Deceased: 1	Necrotic contents within sac. Unable to stage embryo as no easily distinguishable features.
	Unknown (n=0)	Alive: 0	NA
		Deceased: 0	NA
E15.5	*Cdk13*[+/+] (n=6)	Alive: 5	Normal
		Deceased: 1	Empty Decidua. Necrotic tissue only, small placenta, uterine cavity and oligohydramnios
	Cdk13[tm1b/+] (n =21)	Alive: 20	18 Normal phenotype. 1 proportionately small but otherwise normal, 1 small and anteverted ears
		Deceased: 1	Empty Decidua. Necrotic tissue only, small placenta, uterine cavity and oligohydramnios
	Cdk13[tm1b/tm1b] (n=2)	Alive: 2	Normal. CRL 0.9cm, 1.1cm
		Deceased: 0	NA
	Unknown (n=12)	Alive: 0	NA
		Deceased: 12	Empty Decidua. Necrotic tissue only, small placenta, uterine cavity and oligohydramnios

Table 6.10 Gross external assessment of *Cdk13*[tm1b(EUCOMM)Hmgu] embryos based on gestation and genotype.

Homozygous embryos were found alive at E15.5.

Interestingly, the *Cdk13*[tm1b/tm1b] embryos that were alive at the time of harvest at E11.5 and E15.5 looked apparently normal. It is possible that they lack some of the more gross defects that result in earlier demise. A *Cdk13*[tm1b/tm1b] embryo

found deceased at E12.5 showed a looped heart, but minimal development after that. This might explain why the embryo had survived the earlier stages in utero, but failed when the heart could no longer provide adequate perfusion.

6.3.6.5 Phenotype of Adult *Cdk13*^{tm1b(EUCOMM)Hmgu} Mice

As part of routine care of the mice, they were checked for any gross abnormalities by external examination (Table 6.11). Compared to the *Prkd1* mouse line I also studied, the *Cdk13*^{tm1b/+} mice were subjectively smaller. A single *Cdk13*^{+/+} and single *Cdk13*^{tm1b/+} mouse appeared significantly smaller still. Microphthalmia and anophthalmia were recorded in three sibling *Cdk13*^{tm1b/+} mice. Whilst there are reports of strabismus in the human individuals with damaging *CDK13* mutations, this type of abnormality is associated with the C57Bl6 strain of mouse (https://www.jax.org/news-and-insights/1995/october/microphthalmia-and-ocular-infections-in-inbred-c57-black-mice) [891]. The fact that it was seen in three siblings also suggests that there may be another mechanism contributing here rather than *Cdk13*.

The staff in the Bio Support Unit noted that the *Cdk13*^{tm1b/+} mice were more nervous and "stressed" on arrival. The colony took much longer to settle than others I worked with. They were provided with extra environmental stimulation. Interestingly, an exaggerated startle response seen in heterozygotes in the IMPC data.

Gestation	Genotype	Phenotype
Adult	*Cdk13*[+/+] n = 47	Normal 46 Small 1
	Cdk13[tm1b/+] n = 52	Normal 47 Small 1 Abdominal swelling (renal cyst) 1 Micro/anophthalmia 3
	Cdk13[tm1b/tm1b] n = 0	NA

Table 6.11 Abnormalities seen in adult *Cdk13*[tm1b(EUCOMM)Hmgu] mice.

Abnormalities revealed during gross external assessment of adult mice as part of routine care.

There was higher than expected neonatal losses of pups with *Cdk13*[tm1b/+] mothers. The female mice would give birth normally, but the pups died within the first 24 hours. This was thought to be due to maternal neglect, as no milk could be seen within the pups' stomachs. Of the 28 pups born to heterozygous mothers, 20 died in the early neonatal period (*Cdk13*[+/+]=5, *Cdk13*[tm1b/+] = 9, unable to genotype =6). Usually the pups had been cannibalised which made phenotyping and genotyping difficult or impossible. This was a fairly consistent problem which meant that I had to use female *Cdk13*[+/+] mice and cross them with male heterozygous stud mice to continue the colony. The maternal neglect of newborn pups by heterozygous females could possibly represent some degree of neurodisability, but I did not carry out any assessment to confirm this.

Internal and external examination was carried out on all *Cdk13*[tm1b/+] female mice that were sacrificed as part of the procedure to harvest embryos. All were normal with the exception of one (n=1/10). I dissected the mouse and found a

large multiloculated cyst within the abdominal cavity on the left side (Figure 6.11). It was adherent to peritoneum, and I was unable to identify the left kidney in its normal position. I concluded it was most likely to be a renal based cyst and I was able to trace the ureter down into the pelvis as there were a number of cysts along it. The transverse colon was very full, probably due to an element of obstruction secondary to the mass.

Renal abnormalities have been identified in three human individuals (duplicated and dilated collecting systems, fused renal ectopia) and null mice have delayed kidney development and reduced numbers of glomeruli [833, 885]. Novakova et al. identified Cdk13 expression in the renal pelvis in adult mice; this could be relevant given that the cystic mass appeared to continue along the ureter, and the renal pelvis and ureter contain similar epithelial cells. Given that this was only seen in a single mouse, it is unclear if the cystic abnormality directly is related to *Cdk13* function. It would be informative to examine all adult *Cdk13*[tm1b/+] mice once they have been culled for similar abnormalities.

Figure 6.11 Large multiloculated cyst found in *Cdk13*^{tm1b/+} female mouse abdomen.

This was presumed to be renal in origin. Further small cysts can be seen along the course of what might represent the ureter.

In addition a single adult male mouse had to be culled due to an overgrown tooth and was examined internally and externally. No other abnormalities were found.

6.3.6.6 HREM cardiac assessment of E15.5 *Cdk13*^{tm1b(EUCOMM)Hmgu} Embryos

26 E15.5 embryos were collected and processed for cardiac phenotyping using HREM (Table 6.12). Two hearts were excluded from further analysis due to damage.

In the normal hearts there were a number of consistent features at E15.5 similar to those described by Geyer et al. [224]. The TV is sited slightly inferiorly to the MV, and the septal leaflet is adherent to the septum still. The apex of the heart is slightly notched, with a groove between the ventricles. The PV is superior and to the right of the AV. The pulmonary arteries arise after a short pulmonary trunk. There were no hearts with an immediate bifurcation and no common pulmonary artery, as described by Geyer et al. [224]. The ductus lumen was comparable to the feeding arteries with a normal insertion point. The RV was larger than the LV, with a thinner compact myocardium at the insertion point of the IVS to the anterior RV free wall. Examples of normal features are shown in chapter 4, section 4.5.6 figures 4.12-4.15.

No abnormalities were detected in the *Cdk13*[+/+] hearts (0/3). The majority of *Cdk13*[tm1b/+] mice were normal (16/19) and both *Cdk13*[tm1b/tm1b] hearts were abnormal (table 5.11).

Genotype	*Cdk13*[+/+]	*Cdk13*[tm1b/+]	*Cdk13*[tm1b/tm1b]
Normal heart	3	16	0
Abnormal heart	0	3 DORV and IAVCD (n=1, 8145C) Coronary artery abnormality (n=1, 8145B) Abnormal AV (n=1, 8145A)	2 CAVCD (n=2, 8144A, 8144B)
Total	3	19	2

Table 6.12 Summary of abnormalities detected in the *Cdk13*[tm1b(EUCOMM)Hmgu] E15.5 embryos by genotype.

DORV: double outlet right ventricle, IAVCD: Intermediate atrioventricular canal defect, CAVCD: complete atrioventricular canal defect. AV: Aortic valve

Figures 6.12 -6.20 demonstrate the abnormalities seen in *Cdk13*^{tm1b/+} mice. This included one mouse with IAVCD and DORV (8145C), one with a single origin of the coronary arteries (8145B), and one with an abnormal AV (8145A).
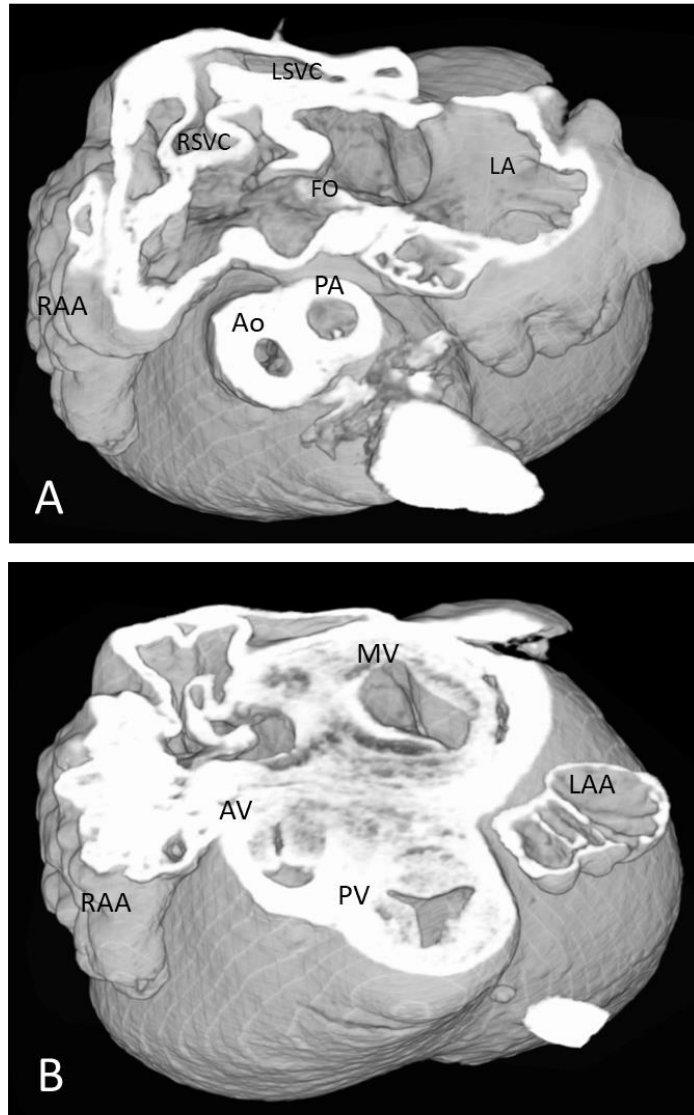
Figure 6.12. Axial cuts through Cdk13tm1b/+ mouse 8145C from 3D reconstruction.
A: Axial cut superior to aortic and pulmonary valves. Right and left SVC can be seen entering the right atrium. The foramen ovale, the aorta and pulmonary artery are shown. The aorta is situated further over to the right hand side than normal.
B: axial cut through aortic and pulmonary valves. The position of the aortic and pulmonary valves is not entirely normal. The aortic valve is more anterior and to the right than normal, compared to the pulmonary valve.

Ao: Aorta AV: Aortic valve, FO: foramen ovale, LA: left atrium LAA: left atrial appendage LSVC: left superior vena cava MV: mitral valve PA: pulmonary artery, PV: pulmonary valve, RAA: right atrial appendage RSVC: Right superior vena cava
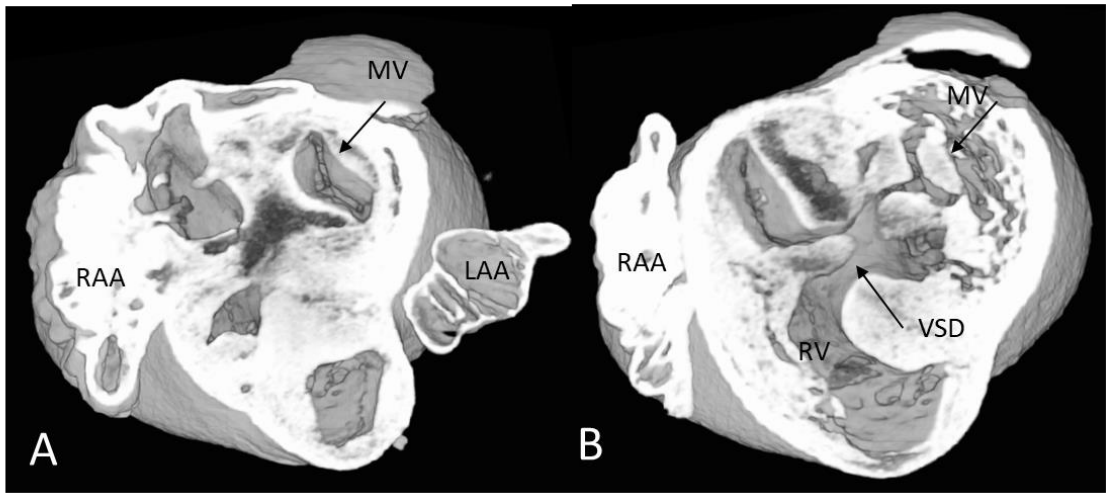
Figure 6.13. Axial cuts through Cdk13tm1b/+ mouse 8145C from 3D reconstruction.
A: Slice just below level of mitral valve annulus. Two commissures can be seen with a suggestion of a further commissure towards the IVS.
B: More inferior slice through mitral valve appears to show three leaflets to the valve. A VSD is also apparent.
LAA: left atrial appendage MV: Mitral valve RAA: Right atrial appendage, RV: Right ventricle, VSD: ventricular septal defect
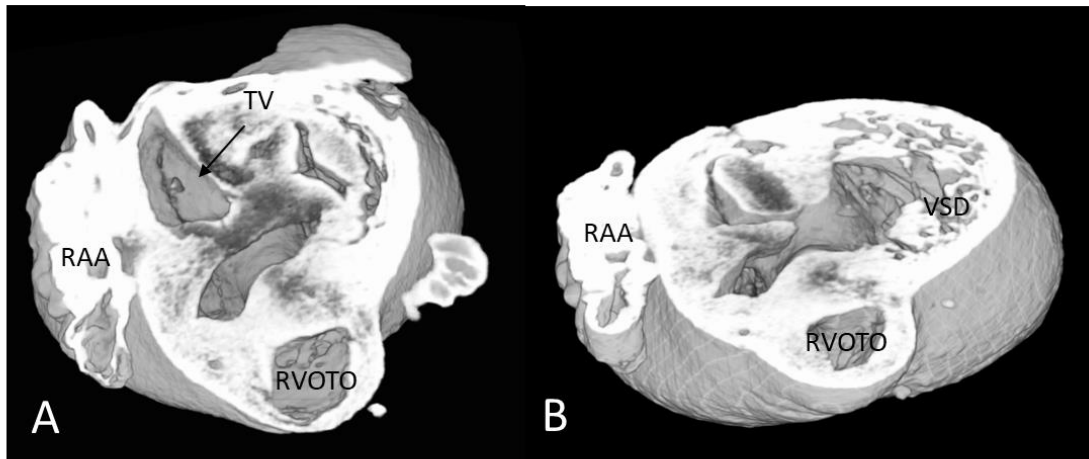
Figure 6.14. Axial cuts through Cdk13tm1b/+ mouse 8145C from 3D reconstruction.
A: Slice to demonstrate tricuspid valve.
B: Ventricular septal defect

RAA: Right atrial appendage, RV: Right ventricle, RVOTO: Right ventricular outflow tract, TV: Tricuspid valve, VSD: ventricular septal defect
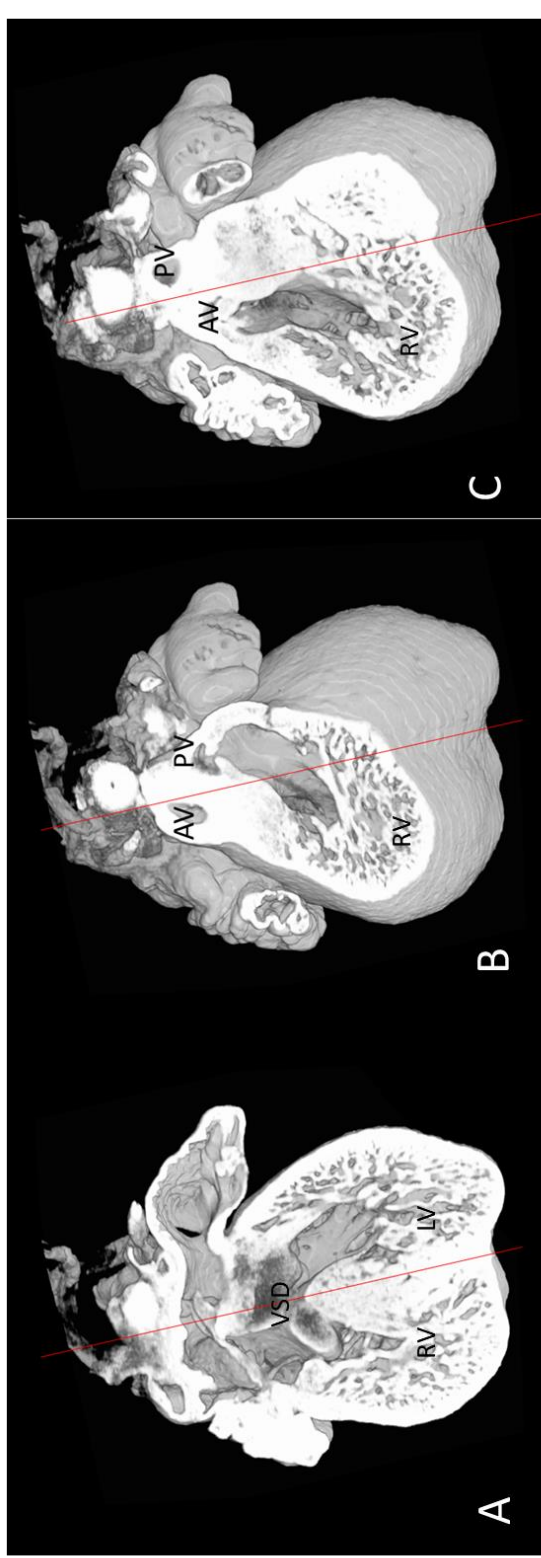
Figure 6.15. Coronal cuts through Cdk13tm1b/+ mouse 8145C reconstructed 3D reconstruction.
A: Axis through IVS at the site of the VSD. B: More anterior slice with same axis. C: Most anterior slice showing greater than 50% override of aortic valve.   AV: Aortic Valve, PV: Pulmonary valve, LV: Left ventricle RV: Right ventricle, VSD: ventricular septal defect
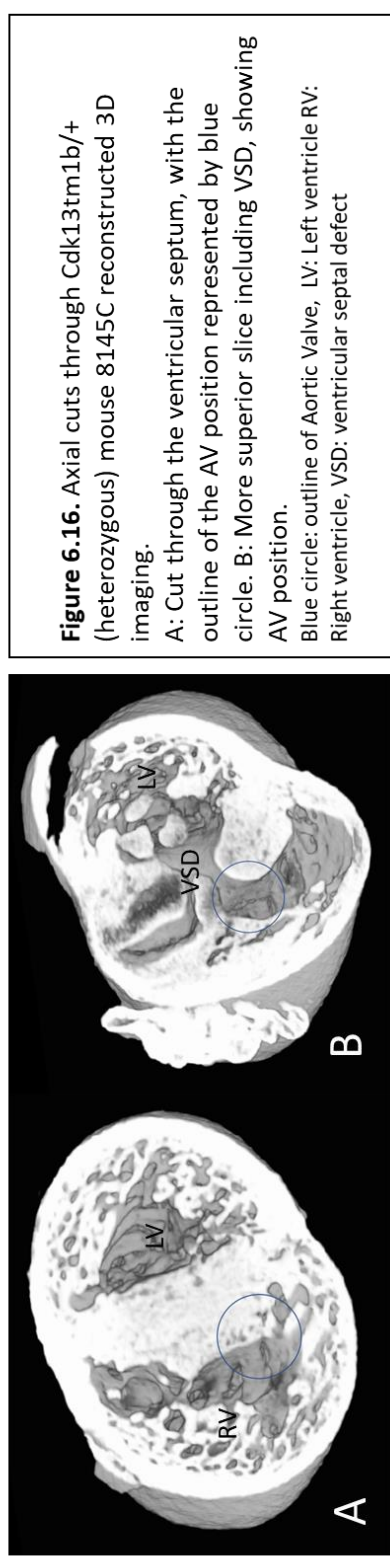
Figure 6.16. Axial cuts through Cdk13tm1b/+ (heterozygous) mouse 8145C reconstructed 3D imaging.
A: Cut through the ventricular septum, with the outline of the AV position represented by blue circle. B: More superior slice including VSD, showing AV position.
Blue circle: outline of Aortic Valve, LV: Left ventricle RV: Right ventricle, VSD: ventricular septal defect

Figure 6.17. Axial cuts through Cdk13tm1b/+ mouse 8145C raw slice photos.

A: Inferior to D superior cuts demonstrating no mitral aortic continuity.

A: Abnormal MV shown with apparently three commissures. VSD also present.

B: MV and RVOTO visible.

C: Lower sections of AV and PV seen.

D: Slice through cusps of AV and PV

AV: Aortic Valve, PV: Pulmonary valve, LAA: Left atrial appendage, MV: Mitral valve, TV: tricuspid valve VSD: ventricular septal defect

One *Cdk13*^{tm1b/+} heart shows what appears to be an IAVCD with DORV (8145C, Figures 6.12- 6.17). A large VSD is seen with two separate atrioventricular valves. The TV looks relatively normal, but the MV has the appearance of a cleft valve. There is greater than a 50% override and relatively normal positioning of the AV and PV.
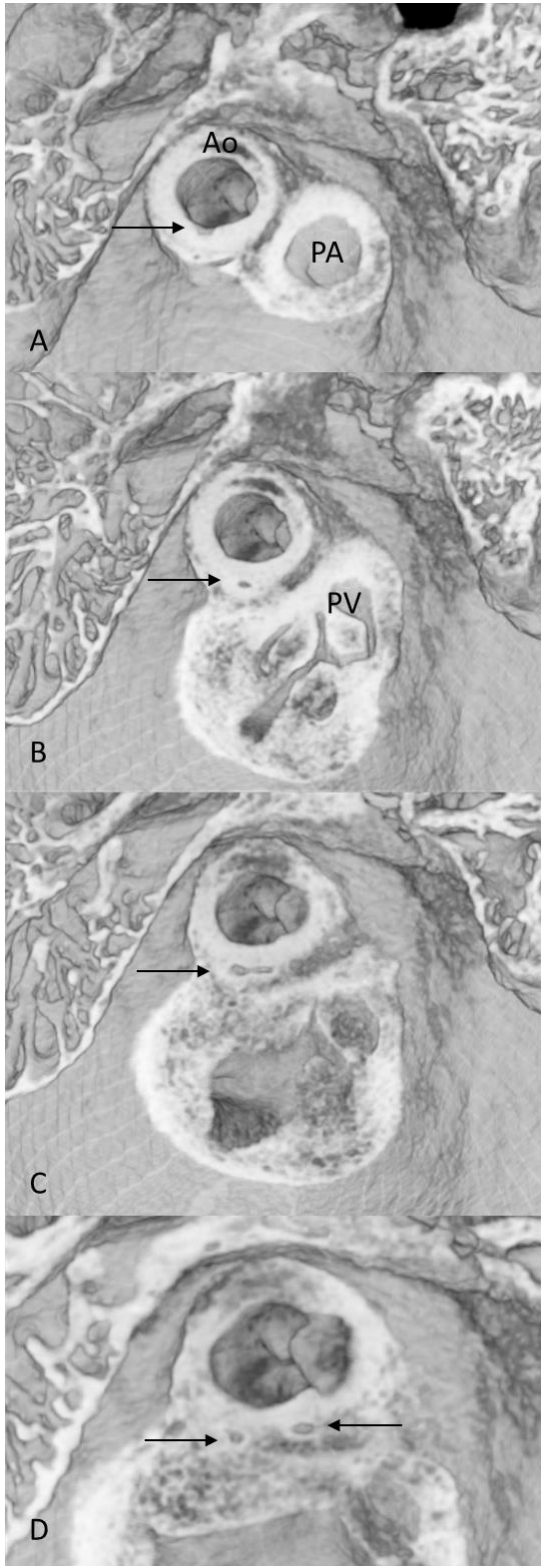
Figure 6.18. 3D reconstruction of E15.5 Cdk13tm1b/+ mouse (8145B). Single origin of both coronary arteries. It is not possible to follow the course of both arteries fully to distinguish if this could be a haemodynamincally significant lesion or not.

A. Axial cut through the aorta just superior to the valve at the origin of the coronary arteries. The arrow indicates a single coronary as it arises from the aorta.
B. Axial cut showing the single coronary artery moving anteriorly (arrow). The pulmonary valve leaflets are visible.
C. The single coronary artery is seen to bifurcate into left and right branches (arrow).
D. The two branches move to the left and right (arrows).

Ao: aorta, PA: pulmonary artery, PV: pulmonary valve.

A single *Cdk13*^{tm1b/+} heart demonstrated a common coronary artery trunk that appears the divide into left and right branches at the point of exiting the aorta (8145B, Figure 6.18). It arises from the right coronary cusp. This is the only heart identified with coronary abnormalities, all other hearts had a normal insertion of separate left and right arteries.
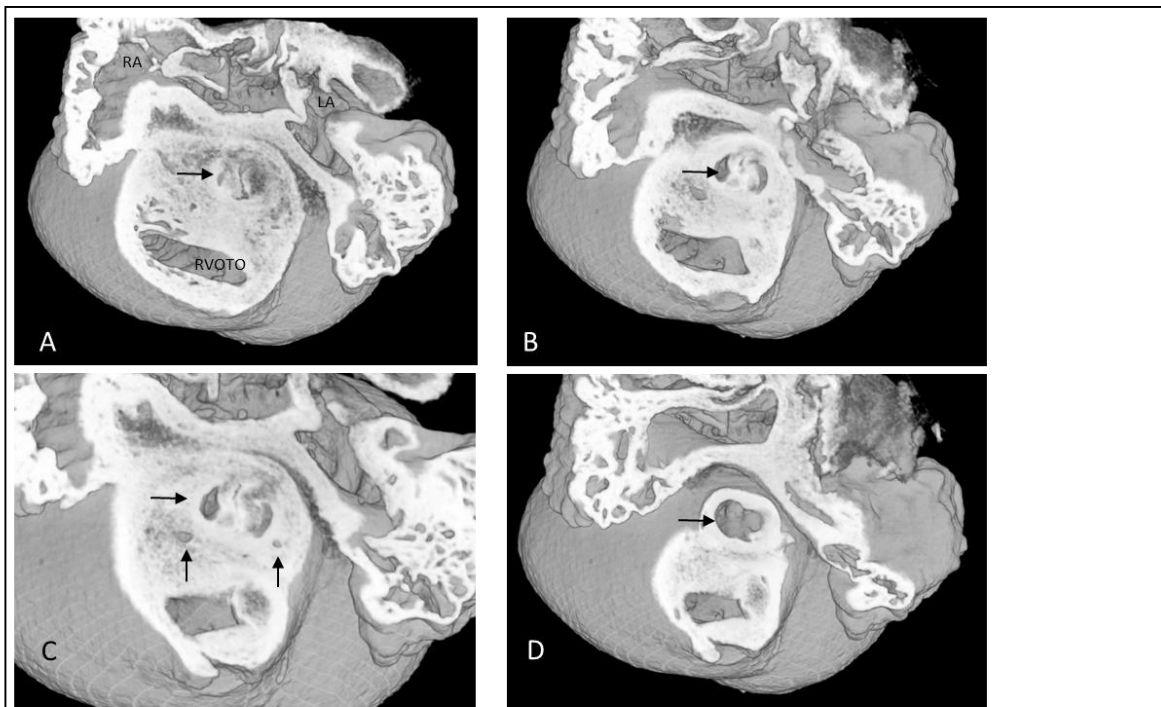
Figure 6.19. Axial 3D reconstruction of Cdk13tm1b/+ E15.5 heart (8145A) from inferior to superior, showing with possible bicuspid aortic valve or fused commissures. It is difficult to identify three distinct commissures and two leaflets may be fused.
A: Base of aortic valve, appearance of two cusps (black arrow).
B: Perpendicular cut through the valve leaflets (arrow).
C:More superior cut through valve leaflets. Coronary arteries seen (vertical arrows).
D: View looking down onto the valve (arrow).

LA: left atrium, RA: right atrium, RVOTO: right ventricular outflow tract
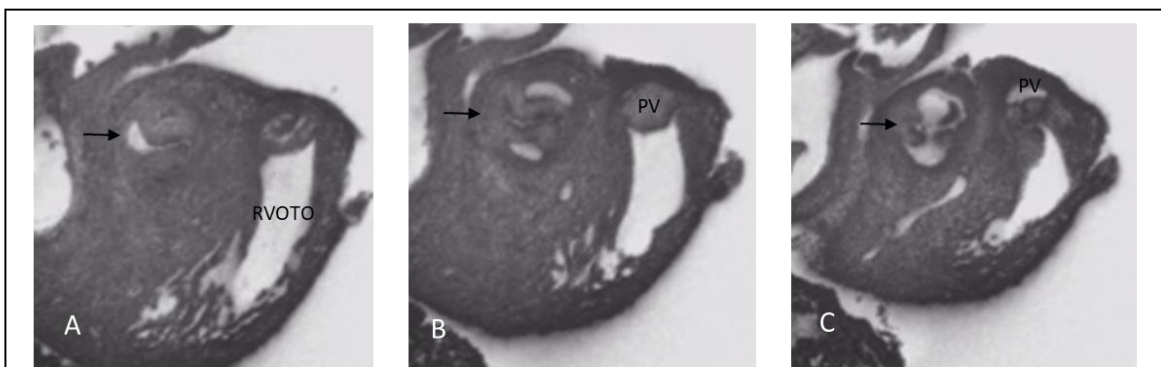
Figure 6.20. Axial cuts from raw slices, Cdk13tm1b/+ (8145A) E15.5 heart . Inferior to superior showing with possible bicuspid aortic valve or fused commissures (black arrows).
A: Axial cut through lower section of valve leaflets.
B: Image through middle of valve.
C: Cut through superior part of valve leaflets
PV: pulmonary valve, RVOTO: right ventricular outflow tract

An apparent abnormality of the cusps of the AV are seen in a *Cdk13*^{tm1b/+} mouse embryo (8145A, Figure 6.19 - 6.20). On static imaging it appears to have two leaflets, but does not look like a true bicuspid valve (type 0). Due to lack of real time imaging it is difficult to define the exact AV abnormality.

Only two homozygous *Cdk13*^{tm1b/tm1b} were found at E15.5 (8144A, 8144B). Both were alive at this gestation, and displayed cardiac abnormalities. Both have a common atrioventricular canal defect (Figures 6.21 – 6.28). It is difficult to identify 5 valve leaflets even on the raw images in one due to failure to remove blood adequately, but the defect does appear to be a common AV valve (mouse 8144B, Figures 6.27 – 6.28).
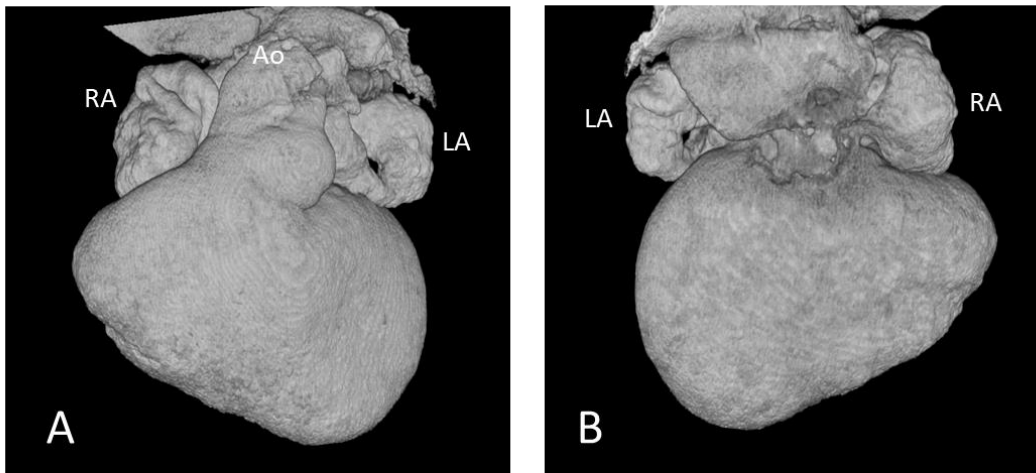
Figure 6.21. External view of 3D reconstruction of CDK13tm1b/tm1b E15.5 heart (8144A).

A. Anterior view. The truncus and area of the RVOTO and infundibulum is more prominent and abnormally angulated.

B. Posterior view.

Ao: aorta, LA: left atrium, RA: right atrium

Figure 6.22. 3D reconstruction of CDK13tm1b/tm1b E15.5 heart (8144A). Oblique cut through aortic and pulmonary valves and right atrium. Normal relationship of aortic and pulmonary valves. The PV is superior and just to the left of the AV. A single coronary artery is visible on this view (arrow to lumen). The pulmonary artery was confirmed to bifurcate normally.

AV: aortic valve, CA: coronary artery, PV: pulmonary valve, RA: right atrium
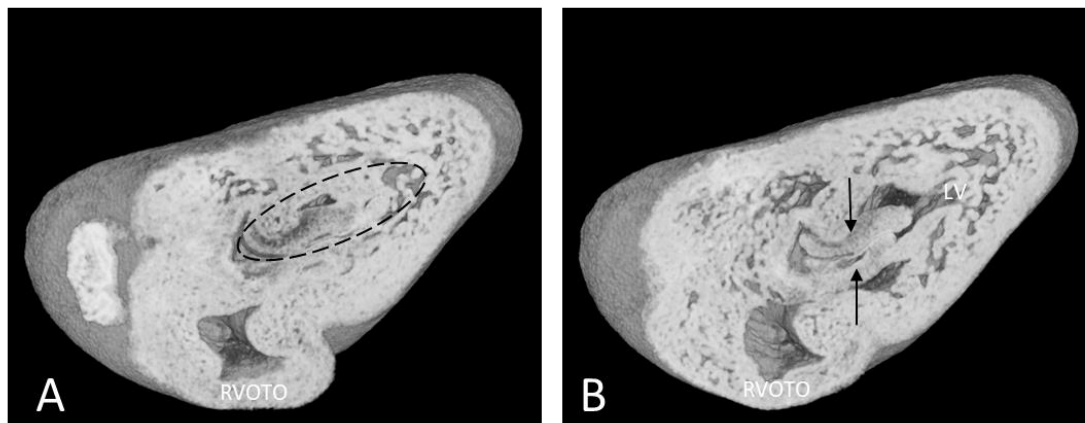
Figure 6.23. External view of 3D reconstruction of CDK13tm1b/tm1b E15.5 heart (8144A). Axial cut at level of atrioventricular valves.

A. A single atrioventricular valve is seen. The edges of the valve are marked by the hatched oval.

B. A more inferior axial cut to image A, illustrating what appears to be a superior and inferior bridging leaflet stretching across the heart (arrows).

Ao: aorta, LA: left atrium, LV: left ventricle, RA: right atrium, RVOTO: right ventricular outflow tract

Figure 6.24. 3D reconstruction of CDK13tm1b/tm1b E15.5 heart (8144A).

A. Sagittal section through the common atrioventricular valve showing the superior and inferior bridging leaflets (arrows). View is from the left side of the heart into the left ventricle.

B. Axial section through the ventricles at the site of the VSD (arrow). The lower part of the common AV valve is evident just above the VSD in the image.

LV: Left Ventricle, VSD: ventricular septal defect
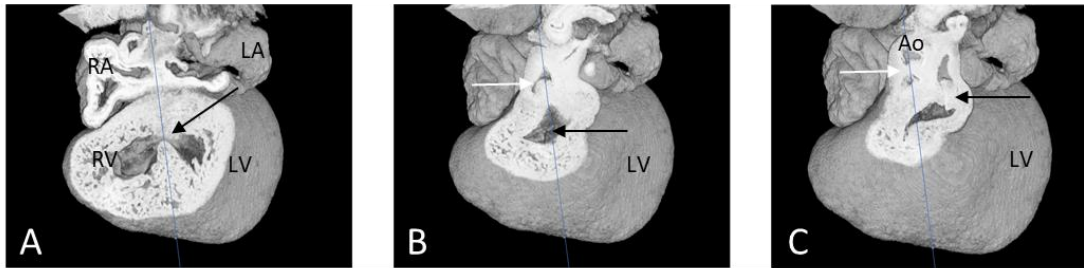
Figure 6.25 External view of 3D reconstruction of CDK13tm1b/tm1b E15.5 heart (8144A).

Coronal sections through the heart from the middle of the heart (A) to the more anterior aortic valve (C) to demonstrate the degree of aortic override. The blue line represents the axis of the heart through the long axis of the IVS.

A. The VSD is shown as marked by the black arrow.
B. Black arrow indicates VSD which is just visible through the window. White arrow indicates the aortic valve, which is bisected in the midline by the blue line, indicating 50% override.
C. White arrow indicates aortic valve and black arrow shows pulmonary valve which is usually set more superiorly. They appear to derive from a common infundibulum. It appears that there is 50% override which makes the delineation between DORV and overriding aorta more difficult.

Ao: aorta, IVS: interventricular septum LA: left atrium, LV: left ventricle, RA: right atrium, RV: right ventricle,
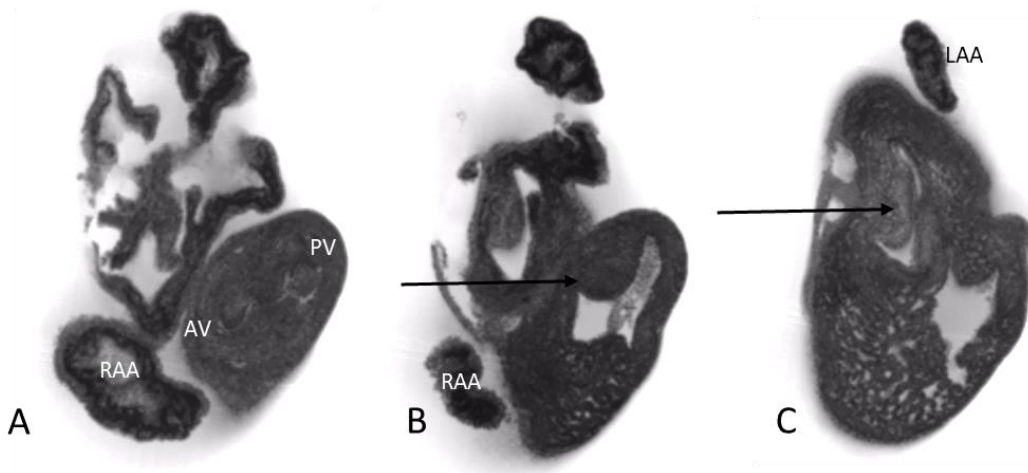
Figure 6.26. Raw section images of CDK13tm1b/tm1b E15.5 heart (8144A).

Axial sections through the heart.

A. At the level of the aortic and pulmonary valves.
B. Inferior to the aortic valve, through the common infundibulum. Loss of fibrous aortico-mitral connection (black arrow) suggesting DORV rather than just overriding aorta.
C. Through the common AV valve (black arrow).

AV: aortic valve LAA: left atrial appendage, PV: pulmonary valve, RAA: right atrial appendage

Cdk13 null mouse 8144A has a CAVCD and likely DORV (Figures 6.21-6.26). The fusion of the cusps is difficult to demonstrate on these images, however a number of segments can be visualised on video. Papillary muscles are visible on both sides of the septum. There is 50% override of the aorta. It is difficult to establish if there is any pulmonary stenosis given the static imaging and under-lysis.
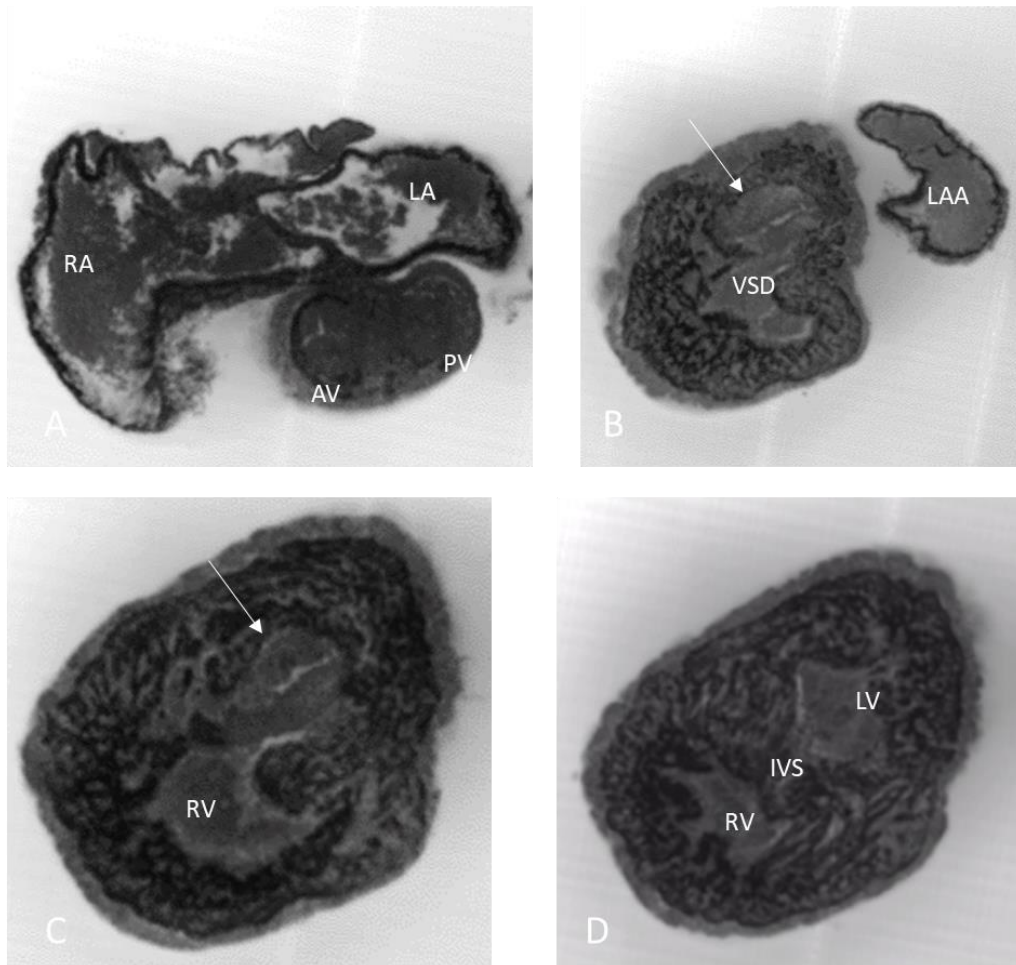
Figure 6.27 Raw slice images of the E15.5 mouse heart ID: 8144B. Axial slices from superior (A) to inferior (D).

Unfortunately the poor lysis and some damage seen in this heart mean that analysis is limited. However, it does appear to show a common AV valve, VSD and no overriding aorta.

A: Image showing atria and aortic and pulmonary valves and atrial septum. Extensive under-lysis is apparent and blood is visible in the atria.

B: Only a single valve is seen connecting the atria to the ventricles (arrow). A VSD is also apparent.

C: More inferior cut showing common valve and VSD.

D: IVS is present in the lower section of the heart.

AV: Aortic valve, IVS: interventricular septum, LA: Left atrium, LAA: Left atrial appendage, LV: Left ventricle PV: pulmonary valve, RA: right atrium, RV: right ventricle, VSD: Ventricular septal defect.

Figure 6.28 Raw slice images of the E15.5 mouse heart 8144B. Axial slices from superior (A) to inferior (C). Unfortunately the poor lysis and some damage seen in this heart mean that analysis is limited. However, it does appear to show a common AV valve, VSD and no overriding aorta.

A: Image showing atria and aortic valve. The site of the aortic valve is shown in the blue circle on the next images, to illustrate that there is no over-riding.

B: Blue line demonstrates the axis of the interventricular septum.

C: More inferior cut showing that the aortic valve does not over-ride at all.

AV: Aortic valve, IVS: interventricular septum, LA: Left atrium LV: Left ventricle RA: right atrium, RV: right ventricle.

Cdk13 null mouse 8144B also appears to demonstrate a CAVCD (Figures 6.27-6.28). It is a difficult heart to interpret as there is extensive under-lysis and some damage. However, only a single AV valve is seen along with a VSD which leads me to suggest this is a CAVCD.

Whilst specific *CDK13* expression patterns in the human and mouse foetal heart are still unknown, there is low expression in the adult human heart including the atrioventricular valves [886]. The BGEE database confirms expression of *CDK13* in the human foetus at three and five months post conception, but the heart is formed by this point and earlier expression information is not available [903]. From embryonic hearts collected by myself, T Ghosh in the Brook lab has shown that CDK13 is expressed in the heart at E11.5, E13.5 and E14.5 using western blotting. This provides support for the role of *CDK13* in cardiogenesis.

Both of the *Cdk13*tm1b/tm1b embryos isolated exhibited cardiac abnormalities. Both had a complete atrioventricular canal defect and one had DORV, which is not an uncommon association. The other heart does not appear to demonstrate any over-riding of the aorta, however analysis of this heart was difficult due to extreme under-lysis resulting in some additional damage. Both of these embryos were stages S22 and S22+, when it is uncommon to find an interventricular communication so these findings are likely to represent a true abnormality [224].

Given AVCDs have been seen in a *Cdk13*tm1b/+ and *Cdk13*tm1b/tm1b mice, *Cdk13* may play a role in endocardial cushion development in the mouse. This is a different type of abnormality to that seen in the human individuals with heterozygous *CDK13* mutations. AVCD is not seen, and instead septal defects predominate. A single individual has been reported with TOF, which can be seen in combination with AVCD [906].

From the known protein interactions with CDK13 [549], there are no genes known to be specifically linked with endocardial cushion formation in the mouse (http://www.informatics.jax.org/go/term/GO:0003272) or AVSD in the human https://omim.org/phenotypicSeries/PS606215. CDK13 is thought to interact with MMP21, which has a role in left right asymmetry and is thought to be a negative regulator of NOTCH[549]. This may be important given the association between AVSD and heterotaxy, and the important role of NOTCH signalling in left right patterning and the heart [907].

The range of defects seen in the *Cdk13*[tm1b/+] embryos is more diverse. The first mouse had a single coronary artery arising from the aorta, which then bifurcates into a left and right coronary artery (8145B). It is difficult to follow the course of this to decide if it is likely to have had any haemodynamic impact or not. A solitary coronary ostium in the aorta is seen in 6% of C57BL/6N mice so it is likely that this is not related to loss of *Cdk13* [908].

The second *Cdk13*[tm1b/+] embryo had an intermediate AVCD and DORV. The atrial septum appeared deficient anteriorly, and there is a VSD. This embryo was stage S23+ indicating a low chance of interventricular connection (2%) [224]. The mitral and tricuspid valves appeared to have fused anterior and posterior bridging cusps and there was reduced trabeculation in the left atrium. The third heart exhibited an abnormality of the aortic valve, possibly related to fused commissures (8145A).

There are a few similarities between the CHD seen in the two species, in those heterozygous for normal *CDK13*. One individual with the commonest mutation p.N842S has BAV as does one of the mice [833]. However, the reported individual has LVNC in addition to BAV. An individual with a duplication with a breakpoint in CDK13 has BAV, although the significance of this CNV is uncertain [522]. Another *Cdk13*[tm1b/+] mouse has an abnormal origin of the coronary arteries and moderate *CDK13* expression was seen in the coronary arteries of human post-mortem samples [887].

All three of the $Cdk13^{tm1b/+}$ mice with abnormalities described above were from the same litter. Given no abnormalities were seen in other heterozygous mice, we should consider whether this may be the result of another genetic or environmental factor, in addition to the abnormal $Cdk13$ product. Novakova et al. did find abnormalities in 2-8% of $Cdk13^{tm1a/+}$ embryos (hypervascularisation, nuchal oedema and pericardial oedema) but they did not specifically evaluate for structural CHD [885]. Further heterozygous mice should be examined.

One factor that could result in different lesions in the mouse and human heart is length of gestation. In the human there is a much longer period between completion of heart structure and birth. Despite the differences in CHD types seen, there are some shared features between humans and mice including palate and renal abnormalities. Submucous clefts have been reported, as have renal abnormalities [833, 835]. However there is no evidence of lung or liver abnormalities in the human population. Further clinical reporting as more individuals are diagnosed may provide more information.

6.3.6.7   Comparison with $Cdk13$tm1a/d mouse

Whilst the cardiac assessments across the $Cdk13$tm1a/d and tm1b lines are not comparable, there is evidence of cardiac abnormalities in all. Novakova et al. considered the tmb1a allele to be hypomorphic, and more severe abnormalities were seen in the tm1d allele mouse [885]. The tm1a allele produces a truncated protein and they suggested that the milder phenotype could occur for two reasons. The first is that the neomycin cassette is in a non-coding region that does actually have some influence on the expression of $Cdk13$. The second is that an abnormal but functional protein is produced as a result of post-transcriptional exon shuffling, which is known to occur at very high frequencies in CDK13 [909]. Certainly it is known that the tm1a allele may not be a true knockout and can be phenotypically hypomorphic or wild type

().

The tm1b allele is considered to be a true knockout so I would expect a more severe phenotype than the tm1a mouse, and comparable effects in the tm1d mouse. However, I did not study the extra-cardiac parts of the embryo in sufficient detail to compare the incidence of palatal or hepatic abnormalities. Novakova et al. have used different methods to evaluate the heart, looking at histology and function whereas my analysis has focussed on structural malformations and myocardial measurements could not be taken [885]. It would be interesting to perform histological analysis of the mouse line I have studied to see if there is also a reduction in myocardiocytes and to assess function of the heart in both null and heterozygous mice.

6.4 Limitations and Future Work

CHDFIDD/*CDK13*-realted disorder is a newly described syndrome, so it is unlikely that we have seen the full mutational and phenotypic spectrum. All individuals in Sifrim et al. were recruited through either DDD or the undiagnosed diseases network, and therefore we should consider the possibility that there may be a degree of ascertainment bias in the phenotypes collected [12]. Certainly for recruitment to DDD, developmental delay will be reasonably significant or the patient will have multiple congenital abnormalities, so we could have missed a milder *CDK13* phenotype.

The natural history of CHDFIDD/*CDK13*-realted disorder is unknown. The majority of the cohort with *CDK13* mutations is still young, and time will reveal the natural history of this condition. Important questions remain such as long term prognosis, and whether there is any risk of malignancy. Reporting of progress and complications in these individuals will provide clinicians with more information to pass onto affected individuals. Van den Akker et al. set up a website where any individuals can be reported to help delineate the clinical

phenotype (www.humandiseasegenes.com) and there is an active Facebook group.

Although some individuals, such as those reported by Bostwick et al. did have echocardiograms to exclude structural heart disease, many others will only be offered this investigation if there are concerning signs or symptoms [833]. Until echocardiography at diagnosis becomes routine, the prevalence of CHD and full range of abnormalities may not be apparent. Certainly BAV might not present until later life, and it is currently unclear if this is part of the spectrum of CHD caused by mutations in *CDK13* or not.

One of the main limitations of this work is the lack of expression and protein studies in the mouse, including assessments of kinase activity. I have collected a number of tissues from embryonic and adult mice for protein and expression studies in the heart at various time points by other members of the Brook Laboratory. Further functional work is also being carried out in the Brook lab, including phosphorylation assays.

The number of mice I have analysed here is small. There are further embryo hearts I collected that are embedded, but not yet sectioned. Larger numbers of embryos are required for meaningful statistical analysis and to be sure of the occurrence of CHD in the mice given several affected embryos were from the same litter. Ideally a larger WT cohort would be analysed. Fortunately, Geyer et al have already analysed the same strain at the same stage so we can use this as control data[224].

The majority of humans with mutations in *CDK13* and CHD, have an ASD (60%). In the embryonic mouse it is difficult to diagnose ostium secundum defects given we expect patency of the foramen ovale. Lack of dynamic imaging means flow, the flap and orifice size cannot be assessed. A larger PFO might suggest a secundum ASD, but I think it is impossible to tell at this stage, especially when

we cannot assess for additional sequelae such as pulmonary hypertension. Lack of dynamic imaging also means other lesions including VSD and valve abnormalities might be missed.

HREM itself requires use of fixatives and dehydration of the samples. There is potential for shrinkage and damage to structures during this process, as there would be with histological analysis. Manually transferring delicate embryo hearts during the embedding process is also difficult and I have tried to minimise the possibility of damage as much as possible. All hearts were assessed for damage on HREM imaging and those that had damage making the images uninterpretable were removed from further analysis.

Examination of the mouse heart at different time points in embryogenesis and correlation with development and evidence of heart failure may help us understand why null mice died at different stages. Additional whole embryo HREM would allow assessment of the whole mouse to identify extra-cardiac abnormalities associated with loss of *Cdk13*. Ideally dynamic studies such as echocardiography would be carried out in addition. Given the irregular heartbeats seen in the Cdk13tm1d embryos, and an individual with arrhythmias possibly secondary to BAV and LVNC, ECGs in the mice may also be informative.

More detailed *Cdk13* expression within the heart may add support to the development of CHD in mice and individuals with *CDK13* mutations. Expression within the endocardial cushions would be especially interesting. I noted any gross placental defects in the embryos I harvested, but other additional work could include formal histology of the placenta. Abnormal function can result in growth restriction and embryo death and might be an important factor in demise of the embryos.

In addition to the *Cdk13* knockout mouse, the WTSI attempted to produce a mouse with a humanised mutation, but unfortunately this was not successful.

The mutation pAsn842Ser was chosen as this was the most commonly reported mutation (9 individuals with CHD, 9 without). It was thought that the CRISPR process was too efficient and this humanised mutation mouse could not be produced. An application was submitted to the Genome Editing of Mice for Medicine Programme (GEMM, MRC Harwell) for this same mutation and p.G717R. This is the second most commonly reported mutation (four individuals, one with CHD). We have secured one mouse and have been asked to apply for another.

As yet, the mechanism or mechanisms by which *CDK13* mutations result in abnormalities is unknown. Gain of function, haploinsufficiency and loss of function, and dominant negative mechanisms have been suggested. This makes it difficult to interpret mutations that have been reported in other areas of the gene, and there are a number of mutations reported on Decipher and ClinVar that are of uncertain significance. Functional work will help with interpretation.

The role of *CDK13* in the context of acute myocardial infarction may also be of interest in the future, given its expression in the adult heart and evidence that miRNAs that target *CDK13* are dysregulated in this situation [910]. Its role in progenitor cell differentiation also makes it an interesting candidate for regenerative medicine. Potentially it could it play a role in the damaged and failing adult heart.

6.6 Conclusions

CHDFIDD/*CDK13*-related disorder is a recognisable syndrome. The phenotype is likely to be delineated further as more individuals are reported. The precise mechanism(s) as to how heterozygous mutations cause clinical abnormalities in humans is still unclear. Functional studies are required.

CHD is reported in around 40% of these individuals. ASDs predominate. Mutations in the protein kinase domain before the start of the active site, affecting the nucleotide binding domain, the ATP acceptor site and the PITAIRE locus are associated with CHD. The risk of CHD with LoF mutations and those outside the protein kinase domain is uncertain currently.

$Cdk13^{tm1b/tm1b}$ mice are embryonic lethal, with death occurring up to at least E15.5. Given that this mouse and the $Cdk12$ null mouse are embryonic lethal at different time points, it is likely that despite both binding with CCNK and their structural similarities these genes are non-redundant in embryogenesis. It is likely that homozygous mutations affecting $CDK12$ and $CDK13$ function in humans results in embryonic lethality as seen in mice.

Endocardial cushion type defects are seen in $Cdk13^{tm1b/tm1b}$ mice. $Cdk13$+/- mice may also exhibit CHD. The large number of apparently healthy heterozygotes born suggests that there is no survival disadvantage. Further more detailed phenotyping in null and heterozygous mice including assessment for neurodevelopmental disability may provide further information.

Genetic testing for $CDK13$ should be offered to individuals with ASD and NDD, or any of the other associated congenital malformations that have been reported or similar facial features. Clinical recommendations should be updated with continuing follow up of affected individuals and information from newly diagnosed families.

The initial individuals reported in Sifrim et al. suggest that pathogenic missense mutations in $CDK13$ result in a recognisable clinical syndrome [12]. Since then, further reports have been published allowing us to refine the phenotype. Whilst CHD is seen in just less than 40% of affected human individuals, our mouse model suggests that CHD is a feature of abnormal function of $CDK13$. The

mechanism by which alterations in *CDK13* produce clinical abnormalities is not yet understood.

Chapter 7 Confirmation and Feedback of results to participants and associated ethical issues

7.1 Introduction

WES (whole exome sequencing) studies in cohorts of individuals with CHD has proved invaluable for identifying new CHD genes [12, 14, 16]. WES and whole genome sequencing (WGS) generate vast amounts of data, much of which will not be relevant to CHD. Whilst this approach works well for gene discovery, it does leave the researcher to consider which other results should be fed back to participants. More specifically, should we feedback results in clinically important genes that are separate to the study question? These results are often referred to as secondary, or incidental findings.

Participants involved in the WES study described in chapter 3 [12] had been recruited over the last 20 years. Participants had the option to opt in or out for feedback of pathogenic results relating to the cause of their CHD. At the time that participants were consented to join the study, use of WES on a large scale was not possible. Therefore, a discussion about secondary findings was not included in the consent form. This meant that we had to consider whether in addition to pathogenic CHD mutations, we should look for secondary findings. This is a complex issue for many reasons, primarily because of the benefits and disadvantages of feeding back secondary findings varies depending on the clinical consequences of mutations in that specific gene. In this chapter I consider the latest guidance regarding secondary findings for clinicians and researchers, and illustrate the confirmation of pathogenic results that were fed back to the NHS clinician for individuals in our WES study [12].

7.1.1 Duties of Clinicians and Researchers

Historically, genetics research and healthcare have been separate entities.

Usually research is driven by the aim of increasing our knowledge about a subject. Researchers may have little or no contact with the participants and traditionally have not had a duty of care to those individuals. Health care and clinicians aim to benefit their individual patient, work closely with them and have clear duty of care guidance. More recently, healthcare and research have been successfully integrated through large projects such as the DDD Study [522].

Projects like the DDD study [496], and 100 000 Genomes Project [911] incorporate research into healthcare by recruiting participants from, and feeding results back through the NHS. In the future it is likely that more and more NHS patients will have WES and WGS, and have their genetic data included in repositories for use in research [912]. Whilst these projects will benefit researchers, clinicians and patients, it is important that each group's responsibilities are clear. Individuals in a study may believe that the researcher is specifically working to improve the care of that individual, and not be driven by other motives. This is termed therapeutic misconception [913]. In addition, it is necessary to make the distinction between healthcare and research clear to participants as failure to do this can result in study participants recalling research as clinical care [914].

Clinicians in research are still bound by the General Medical Council guidelines and have a duty of care to the patients involved in research (Good practice in research and consent to research, 2010 http://www.gmc-uk.org/guidance/research accessed 21.3.13, and Patients and doctors making decisions together 2008 http://www.gmc-uk.org/static/documents/content/Consent_-_English_0911.pdf accessed 21.3.13), and it has been suggested that they should be expected to spot and deal with incidental findings at a level expected of their training [915]. Researchers also have professional obligations under the Helsinki and Nuremburg Declarations, especially in cases where clinical trials form part of medical practice [916, 917] and in some countries they have a legal duty of care [918]. With respect to genetic studies, it has been suggested that a researcher's

duty of care will be passing genetic information to the appropriate clinical team after consultation with a clinician, and that the principle investigator is ultimately responsible for feeding back incidental findings [915].

7.1.2 Secondary Findings generated by WES

Genetic testing is becoming more complex and single genes tests are no longer the norm. Virtual panels of genes, whole exome and whole genome sequencing are now more common place and can identify pathogenic mutations for disorders including intellectual disability in 20-30% of individuals, and up to 57% of trios [919, 920] (and https://www.exeterlaboratory.com/test/exome-sequencing-services). Whilst this is beneficial in terms of increasing diagnostic rates, it means that the possibility of identifying a secondary finding is around 1-7% [921].

Secondary findings cannot be considered as a single entity (Figure 7.1). Pathogenic variants found will represent a wide spectrum of clinical consequences. Examples include a mutation in a tumour suppressor gene or being identified as a carrier of a Cystic Fibrosis mutation. Reduced penetrance and variable expressivity of genes complicates the matter further. Feeding back a result to an individual who is completely asymptomatic and may never develop the disease, puts the individual at risk of harm from increased anxiety, unnecessary and potentially invasive medical tests, screening and risk management [922] and may not be in their best interests.
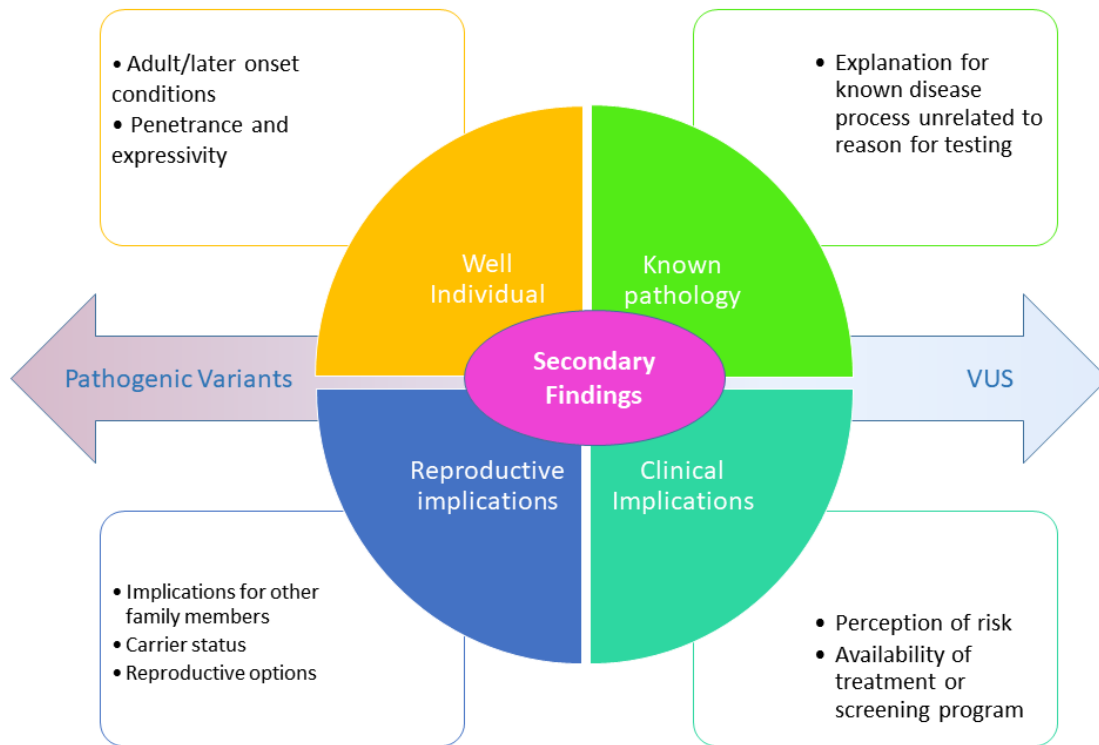
Figure 7.1 Examples of secondary findings.

This might include known pathogenic mutations, or variants of uncertain significance (VUS). In a well individual, mutations may be found in genes that are associated with disease at a later time in life.

There will also be a number of variants of uncertain significance (VUS) detected. Determining the pathogenicity of a VUS is difficult as many genetic conditions are rare and there may be limited information about the genetic mechanism causing disease. This wide spectrum of results makes it very difficult to devise a single guideline to deal all of these possible results, and to discuss them with research participants and record their wishes. On a research basis, sequencing has identified pathogenic mutations in 6-14% antenatally and in neonates and paediatric individuals with CHD. However, a VUS was also identified in 26 - 80% [537, 923]. Whilst this test may be advantageous in terms of making a diagnosis (particularly in S-CHD), the significant chance of identifying a VUS should be considered and form part of the informed consent process.

Discussing the implications of a secondary finding is complicated by variable penetrance and expressivity. Even more difficult is the situation when there is no proven screening or treatment, such as in conditions such as the cancer predisposition Li Fraumeni Syndrome, and certain neurodegenerative conditions. Some lower penetrance conditions have suggested screening, although many people will remain well and screening may therefore ultimately prove to be unnecessary (*SDH* gene cancer predisposition syndromes). Some genetic findings have implications for reproduction, including being identified as a carrier. Counselling and testing options will depend on carrier prevalence within the population. In all of these situations, we also need to consider whether there might have also been transmission within the family and whether others are at risk.

Central to the issue of feeding back secondary findings is informed consent. It is a legal requirement for not only the collection of tissue, blood and saliva, but also the storage and use of genomic data. Gaining fully informed consent to receive secondary findings is difficult because of the number of outcomes that must be discussed including carrier status, reduced penetrance, insurance and those discussed above. In the relatively short time in a genetics clinic, or during consent for research, it is challenging to discuss all of this in sufficient depth to gain fully informed consent for all possible findings. It has been suggested that this would requires six hours of discussions for consent and 5 hours of direct discussion of the results [924, 925]. Clearly this is not feasible.

A number of strategies have been suggested to help minimise the difficulties of dealing with secondary findings. This includes lists of important genes where mutations should be fed back, and simplifying choices by grouping genes into classes or "bins" and making a decision about each group [926]. Some of these recommendations for dealing with secondary findings are discussed further here. Although mainly directed towards clinicians, they are important for researchers too. Whilst there is no UK law regarding the duty to provide secondary findings or current test case, it seems prudent for all researchers carrying out untargeted

testing to consider their duty to their research participants and how they might deal with these results.

7.1.3 Current guidance regarding secondary findings

7.1.3.1 Guidance for Clinicians

The American College of Medical Genetics and Genomics (ACMG) have recommended review of sequencing results, for a specific set of genes that have been identified as potentially important. They recommend feeding back class 4 and 5 variants, in 56 genes [927]. They are cancer predisposition, cardiomyopathy, arrhythmia and aortopathy genes in the main, which would require increased medical care. However some of these genes exhibit reduced penetrance and variable expression. Unless the patient has specifically opted out of feedback of these results, the ACMG recommends that these results should be passed onto the patient [928].

The European Society for Human Genetics (ESHG) takes a different approach to that of the ACMG. Their guidance aims to minimise the production of secondary findings. Rather than using all WES and WGS data generated, they suggest targeted testing of relevant genes initially. Only if no genetic explanation for the individual's symptoms has been found, should untargeted sequencing be used [929]. If the individual concerned has opted in for feedback of secondary findings then the health care practitioner should report results that cause serious health problems as defined by the ESHG [930].

7.1.3.2. Guidance for Research

The MRC Framework on the Feedback of Health-related findings in research provides useful and practical advice around dealing with secondary findings for researchers https://wellcome.ac.uk/sites/default/files/wtp056059.pdf . The

guidance suggests that incidental findings should be considered on a case by case basis. The cost of feedback, subsequent tests and extra clinical interactions in the NHS must be in proportion to the benefits gained by the participant. The benefits of feeding back a result should outweigh the disadvantages, and a pathway to do this should be factored in to any research plans. The consent process must be designed to include information about feedback of secondary results. This should include whether they will be available or not, the scope of the feedback, whether their primary healthcare provider will be involved, and whether they can decide which results they would like.

7.1.3.3 Current Examples: The 100 000 Genomes Project

Recent guidance for feedback of significant results has been developed for the 100 000 Genomes Project [931]. Participants can opt in or out of receiving secondary findings, defined as mutations that cause serious and life threatening medical conditions for which there is an option of prevention or treatment available within the NHS. A short list of high penetrance cancer predisposition genes, and familial hypercholesterolaemia genes has been compiled. *CFTR* is the only gene for which carrier status is offered as an option for feedback. A pathway is in place to pass findings through to an NHS clinician who will relay the results and organise necessary clinical interventions. With respect to children, only childhood onset conditions will be looked at, in line with current NHS practice. VUS are withheld, but clinical judgement can be used and clinicians can feedback a VUS if more investigation is planned to determine its significance. The 100 000 Genomes project experience will be a helpful model for new genetics based research studies.

7.2 Aims

The aims of this chapter are to:

- Identify and confirm pathogenic mutations identified in individuals with CHD recruited locally and feed the results back to the families.
- Determine how to deal with secondary findings.

7.3 Results and Discussion

7.3.1 Confirmation of Pathogenic Mutations

De novo and inherited variants in monoallelic tier 1 and tier 2 CHD genes were identified by A Sifrim (see section 3.3.2 for description of tier 1 and tier 2). Several patients had multiple mutations in a single gene. 4 identical mutations in *RAI1* were identified in 4 separate individuals. This was an insertion, which was thought to be a false call due to difficulty in aligning the reads. These *RAI1* insertions were excluded from further analysis.

Of the tier 1 and tier 2 de novo variants, seven were VUS and were not considered for feedback. 5 individuals were identified with de novo mutations that had previously been reported as pathogenic in known monoallelic CHD genes. Two individuals with *PTPN11* mutations had been diagnosed in the NHS since recruitment, so the results did not require reporting. This left three pathogenic mutations for confirmation and feedback to the participants (Table 7.1).

Gene	Mutation	Genotype based interpretation
GATA6	De novo c.1367G>A p.R456H	*GATA6* Causes AD CHD with or without pancreatic aplasia. This mutations has been reported as pathogenic [932]. All mutations occur in same DNA binding site.
PTPN11	De novo, c.992A>G p.Asn308Asp	Causes AD Noonan Syndrome. Reported in the literature [329]and HGMD as pathogenic
PTPN11	De novo c.188A>G p.Tyr63Cys	Reported as pathogenic [329, 933, 934] (rs121918459)

Table 7.1 Review of de novo variants for feedback.

All inherited mutations in tier 1 and tier 2 genes were VUS, and were not suitable for reporting. Two individuals were identified with 22q11.2 deletions. This diagnosis had also been made in the NHS since recruitment and therefore did not require reporting.

7.3.1.1 Confirmation of *GATA6* mutation

Sequencing identified the de novo missense mutation c.1367G>A, p.R456H in *GATA6* based on transcript: ENST00000269216 (rs387906819). Both SIFT and Polyphen predicted it to be pathogenic [493].

Mutations in *GATA6* are associated with pancreatic agenesis and congenital heart defects, as well as isolated ASD, AVSD, persistent TA and TOF [302, 303]. The identified p.R456H mutation, has been reported in a single patient with pancreatic aplasia, PDA, VSD and a hypoplastic left pulmonary [932]. The mutation was identified by targeted sequencing. A further mutation at this

codon, p.Arg456Cys, was also found in association with pancreatic agenesis and CHD with a congenital diaphragmatic hernia [932].

Confirmation of the mutation is shown below in figure 7.2.
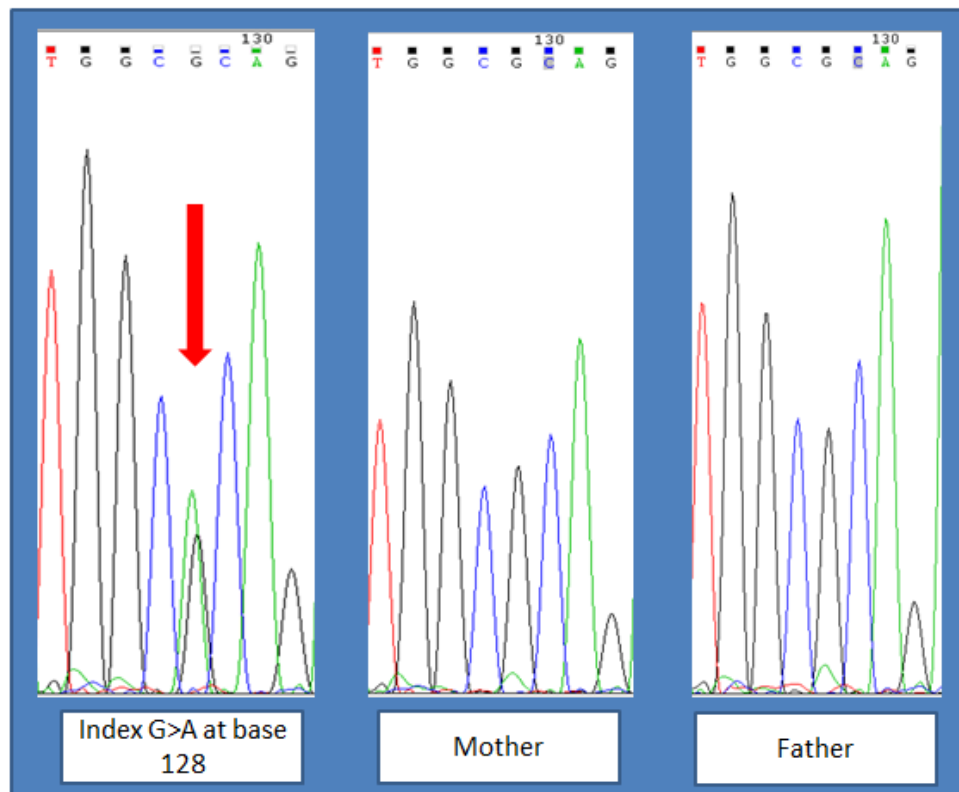
Figure 7.2 PCR was carried out to amplify the region of interest in *GATA6*.

A 244bp product was confirmed by running the PCR products on a 1% agarose gel initially. The PCR was then repeated and sent for Sanger Sequencing. The sequencing results for the trio show the G>A mutation in the index (red arrow), but neither parent.

7.3.1.2 Confirmation of *PTPN11* mutation p.Asn308Asp

WES identified the de novo mutation p.Asn308Asp in *PTPN11,* which has been reported as pathogenic previously [329]. Mutations in *PTPN11* are associated with Noonan Syndrome and related disorders. The individual had pulmonary stenosis, the common type of CHD seen in this condition. The result was confirmed (Figure 7.3).
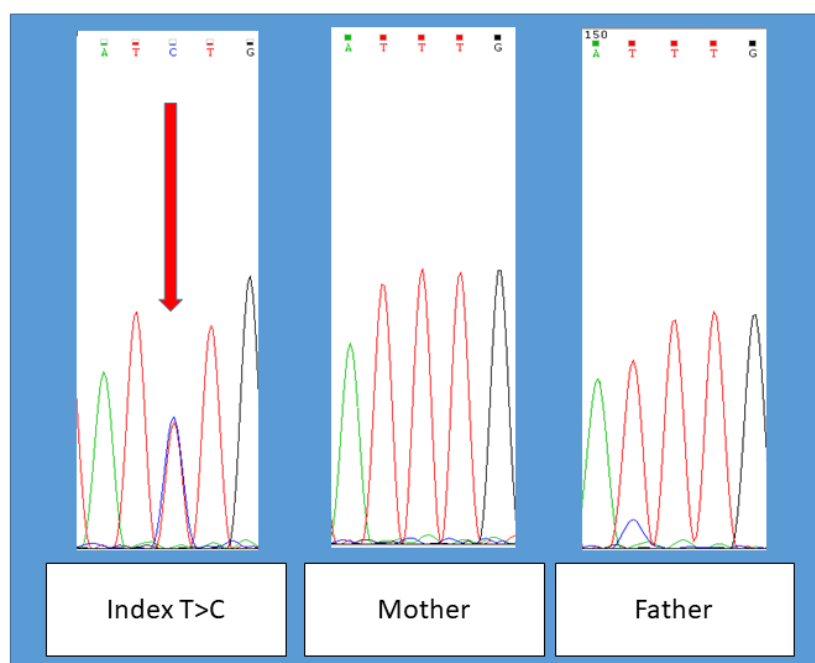
Figure 7.3 Sequence results for the *PTPN11* mutation.

Sequence amplifying the region surrounding mutation c.992A>G, p.Asn308Asp identified in the index (red arrow). The reverse sequence is shown as it was clearer.

7.3.1.3 Confirmation of *PTPN11* mutation p.Tyr63Cys

The individual with the p.Tyr63Cys mutation in *PTPN11* had pulmonary stenosis. The result was confirmed by PCR and subsequent Sanger sequencing (Figure 7.4).
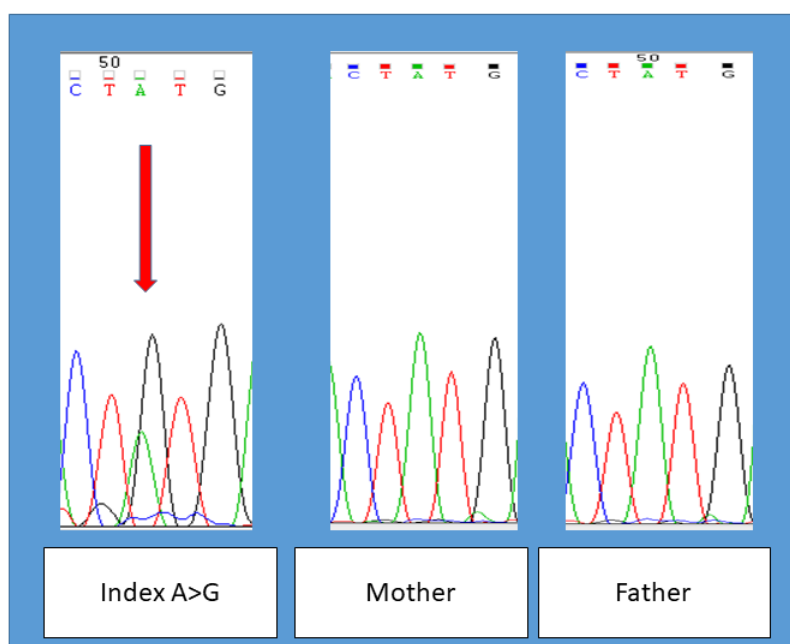
Figure 7.4 *PTPN11* sequence amplifying the region of interest.

Forward Sequencing results to confirm mutation p.Tyr63Cys. The A to G substitution can be seen at position 51 in the index (red arrow), but not in either parent.

7.3.2 Consideration of Secondary Findings

Guidance produced by the MRC suggests that investigators consider the chances, challenges and consequences of identifying secondary findings. Table 7.2 outlines the implications for feeding back secondary findings to individuals involved in the WES study.

Consideration	Implications for our study
Chance of significant unexpected result	3% in WES [935], 1% for Array-CGH [936]
Severity	Variable, could be life threatening in case of certain tumour suppressor genes e.g. *APC*, *BRCA*, *TP53*.
Chance of false positive	Mutations would need to be verified in the lab by a second method, before feeding back. A referral to NHS Clinical Genetics Services and confirmation in an NHS approved laboratory would be required in addition.
Clinical Utility	Will vary depending on the gene involved.
Screening and Management options	Depends on gene, certain such as *APC* and *BRCA* have definite management plans, others such as *TP53* are less clear. This screening can vary depending on NHS health trust.
Vulnerable populations	Children: Many participants would have been recruited as children, therefore it would not be appropriate to feedback information about adult onset conditions. Those recruited as children who are now adults would need to be re-consented. Individuals with NDD: Some individuals were recruited with learning difficulties.
Benefits of feedback	For the individual: a mutation may explain other pathologies they have, or might identify other organs systems that may need surveillance. The option of prenatal testing may be available in some cases, and lifestyle modifications or treatment options may be available for some genetic conditions. For the family: screening within the family to identify others at risk or reassure them.
Harms of feedback	Worry and anxiety, especially about later onset conditions or those where there is no cure or limited treatments. It is very difficult to provide counselling

	for genes which have variable penetrance and age of onset. Implications for insurance would require discussion. Communication of results in the family may cause upset [937].
Logistics and Costs of Feedback	Review of all consent forms to identify those that opted to receive results, and a further contact to discuss whether they would like to receive secondary findings. There was no clinician time allocated for additional discussions of secondary findings. Any results would require a referral to the local Clinical Genetics team and confirmation of the result in a NHS approved laboratory. This would require an additional blood sample and use of phlebotomy services.
How the results will be verified	In the laboratory initially, PCR and sequencing. Mutations would then need confirmation in an NHS accredited laboratory as well.
How and when feedback will occur	Ideally via a clinical geneticist, however there is no route other than cardiology in place currently. There is currently no guidance on what timeframe these results should be identified and verified within. We would need to consider the impact on local NHS services if all results were passed back to study participants at once.
Who will deliver the feedback and are they experienced enough and appropriately trained to do this	Consultant cardiologist but clinical genetics would be essential. A new results pathway would need to be planned and financed. Some participants may no longer be in the area and would require feedback from a service more local to them or telephone feedback.

Table 7.2 Completed table from the MRC guidance. Consideration of the implications of feeding back secondary findings to individuals in this WES study. (The MRC Framework on the Feedback of Health-related findings in Research https://wellcome.ac.uk/sites/default/files/wtp056059.pdf).

This assessment highlights a number of issues. The main difficulties would have been the logistical aspect of actually gaining consent retrospectively, and lack of involvement of a clinical geneticist. The majority of participants also fell into vulnerable categories, given most were very young children when recruited and may require careful counselling adapted to their needs.

Professor Brook and I met with the local Clinical Genetics consultants in 2015 to discuss feedback of secondary findings further, given that participants would have potentially been referred to their service. The consent forms and information sheets were available for review. It was decided that it would be appropriate to feedback pathogenic mutations in definite CHD causing genes, but not secondary findings or VUS. This is because these findings did not form part of the original consent process. The option of re-contacting patients to discuss secondary findings was considered, but it was not thought practical given the number of participants and the amount of time that has passed since they were recruited to the study.

We also considered genes that cause CHD and other cardiac phenotypes. A number of genes can cause both aortic dilatation and CHD. It was decided that it would be appropriate to feed these back. The same applies to genes which cause structural CHD and cardiomyopathy later in life. We then discussed genes that cause a purely aortic dilatation or cardiomyopathy phenotype. Although these processes relate to the heart, we decided that it was a separate issue to congenital structural CHD, and would instead fall into the category of secondary findings.

Given that the WES data had only been assessed by the WTSI pipeline, the calls in genes suggested for reporting of secondary findings had not been analysed in sufficient detail to identify pathogenic mutations. It was decided that it would not be appropriate to proceed with these any further.

Moving forward however, it is necessary to plan for secondary findings. What is essential is that consent is informed and should include discussions around what feedback is available. It was decided that in the future, new study participants should be offered a choice as to whether they would like to receive information on secondary findings or not if bioinformatics resources were available to identify them, or they would be advised that we would not feedback secondary findings. I updated the consent forms accordingly, but this will require a substantial ethics amendment to effect the change. Feedback of secondary findings will require careful planning and allocation of resources to set up a MDT to discuss which results to feedback, to allow time and laboratory resources to confirm the results, and an appropriate clinician to relay the results back to participants.

7.4 Conclusions

We identified pathogenic mutations in three individuals with no prior genetic diagnosis. A MDT to discuss potentially pathogenic results and the possibility of functional work for VUS, could also increase this number further. Larger studies would increase the chance of identifying more novel CHD genes, and would hopefully identify more results for participants.

When considering further patient recruitment, we should think about what results might be generated as part of the research. Patient involvement in research is being encouraged more and more, and given they provide the data in many studies it seems only fair to involve them in decision making, and ask which results they feel are important [938]. Close collaboration between researchers and clinicians will be required to ensure that important variants are identified and that there is an appropriate pathway to return results.

Some of the most successful studies seamlessly integrate clinical care with research. Most recently the 100 000 Genomes project recruited individuals from the NHS and may provide them with a diagnosis, but at the same time the data is held securely, strictly anonymised and available to researchers and industry. This represents the ideal situation where the individual acts in an altruistic fashion to provide information for research and potentially answers for others, but may also benefit directly by receiving a diagnosis for themselves too. Perhaps as we move towards WES and WGS for patients, this will become the normal and all NHS patients will also be research participants. As the boundaries between research and clinical care grow ever closer it is likely that it will become part of researcher's duty of care to inform individuals of secondary findings [939].

Chapter 8 Final Conclusions

CHD is the most common congenital abnormality, and presents significant costs to both affected families and health services. Improved survival means that this cohort require specialist lifelong care. Understanding the genetic basis of CHD is beneficial at the individual and population level. It allows us to provide tailored care, prognostic information and prenatal options. It may also help us identify therapeutic targets for a wide range of cardiovascular conditions, making it relevant to an even wider population.

The majority of individuals with CHD, especially those with NS-CHD, do not receive a genetic diagnosis. Previous research has identified a number of genes that cause CHD by analysis of families displaying Mendelian inheritance. These families are rare however, and CHD is usually a sporadic occurrence. It is likely therefore, that these results are not representative of the wider CHD population. The WES work carried out as part of this thesis has considered a more appropriate CHD cohort, and the results have confirmed the role of de novo variants in both S-CHD, and NS-CHD to a lesser extent. It has also strengthened the association with genes involved in chromatin modification.

Novel findings in this thesis include the identification of three new genes which cause S-CHD: *CHD4*, *PRKD1* and *CDK13*. We have also identified for the first time, that inherited rare PTVs with reduced penetrance, contribute to NS-CHD. We have also been able to expand the phenotypes of known CHD genes (*TAB2*) and a number of genes known to cause developmental disability, but not CHD previously.

14 individuals have now been reported with mutations in *CHD4*, in addition to the 5 we described, and a consistent phenotype including NDD, hearing loss, skeletal, palatal and genitourinary abnormalities and recognisable facial features is now apparent. It is now included in NHS genetic testing, on a number of

panels. Evidence from mouse models, and the fact that over half of the individuals reported with mutations in this gene have CHD, confirms that it is an important contributor to CHD.

The case of *PRKD1* is less clear. The small numbers of individuals with mutations in this gene reported so far, mean that it is difficult to determine the associated phenotype and suggest that it is not a common cause of CHD. Studies of embryonic survival in the *Prkd1*[Em1(IMPC)Wtsi] and *Prkd1*[Em2(IMPC)Wtsi] mice, show homozygous lethality with no definite evidence of worsened survival in heterozygotes. The mouse models did not conclusively support the role of *Prkd1* in CHD, but no *Prkd1*[Em1/Em1] mice, and only a single *Prkd1*[Em2/Em2] mouse (normal cardiac phenotype) were available for analysis by HREM. The implications of finding an IAA and a muscular VSD in heterozygous mice from the two lines are unclear. Despite supporting evidence of CHD in other mouse models with knockout of genes known to interact with *Prkd1* (*Tbx5, Mef2c, Hdac5/9*), a larger scale analysis including earlier gestations or other functional work are required to understand the role of *PRKD1* in CHD.

PRKD1 presents an interesting target given its role in hypertrophy. It could be an important gene to consider in other cardiac diseases that result in a hypertrophic response. Given the *Prkd1*[Em2/+] mice had no evidence of structural cardiac disease on MRI, the finding of significant ventricular wall thickening in the RV segment and a trend towards an increase wall thickness in the majority of ventricular wall segments was surprising. Further work could investigate the consequence of the mutation on function of the protein and consider the response of the heart in the same line under stress. Careful monitoring of the individuals reported with mutations in *PRKD1* and CHD will help determine if hypertrophy is a feature of the p.Gly592Arg mutation.

After the initial report of seven individuals with mutations in *CDK13*, a further 37 individuals have been reported, allowing delineation of the phenotype and recommendations for clinical care to be produced. Like *CHD4* and *PRKD1*, it is

now included in NHS genetic testing. There is a recognisable facial gestalt and key features include developmental delay, structural brain abnormalities and CHD. *Cdk13*^{tm1b/tm1b} mice exhibit complete embryonic lethality, but over a wide range of embryonic stages. HREM studies support a role in CHD, with a preponderance for endocardial cushion defects. Functional work may help ascertain the mechanism by which mutations result in CHD.

The work reported here shows that much larger sample sizes will be required to identify the remaining novel CHD genes. The work carried out for this thesis to phenotype the cohort and create a list of genes we are highly confident cause CHD, has provided a starting point for a larger WES study that is in progress. Confirming pathogenic variants and feeding back results to participants is possible, but additional resources need to be planned to deal with secondary findings and evaluation of other potentially pathogenic variants going forward.

Appendix A. Overview of cardiovascular development in the mouse

The gestational period in the mouse is 19-21 days. Despite the shorter timeframe, formation of the mouse heart follows a similar course to that described in the human in 1.2.1. The mouse heart is the first organ to form, to meet the oxygen requirements of the developing embryo. Oxygenated blood flows into the heart through the umbilical veins from the placenta, and deoxygenated blood arrives from the systemic circulation via the cardinal veins and sinus venosus, and via the vitelline veins from the yolk sac [889]. An overview of development is shown in figure A.1 and table A.1 below. Mouse embryogenesis is staged by embryonic day (E), and by Theiler stage defined by morphological development.
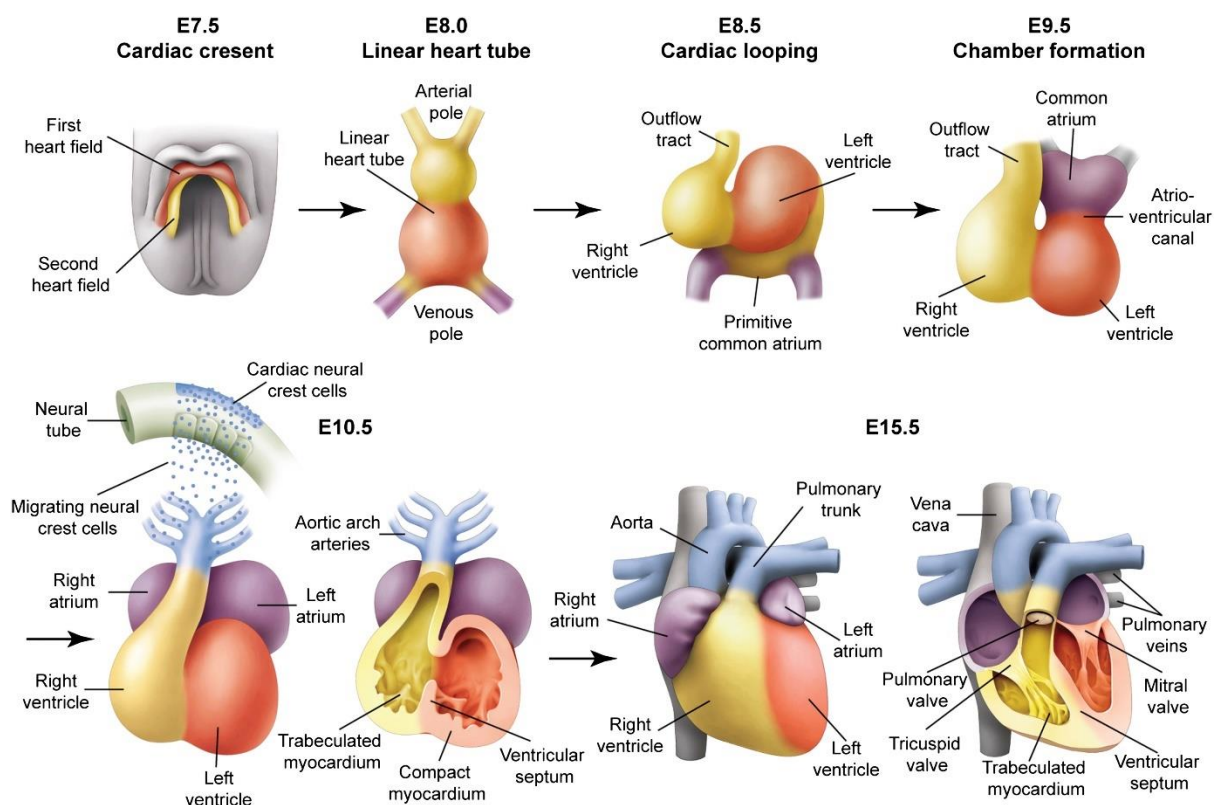
Figure A.1 Overview of mouse heart formation including looping.

Reproduced with permission from Science Direct and Epstein et al [940]. Overview of mouse heart formation including looping. E; Embryonic day.

Stage	Development
E7.5	Cardiac crescents form, made up of cells from the first and second heart fields [889].
E7.5-9.0	Contractions can be seen from E7.5, and a regular sinus rhythm originating from the venous pole is present from E9.0 [215, 217, 941]
E8.25-E10.5	Heart tube looping [942]
E8.5 – E13	Pharyngeal/aortic arch arteries. There are six bilaterally, and they join to the two dorsal aortae [889]. The first and second are seen from E8.5, and become the mandibular, stapedial and hyoid arteries [943]. Others are seen at a slightly later stage, from E9.5 onwards. They eventually become the aortic arch and proximal pulmonary arteries and resemble the postnatal arrangement by E13 [943].
E10.5-E13.5	Atrial septation [218]. The septum secundum forms from tissue at the posterior of the atrium and is present from E11.5. Fusion with the superior and inferior cushions occurs at E12, to close the ostium primum.
E10.5-E13.5	Ventricular septation occurs between E10.5-13.5 [218]. The atrioventricular and outflow tract cushions (ridges) which become evident at E9.5, contribute to the valves, ventricular septum and division of the outflow tract [944, 945]. Evidence of a muscular ventricular septum is seen from E10.5, but this is variable, so E11.5 is often considered to be the point at which septation truly starts. Closure of the interventricular foramen is seen at E13.5.
E10.5-E17.5	Atrioventricular valve formation begins at E10.5 and progresses until E17.5. The four endocardial cushions in the atrioventricular canal come together at E11.5 to produce the atrioventricular septum, and by E12.5 the two atrioventricular valves are present [946]. The valves and chordae continue to develop until

	E17.5, but the tricuspid valve does not delaminate and may not separate fully from the septum [218].
E11.0	At E11.0 ventricular trabeculations are present [217].
E11.5- E13.5	Outflow septation. The septum forms from a number of sources. The neural crest cells begin septation and the two cushions (mesenchymal ridges), which are seen from E9.5, spiral and separate the outflow tract proximally into the two arteries by E12.5.
E11.5 – E15.5	Venous drainage. At E11.5 the IVC, ductus arteriosus and coronary sinus (previously left sinus horn) are established. The left SVC is developing [889]. The IVC enters the prospective right atrium at the sinus venosus [217]. At E12, the left and right SVC are present, but the right SVC empties into the IVC at this point, rather than the RA [946]. The left SVC and IVC are connected by the coronary sinus which does not drain directly into the RA until E15.5, and the IVC will maintain its own entry point into the RA caudally [946, 947].
E11.5- E18.5	The semilunar valves are mainly derived from the endocardial cushions and are beginning to form at E11.5. The third valve cusps are produced from the walls of the truncus. They thin and mature between E12.5 and E18.5.

Table A.1 Overview of some of the main events in mouse heart development.

Appendix 2. *Cdk13*^{tm1b(EUCOMM)Hmgu} Mouse Sequence

The relevant sequences of the *Cdk13*^{tm1b(EUCOMM)Hmgu} mouse, including the cassette, are outlined below in figure 2. The wild type sequence just prior to exon 3 with the two relevant primers for identification of the wild type mouse are shown in figure 3.

```
 9061 ttagcactct acaaagttta atttggttac tgactaaatg ttaaattatt tttcatcagacgatctgcca aatcccgaag cagaagccca tattcatcta ggcactcaag atctcggagc
 9181 aggcacagat tgtctagatc aagaagtcgt cattcaagca tttctcctag cacactaactctgaagagta gcttggcagc tgaattgaac aagaataaaa aggcacgagc tgcagaggca
 9301 gcaagagctg cagaggcagc aaaagctgca gaagctgcta agcagctgctga agcagctgccaaagctgcca aagcctcaaa tgcttctaca cctaccaagg ggaacacaga aactggtgcc
 9421 agtgtctcac agacaaacca tgtgaaggaa gtcaaaaaac ttaaaactga gcatgcaccttctccttcaa gtggtgggac cgtcaaaagc gacaaagcaa aaacaaagcc accgcttcaa
 9541 gtaacaaagg tagacaataa tttgacagta gagaaagcca ccaagaaaac agttgttgggaaggagagta aacctgctgc tacaaaggaa gaaccagttt ccactaaaga gaaaagcaag
 9661 ccactcacac caagcacagg agccaaggag aaggagcagc atgtggcttt agtgacctctacgttaccgc cattaccttt gcctcccatg ctgcctgaag ataaagatgc tgataggtaa
 9781 gtccagacaa atttgaaatt agtaactctt caaagaaaaa gtctgcttat tttgctgtt .............................
14761 tgattctttc tttttaattc ataattttga aagttagagc tattttttt taagtttaaa ctttgttatt tgtttgtgtg gcatgcaata ctgctggtat aggctctaag
14881 ggcaaccttg aatgttgctt ctgaccttgt tttaggcaac ttcactttgt tactccaata taaacaccag ggtagctggc caatgagctt tcagggattc tcctgcctcc tcccctcccc
15001 tcagatgagg gttactgaga ttgtagactg tgctgctgtg cccagcttta aaggcgcata acgataccac gatatcaaca agtttgtaca aaaaagcagg ctggcgccgg aaccgaagtt
15121 cctattccga agttcctatt ctctagaaag tataggaact tcgaacccttt tcccacacca ccctccacac ttgccccaaa cactgccaac tatgtaggag gaaggggtg ggactaacag
15241 aagaacccgt tgtgggggaag ctgttgggag ggtcactta tgttcttgcc caaggtcagt tgggtggcct gcttctgatg aggtggtccc aaggtctggg gtagaaggtg agagggacag
15361 gccaccaagg tcagccccccc ccccctatcc catagggagcc aggtccctct cctggaggccg acgggtgtt actcctcac atttaatgtt gatgaaagct gatgctggg aggccgagcg
15481 caaggaaaca aacttggcct caccaggcct cagccttggc tcctcctggg aactctactg cccttgggat cccctttgtag ttgtgggtta cataggaagg gggacgggat tcccccttgac
15601 tggctagcct actcttttct tcagtcttct ccatctcctc tcacctgtct ctcgacccttt ccctaggat agacttggaa aaagataagg ggagaaaaca aatgcaaacg aggccagaaat
15721 gattttggct gggcattcct tccgctagct tttattggga tccccctagtt tgtgataggc cttttagcta catctgccaa tccatctcat tttcacacac acacacacca ctttccttct
15841 ggtcagtggg cacatgtcca gcctcaagtt tatatcacca cccccaatgc ccaacacttg tatggccttg ggcggggtcat cccccccccc accccccagta tctgcaacct caagctagct
15961 tgggtgcgtt ggttgtggat aagtagctag actccagcaa ccagtaacct ctgcccttttc tcctccatga caaccaggtc ccaggtcccg aaaaccaaag aagaagaacc ctaacaaaga
16081 ggacaagcgg cctcgcacag ccttcactgc tgagcagctc cagaggctca aggctgagtt ....................
16741 aaaaacgtct aggcccccg aaccacgggg acgtggtttt cctttgaaaa acacgatgat aagcttgcca caaccatgga agatcccgtc gtttacaac gtcgtgactg ggaaaaccct
16861 ggcgttaccc aacttaatcg ccttgcagca catccccctt tcgccagctg gcgtaatagc gaagaggccc gcaccgatcg cccttcccaa cagttgcgca gcctgaatgg cgaatggcgc
16981 tttgcctggt ttccggcacc agaagcggtg ccggaaagct ggctggagtg cgatcttcct gaggccgata ctgtcgtcgt ccctccaaac tggcagatgc acggttacga tgcgcccatc
17101 tacaccaacg tgacctatcc cattacggtc aatccgccgt ttgttcccac ggagaatccg acggtgtgtt actccgctc atttaatgtt gatgaaagct gatgctggg gctacagga aggccgacg
17221 cgaattattt ttgatggcgt taactcggcg tttcatctgt ggtgcaacga gcgctgggtc ggttacggcc aggacagtcg tttgccgtct gaatttgacc tgagcgcatt tttacgcgcc
17341 ggagaaaacc gcctcgcggt gatggtgctg cgctggaagtg acggcagtta tctggaagat caggatatgt ggcggatgag cggcatttc cgtgacgtct cgttgctgca taaaccgact
17461 acacaaatca gcgatttcca tgttgccact cgctttaatg atgatttcag ccgcgctgta ctggaggctg aagttcagat gtgcggcgag ttgcgtgact acctacgggt aacagtttct
17581 ttatggcagg gtgaaacgca ggtcgccagc ggcaccgcgc ctttcggcgg tgaaattatc gatgagcgtg gtggttatgc cgatcgcgtc acactacgtc tgaacgtcga aaacccgaaa
17701 ctgtggagcg ccgaaatccc gaatctctat cgtgcggtgg ttgaactgca caccgccgac gttgaagcag agcctgcgat gtcggtttcc gcgaggtgcg gattgaaaat
17821 ggtctgctgc tgctgaacgg caagccgttg ctgattcgag gcgttaaccg tcacgagcat catcctctgc atggtcaggt catggatgag cagacgatgg tgcaggatat cctgctgatg
 7941 aagcagaaca actttaacgc cgtgcgctgt tcgcattatc cgaaccatcc gctgtggtac acgctgtgcg accgctacgg cctgtatgtg gtggatgaag ccaatattga aacccacggc
18061 atggtgccaa tgaatcgtct gaccgatgat ccgcgctggc taccggcgat gagcgaacgc gtaacgcgaa tggtgcagcg cgatcgtaat cacccgagtg tgatcatctg gtcgctgggg
18181 aatgaatcag gccacggcgc taatcacgac gcgctgtatc gctggatcaa atctgtcgat ccttcccgcc cggtgcagta tgaaggcggc ggagccgaca ccacggccac cgatattatt
18301 tgcccgatgt acgcgcgcgt ggatgaagac cagcccttcc cggctgtgcc gaaatggtcc atcaaaaaat ggctttcgct acctggagag acgcgcccgc tgatcctttg cgaatacgcc
18421 cacgcgatgg gtaacagtct tggcggtttc gctaaatact ggcaggcgtt tcgtcagtat ccccgtttac agggcggctt cgtctgggac tgggtggatc agtcgctgat taaatatgat
18541 gaaaacggca acccgtggtc ggcttacgcg ggtgattttg gcgatacgcc gaacgatcgc cagttctgta tgaacggtct ggtctttgcc gaccgcacgc cgcatccagc gctgacggaa
18661 gcaaaacacc agcagcagtt tttccagttc cgtttatccg ggcaaaccat cgaagtgacc agcgaatacc tgttccgtca tagcgataac gagctcctgc actggatggt ggcgctggat
18781 ggtaagccgc tggcaagcag taagtgcct ctggatgtcg ctccacaagg taaacagttg attgaactgc ctgaactacc gcagccgagg agcgccgggc aactctggct cacagtacgc
18901 gtagtgcaac cgaacgcgac cgcatggtca gaagccgggc acatcagcgc ctggcagcag tggcgtctgg cggaaaacct cagtgtgacg ctccccgccg cgtcccacgc catcccgcat
19021 ctgaccacca gcgaaatgga tttttgcatc gagctgggta ataagcgttg gcaatttaac cgccagtcag gctttctttc acagatgtgg attggcgata aaaaacaact gctgacgccg
19141 ctgcgcgatc agttcacccg tgcaccgctg gataacgaca ttggcgtaag tgaagcgacc cgcattgacc ctaacgcctg ggtcgaacgc tggaaggcgg cgggccatta ccaggccgaa
19261 gcagcgttgt tgcagtgcac ggcagataca cttgctgatg cggtgctgat tacgaccgct cacgcgtggc agcatcaggg gaaaacctta tttatcagcc ggaaaaccta ccggattgat
19381 ggtagtggtc aaatggcgat taccgttgat gttgaagtgg cgagcgatac accgcatccg gatgtgagtt gttggtcgct gtggaaggac tgggatctgc atggacctcg gtgaaggtgg
19501 ttagggccgc aagaaaacta tcccgaccgc cttactgccg cctgttttga ccgctgggat ctgccattgt cagacatgta taccccgtac gtcttcccga gcgaaaacgg tctgcgctgc
19621 gggacgcgcg aattgaatta tggcccacac cagtggcgcg gcgacttcca gttcaacatc agccgctaca gtcaacagca actgatggaa accagccatc gccatctgct gcacgcggaa
19741 gaaggcacat ggctgaatat cgacggtttc catatgggaa ttggtggcga cgactcctgg agcccgtcag tatcggcgga attcagctg agcgccggtc gctaccatta ccagttggtc
19861 tggtgtcaaa aataataata accgggcagg ggggatctaa gctctagata agtaatgatc ataatcagcc atatcacatc tgtagaggtt ttacttgctt taaaaaaccct cccacacctc
19981 cccctgaacc tgaaacataa aatgaatgca attgttgttg ttaacttgtt tattgcagct tataatggtt acaaataaag caatagcatc acaaatttca caaataaagc atttttttca
20101 ctgcattcta gttgtggttt gtccaaactc atcaatgtat cttatcatgt ctggatccgg aataacttcg tatagcatac attatacgaa gttatggtct gagctcgcca tcagttcaga
20221 tttaggggaa tgtgttcctt ttgtggtgaa atgattatta aaactgttgc tacttaggct......................
21541 tttaaaaggt aggaacagaa atgataaaaa tgttaaaatt atgaaaaatg gcattgtctt ttgatagcta gtaacttgat tttatttaac tgtctaatag gagaaatggt agccttaaag
21661 aaagtacgtc tggataatga aaaggagggt ttcccaatta cagcaattag agaaattaaa attcttcggc aactcaccca ccagagtatc atcaatatga aggaaatcgt gactgataaa
21781 gaagatgctt tggatttaa gaaagacaaa ggtatgtatg cagacttcct aaccaagtct ttgaattcat gttatttcca actggataat tatttaagga ctgggtagat aaaatgtttt
21901 ctaatgtcaa cctgttttaa aatttagtag agttaaagaa gtacatatt acgttagtgg
```

Figure B.1 Sequence of *Cdk13* with relevant sites for generating and genotyping the *Cdk13*^{tm1b(EUCOMM)Hmgu} mouse.

Frt sites are highlighted in green. The lox P site is highlighted in red. The Exons are pink and the introns are turquoise. The cassette is in yellow. The exon at position 9121 is exon 2, and at position 21601 is exon 5. The mouse model therefore lacks exons 3 and 4. Primers for genotyping are highlighted in grey: CDK13_8441_F primer is at position 14821, Cas-R1Term is at position 15001, LacZ small F at 18181 and LacZ_small_R is at 18290. At position 15051, the sequence deviates from the WT.

```
tgggcaagagaatcgtgtgctaattgtagtttgtgatttatttgctaaaaagaaggcat
cattgtctccttgtcataatcaaatactacaaaattatgaggtttcttgcaaagcatatt
gatttgattctttcttttaattcataattttgaaagttagagcttatttttttaagtt
taaactttgttatttgtttgtgtgtattgtatgtgcatgcaatactgctggtatag gctc
taagggcaaccttgaatgttgcttctgaccttgtttaggcaacttcactttgttactcc
aatataaacaccagggtagctggccaatgagctttcagggattctcctgcctcctcccct
cccctcagatgagggttactgagattgtagactgtgctgctgtgcccagctttatgtgga
tttttgagatttgaatccaggttcttacacttgaatgactagcacc acaacaccatcttc
ccagct tcctagactttattttaaaaaatagtttcagatttgcaaagataaaaagatat
tgaatccacatttgtcttgcacttattacatttagtcagtattcatatactactattaat
taaggtctactctatagatttctatggttttacttaattttgtgtcaccttttttctttc
tttcttttttttttttacttcgtggttgaaggataataatctgtttagatggacttgaag
tatgacatatgcctttgcatgctgcttaaattttaaatatgaccatagtttatcgtgta
atcaattacttatggttgatatgacttacacccctcccccttttttttctttag

CTTAAGAGGCAACATTTCTGTCAAAGCAGTTAAAAAAGAAGTAGAAAAGAAACTCCGATG
TCTGCTTGCTGATTTACCATTGCCCCCTGAGTTACCAGGAGGAGATGATCTTTCCAAGAG
TCCAGAGGAGAAGAAAACAGCAGCACAGTTACATAGCAAACGAAGGCCTAA

gtatgtgcttgctttctactggcacttaggttgactggacacactatatcctggataact
aaaaaactagaaaacagtaagatttaaatgaaaagtgtatatttatttgaacataaatat
acaggttggtttagagttctgaaacaaatattcatgaacggtggtacctatttgcagcta
```

Figure B.2 Wildtype *Cdk13* sequence.

CDK13_84881_F and CDK13_84881_R primers in the wild type mouse sequence are highlighted in grey, just prior to exon 3 (capitalised sequence).

Appendix C. Phenotype of the Wildtype C57BL/6N Mouse

The background strain of both the *Prkd1* and *Cdk13* mice is the commonly used C57BL/6N inbred strain. It is a permissive background, with low incidence of tumours, they breed well and have a relatively long lifespan (https://www.jax.org/strain/000664). This strain is used in the DMDD project and is well phenotyped (https:/dmdd.org.uk). Some of the common abnormalities that can arise in this strain are outlined in table 2 below. Occurrence of any of these features should not be attributed to mutant gene expression without convincing evidence.

Abnormality	Details
Susceptible to diet induced obesity, type II diabetes and atherosclerosis	A high fat diet tends to result in obesity, mild-moderate hyperglycaemia and insulinaemia, and atherosclerosis in the aorta.
Microphthalmia and other eye abnormalities	Affecting 8-20%. Unlikely to be relevant in this project, except for suitability for breeding [948].
Hereditary hydrocephalus	Thought to affect 1-4%. Unlikely to be relevant in this project, except for suitability for breeding.
Portosystemic shunts	Thought to affect 5%. Could be relevant in adult studies.
Hair loss associated with over grooming	Will be monitored and action taken if welfare thought to be compromised.
Severe progressive hearing loss	Late onset. Unlikely to be relevant to this project

Table C.1 Characteristics of the C57BL/6 mice.

Information from Jackson Laboratory site https://www.jax.org/strain/000664.

To accurately phenotype the mouse heart, it was important to understand the expected anatomical development of the heart at E15.5. E14.5 and E15.5 embryos tend to have reached Theiler Stages 21-23 [224]. There are rapid changes in the development of the heart between these stages, and staging based on days post coitus and even the Theiler system may not be adequate. Geyer et al. produced a more detailed sub stage system to overcome this problem and examined the heart at each stage with HREM [518]. The table below (table 2) is a summary of the phenotypic information collected. The authors noted that the variable anatomy between the stages of development mean that there is a risk of incorrect interpretation of normal development as pathology. For example, the interventricular foramen may still be open as part of normal development, and be incorrectly described as a perimembranous VSD. Between these stages, there was huge variation in the diameter of vessels and their lumens, meaning that it is hard to comment on vascular stenosis. Another specific limitation of this method is the appearance of confined areas of capillary dilatation that might be misinterpreted as haemangioma, haemorrhage or telangiectasia. The developmental range described was used to inform the HREM analysis of E15.5 hearts.

Anatomy	S21	S22-	S22	S22+	S23-	S23	Abnormal phenotypes in Wildtype Embryos
Atrial Communication	Septum fully formed, only communication is foramen ovale. Large appendices with pectinate muscles	As for S21.	As for S21.	As for S21.	As for S21.	As for S21.	Fewer pectinate muscles, common atrium (with AV malformation)
Ventricular Communication	97% incomplete septation of ventricles	-	-	-	-	S23 23 2% have remnant of the interventricular communication muscular).	VSDs (perimembranous and muscular).
Aortic Valve	Appearance similar to over-riding aorta as aortic valve sits over the interventricular communication. Fully formed, appears as 3 solid tissue protrusions.	-	Fully formed, appears as 3 solid tissue protrusions.	-	-	Aortic valve on left side of the septum in all embryos. Fully formed, appears as 3 solid tissue protrusions.	BAV Common valve with TA
Pulmonary Valve	To the right and sited more rostrally than the aortic valve. Fully formed, appears as 3 solid tissue protrusions.	-	As for S21.	-	-	As for S21.	DORV Common valve with TA
Myocardium thickness (µm)	R: 43 L: 63	R: 52 L: 68	R: 63 L: 91	R:74 L:106	R:87 L: 115	R:90 L: 119	
Coronary sinus	Present	As for S21.	As for S21.	As for S21.	As for S21.	As for S21.	
Coronary arteries: connection of both coronary arteries to ascending aorta (% not connected)	R: 39 L: 58	R: 7 L: 20	R: 0 L: 0	R: 0 L: 9	R: 0 L: 0	R: 0 L: 2	

Table C.2. Results from Geyer et al. Myocardial measurements from axial images at level of the interventricular foramen. Additional measurements were taken including the distance between the ductus arteriosus and subclavian artery. 1.5% of the embryos exhibited remnants of the sixth pharyngeal arch artery, and 50% had remnants of the right dorsal aorta.

Appendix D Embryonic Mouse Phenotyping Protocol

Gross Morphology of E14.5-E15.5 embryos

1. Measure CRL at time of schedule 1.
2. External examination. Use Mammalian Phenotype Ontology [949] codes to document phenotypes
3. Remove heart as per HREM protocol
4. Forelimb stage assessment

ID

CRL (mm)

Heart weight (mg)

(after fixation)

Tibia length (mm)

Heart weight/tibia length ratio (mg/mm)

Forelimb stage

External photos taken ☐

Heart photographed ☐

Sample for DNA ☐

External Examination

Alive	Yes/no
Severely dysmorphic	Normal/abnormal/unobservable
Skin haemorrhage	Normal/abnormal/unobservable
Head shape	Normal/abnormal/unobservable
Facial or palatal cleft	Normal/abnormal/unobservable
Body wall morphology	Normal/abnormal/unobservable
Ear morphology	Normal/abnormal/unobservable
Integument	Normal/abnormal/unobservable
Vibrissae morphology (whiskers)	Normal/abnormal/unobservable
Limb bud morphology	Normal/abnormal/unobservable
Polydactyly	Normal/abnormal/unobservable
Syndactyly	Normal/abnormal/unobservable
Spina bifida	Normal/abnormal/unobservable
Eye morphology	Normal/abnormal/unobservable
Tail morphology	Normal/abnormal/unobservable
Forelimb stage	

Reproduced from Geyer et al
A staging system for correct phenotype interpretation of mouse
embryos harvested on embryonic day 14 (E14.5)
J Anat. 2017 May; 230(5): 710–719
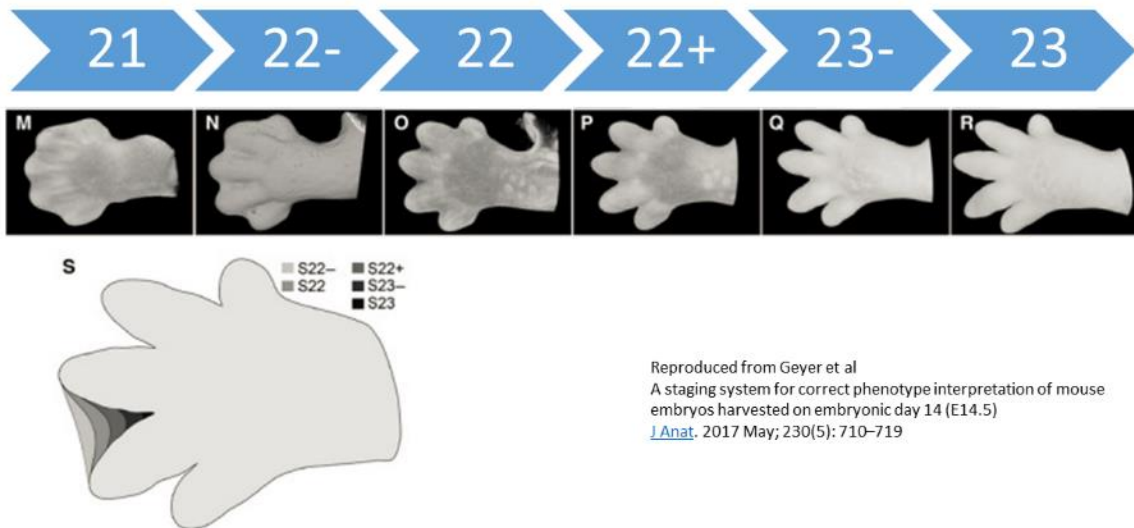
Internal Examination

Haemorrhage	Normal/abnormal/unobservable
Liver situs	Normal/abnormal/unobservable

Cardiac Examination

Outflow tract development	Normal/abnormal/unobservable
Heart morphology	Normal/abnormal/unobservable
Heart situs	Normal/abnormal/unobservable

Placental Examination

Placenta Development	Normal/Abnormal/Unobservable
Pale placenta	Normal/Abnormal/Unobservable

Forelimb staging figure reproduced with permission from Geyer et al John Wiley
& Sons Ltd on behalf of Anatomical Society.

Cardiac Assessment

Site	Features
Atria	Fully septated (except PFO) Appendages and pectinate muscles normal
Coronary sinus	Present or absent
PFO	Present or absent
Tricuspid Valve	Cavitations or full cusps seen Tricuspid
Mitral Valve	Cavitations or full cusps seen Bicuspid
Ventricular Septum	Interventricular foramen should not be present (seen in 2% at S23)
Aortic Valve	Cavitations or full cusps seen Tri/bicuspid Normal position on left, any overriding
Pulmonary Valve	Cavitations or full cusps seen Tricuspid Normal position on right and more cranial
Pulmonary artery	Normal position and branches
Aorta	Normal position and branches
Coronary arteries	Number seen and insertion normal
Ductus	present

Appendix E Adult Mouse Phenotyping Protocol

Overview

1. Schedule 1 the mouse
2. Transfer to warm PBS and measure the CRL
3. Remove heart
4. Internal and external examinations
5. Remove tail and freeze for DNA extraction (-20c)

Use MPO [949] codes to document phenotypes

ID

DOB

DOD

Sex

CRL (mm)

Heart weight (mg)

External photos taken ☐

Heart photographed ☐

Tissue for DNA taken ☐

External Examination

Eyes	Normal/Abnormal/Unobservable
Ears	Normal/Abnormal/Unobservable
Teeth	Normal/Abnormal/Unobservable
Cleft	Normal/Abnormal/Unobservable
Skin and hair	Normal/Abnormal/Unobservable
Whiskers	Normal/Abnormal/Unobservable
Limbs	Normal/Abnormal/Unobservable
Digits and nails (number, syndactyly)	Normal/Abnormal/Unobservable
Haemorrhage	Normal/Abnormal/Unobservable
Abdominal Wall	Normal/Abnormal/Unobservable
Spine	Normal/Abnormal/Unobservable
Tail	Normal/Abnormal/Unobservable
Oedema	Normal/Abnormal/Unobservable

Internal Examination

Liver situs	Normal/Abnormal/Unobservable
Stomach situs	Normal/Abnormal/Unobservable
Haemorrhage	Normal/Abnormal/Unobservable

Cardiac Examination

Situs	Normal/Abnormal/Unobservable
Development	Normal/Abnormal/Unobservable
Outflow tract	Normal/Abnormal/Unobservable

Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing

Alejandro Sifrim,[#,1] Marc-Phillip Hitz,[#,1,2,3] Anna Wilsdon,[4] Jeroen Breckpot,[5] Saeed H. Al Turki,[1,6,7] Bernard Thienpont,[8,9] Jeremy McRae,[1] Tomas W Fitzgerald,[1] Tarjinder Singh,[1] Ganesh Jawahar Swaminathan,[1] Elena Prigmore,[1] Diana Rajan,[1] Hashim Abdul-Khaliq,[10,11] Siddharth Banka,[12,13] Ulrike M. M. Bauer,[11] Jamie Bentham,[14] Felix Berger,[3,11,15] Shoumo Bhattacharya,[16] Frances Bu'Lock,[17] Natalie Canham,[18] Irina-Gabriela Colgiu,[1] Catherine Cosgrove,[16] Helen Cox,[19] Ingo Daehnert,[11,20] Allan Daly,[1] John Danesh,[1,21,22] Alan Fryer,[23] Marc Gewillig,[24] Emma Hobson,[25] Kirstin Hoff,[2,3] Tessa Homfray,[26] The INTERVAL Study,[27] Anne-Karin Kahlert,[2,3,28] Ami Ketley,[4] Hans-Heiner Kramer,[2,3,11] Katherine Lachlan,[29,30,31] Anne Katrin Lampe,[32] Jacoba J. Louw,[24] Ashok Kumar Manickara,[33] Dorin Manase,[33] Karen P. McCarthy,[34] Kay Metcalfe,[13] Carmel Moore,[22] Ruth Newbury-Ecob,[35] Seham Osman Omer,[36] Willem H. Ouwehand,[1,21,37,38] Soo-Mi Park,[39] Michael J. Parker,[40] Thomas Pickardt,[11] Martin O. Pollard,[1] Leema Robert,[41] David J. Roberts,[21,42,43] Jennifer Sambrook,[22,37] Kerry Setchfield,[4] Brigitte Stiller,[11,44] Chris Thornborough,[17] Okan Toka,[11,45] Hugh Watkins,[16] Denise Williams,[19] Michael Wright,[46] Seema Mital,[33] Piers E. F. Daubeney,[47,48] Bernard Keavney,[49] Judith Goodship,[50] The UK10K Consortium,[27] Riyadh Mahdi Abu-Sulaiman,[51,52,53] Sabine Klaassen,[3,11,54,55] Caroline F. Wright,[1] Helen V. Firth,[56] Jeffrey C. Barrett,[1] Koenraad Devriendt,[5] David R. FitzPatrick,[57] J. David Brook,[4] The Deciphering Developmental Disorders Study,[27] and Matthew Hurles[1]

[#]Contributed equally.

Corresponding Author: Matthew Hurles - ku.ca.regnas@hem

[27]A list of members and affiliations appears in the Supplementary Note.

Abstract

We evaluated the burden of high confidence DNMs within S-CHD and NS-CHD trios separately ($N_{\text{S-CHD}}= 518$, $N_{\text{NS-CHD}}= 847$). We classified DNMs into three distinct categories: PTVs (nonsense, frameshift and splice-site variants); missense variants (including in-frame indels); and silent mutations. We compared the observed numbers of DNMs to those expected under a null mutational model[9], across a set of manually curated CHD-associated genes, non-CHD developmental disorder associated genes and all remaining protein coding genes (Supplementary Tables 1-3, Figure 1A). S-CHD probands exhibited the largest excess in *de novo* PTVs (27 variants, OR=81, P=1.21×10^{-43}) and *de novo* missense variants (22 variants, OR=8.6, P=7.35×10^{-15}) for autosomal dominant CHD genes (Supplementary Table 4). S-CHD probands also manifested a burden of *de novo* PTVs in autosomal dominant developmental disorder-associated genes not currently associated with CHD (12 variants, OR=18.4, p=3.49×10^{-13}). In contrast, NS-CHD probands presented with a much lower burden of *de novo* PTVs in CHD-associated genes (4 variants, OR=7.3, P=2.61×10^{-4}). Finally, we found a significant exome-wide excess of *de novo* missense, but not silent mutations (after excluding CHD and developmental disorder genes) in both S-CHD and NS-CHD probands, suggesting additional undiscovered dominant CHD-associated genes. The excess of *de novo* PTVs in S-CHD cases reported here is of the same magnitude as that found in cases with severe developmental disorders without CHD and considerably higher than that found in Autism Spectrum Disorder (Figure 1B, Supplementary Table 5). The observed marked difference in DNM burden between NS-CHD and S-CHD confirms findings in a recent study

by Homsy *et al.*[8] looking at differences in mutational burden in CHD cases with and without neurodevelopmental deficits, which are by far the most common extra-cardiac manifestations. This burden additionally mirrors that observed in Autism between individuals with and without intellectual disability[10].
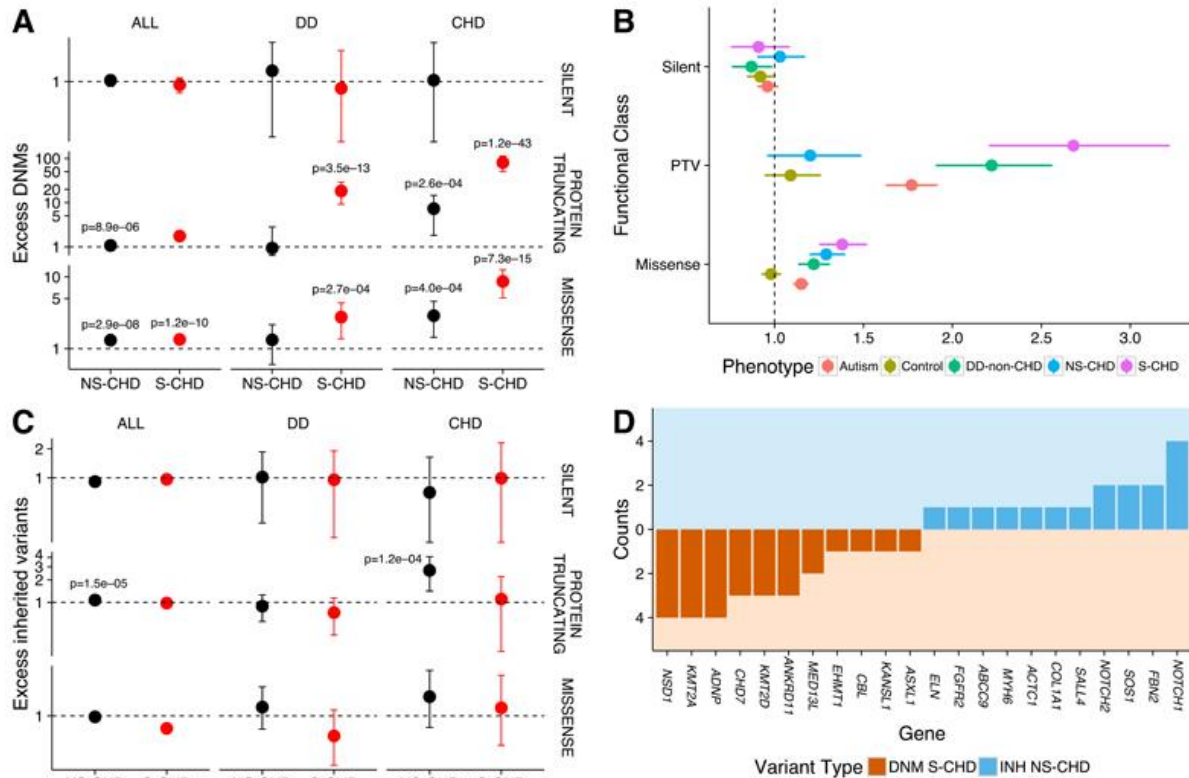
Figure 1

Burden of *de novo* and inherited variants in NS-CHD compared to S-CHD

(A) Excess of DNMs compared to null mutation model. Excess of DNMs was computed as the ratio of the observed number of DNMs over the expectation given random mutation using a null gene-wise mutation rate model. P-values were computed using a Poisson model parameterized by the cumulative mutation rate across the gene set for the same number of probands (n_{S-CHD}= 518, n_{NS-CHD}= 847). We stratify by variant consequence and within known autosomal dominant CHD genes (n=78), autosomal dominant developmental disorder genes excluding autosomal dominant CHD genes (n=203) and all autosomal protein coding genes excluding autosomal dominant developmental disorder and CHD genes (n=17,404). No data is shown for silent variants in CHD genes for syndromic probands as no variants were detected. Error bars represent the 95% confidence interval. (B) Comparison of exome-wide excess of DNMs across different diseases stratified by variant consequence. (C) Excess of rare inherited variants (n_{S-CHD}= 471, n_{NS-CHD}= 663) compared to 12,031 controls of matched ancestry: Excess of DNMs was computed as the ratio of the observed number of rare inherited variants over the expected numbers as seen in controls. (D) Counts of *de novo* PTVs in S-CHD probands and rare inherited (INH) PTVs in NS-CHD probands in known monoallelic CHD-associated genes.

To evaluate the contribution of incompletely penetrant inherited variants, we compared the burden of rare (Minor allele frequency < 0.1%) inherited variants in the three previously described gene sets in the S-CHD and NS-CHD cases of European ancestry, relative to population-matched controls (n=12,031, Supplementary Figure 1, Supplementary Table

6, Figure 1C). We observed a significant excess of rare inherited PTVs in autosomal dominant CHD-associated genes in NS-CHD (17 variants, OR=2.67,p=$1.1x10^{-4}$), but not in S-CHD (p=0.3). The CHD-associated genes with inherited PTVs in NS-CHD (Supplementary Table 7) have previously only been linked with non-syndromic or syndromic presentations with variable presentations, and were non-overlapping with genes with *de novo* PTVs in S-CHD (Figure 1D). Non-syndromic presentations of inherited PTVs in several genes originally associated with S-CHD have previously been described (e.g. *JAG1*[11], *TBX5*[12]). Moreover, we also observed an exome-wide excess of rare inherited PTVs (3,318 variants, OR=1.08, p=$1.51x10^{-5}$) in NS-CHD probands, even after excluding known CHD-associated and developmental disorder-associated genes, suggested incomplete penetrance in additional, novel CHD-associated genes. We did not observe this exome-wide excess in the S-CHD cohort (p=0.8), suggesting a more appreciable role for incomplete penetrance in NS-CHD than S-CHD.

Using a previously described null mutation model[6,9], we evaluated individual genes for an excess of *de novo* PTVs and *de novo* missense variants separately, using a high sensitivity set of candidate DNMs and defining genome-wide significance as p < $1.3x10^{-6}$. When considering all CHD trios (S-CHD and NS-CHD), including cases with mutations in known developmental disorder or CHD-associated genes, we identified 11 genes, with genome-wide significance. When we stratified by syndromic status we found no genes at genome-wide significance in the NS-CHD cohort. Conversely, we found the aforementioned 11 genes and one additional gene at genome-wide significance in the S-CHD cohort, in line with the described increased burden of DNM PTVs in this cohort (Table 1, Supplementary Table 8, Figure 2A). Nine of the 12 genome-wide significant genes were known to be associated with developmental disorders, although not all had previously been implicated in CHD. These findings expand the known phenotypic spectrum of several genes (e.g. S-CHD cases with *de novo* mutations in *TAB2,* a gene previously only described in NS-CHD[13]), however larger genotype-phenotype studies are needed to fully characterise the phenotypic spectrum associated with each gene. To maximise power to detect novel causative genes, we focused on 'unresolved' (i.e. probands without a plausible pathogenic DNM in known developmental disorder and CHD-associated genes) S-CHD trios (n=398) and identified three novel genes: *CDK13*, *CHD4* and *PRKD1*, at genome-wide significance (Table 1, Figure 2B, Supplementary Table 9). All candidate DNMs in these three genes were experimentally validated. We found no genes at genome-wide significance when we performed the analysis on 'unresolved' NS-CHD cases (n=792).
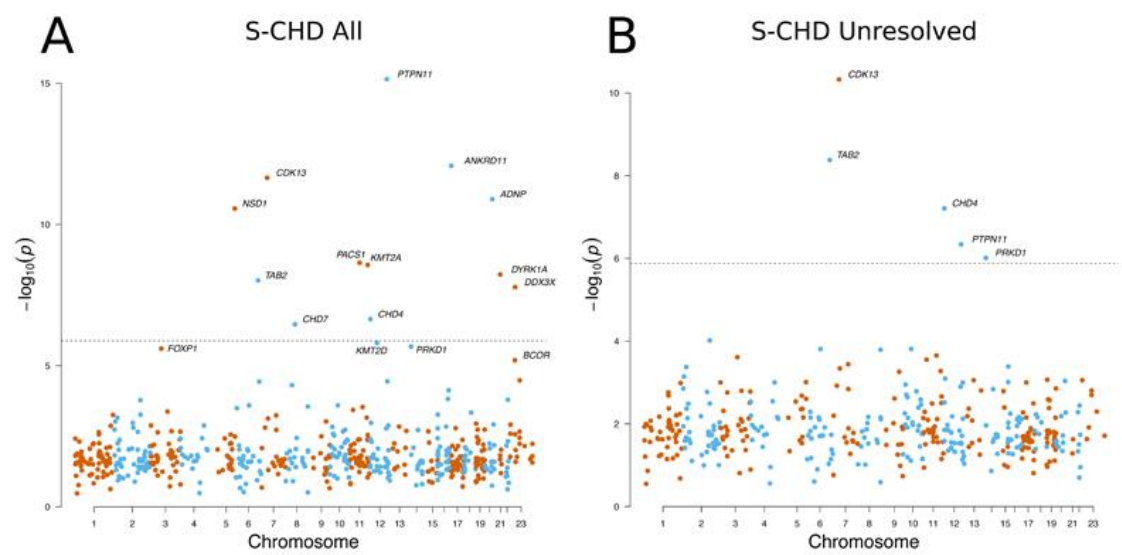
Figure 2

Gene-wise enrichment of *de novo* mutations

Gene-wise DNM enrichment was computed for A) the complete S-CHD cohort (n=518), B) 'unresolved' S-CHD trios without a plausible pathogenic DNM in known developmental disorder and CHD-associated genes (n=398). The probability of enrichment was computed given a Poisson distribution with the rate given by the gene-specific mutation rate multiplied by the number of chromosomes considered. This was performed for *de novo* PTVs and *de novo* missense variants independently. The *de novo* missense-enrichment probability was further combined with the probability of non-random clustering of *de novo* mutations using Fisher's method and the minimum was taken between the combined and the original p-value. The minimum probability (considering either *de novo* PTVs or *de novo* missense mutations) was plotted. The dashed horizontal line represents genome-wide significance (p<1.31x10^{-6}, Bonferronni corrected P-value of 0.05 corrected for 2x19,252 protein coding genes).

Chapter 9 Table 1

Genes with genome-wide significant enrichment of *de novo* mutations in the S-CHD cohort (n=518). Probabilities are also given for "unresolved" S-CHD cases (n=398). Missense mutations are considered significantly clustered if P<0.05.

Gene	DNMs (PTV/Missense)	Missense Clustering	P(S-CHD)	P(Unresolved)
PTPN11[D,C]	7 (0/7)	YES	7.29E–16	NA

Gene	DNMs (PTV/Missense)	Missense Clustering	P(S-CHD)	P(Unresolved)
ANKRD11[D,C]	5 (5/0)	NO	8.50E–13	NA
CDK13	**6 (0/6)**	**YES**	**2.26E–12**	**4.73E–11**
ADNP[D,C]	4 (4/0)	NO	1.29E–11	NA
NSD1[D,C]	6 (4/2)	YES	2.77E–11	NA
PACS1[D,C]	3 (0/3)	YES	2.32E–09	NA
KMT2A[D,C]	5 (4/1)	NO	2.74E–09	NA
TAB2[C]	3 (3/0)	NO	4.19E–09	NA
DYRK1A[D]	4 (3/1)	NO	5.99E–09	NA
DDX3X[D]	4 (2/2)	NO	1.69E–08	NA
CHD4	**5 (0/5)**	**NO**	**2.28E–07**	**6.18E–08**
CHD7[D,C]	4 (3/1)	NO	3.45E–07	NA

Gene	DNMs (PTV/Missense)	Missense Clustering	P(S-CHD)	P(Unresolved)
PRKD1	3 (0/3)	YES	2.13E–06	9.78E–07

[D]Associated with a developmental disorder
[C]Associated with CHD

We identified seven S-CHD individuals ([Figure 3A](#)) with clustered missense variants, six *de novo* variants and one variant of unknown inheritance, in the highly conserved serine/threonine protein kinase domain of cyclin-dependent kinase 13 (*CDK13*), which shows a marked depletion of missense variants in the European population ([Figure 3B](#)). Four probands carry an identical missense mutation (Asn842Ser). These seven S-CHD cases (6 trios and 1 singleton) were characterised by septal defects (VSD n= 2, ASD n= 5), with two also presenting with pulmonary valve abnormalities. Each had a recognizable facial gestalt, significant developmental delay, slight to moderate microcephaly and two had agenesis of the corpus callosum ([Figure 3A](#), [Supplementary Table 10](#)). Modelling of the kinase domain indicates that the observed mutations impair: ATP-binding, binding of the magnesium ion that is essential for enzymatic activity, or interactions with Cyclin K, with which *CDK13* forms a cyclin-dependent kinase complex ([Figure 3C](#)). This Cyclin K/CDK13 complex phosphorylates RNA polymerase II and is necessary for alternative splicing of RNA[14,15]. The knockout mice for *Cdk12*, the closest paralogue for *Cdk13*, both of which have ubiquitous developmental expression patterns, die at post-implantation (E5.5) suggesting a strong developmental effect[16].
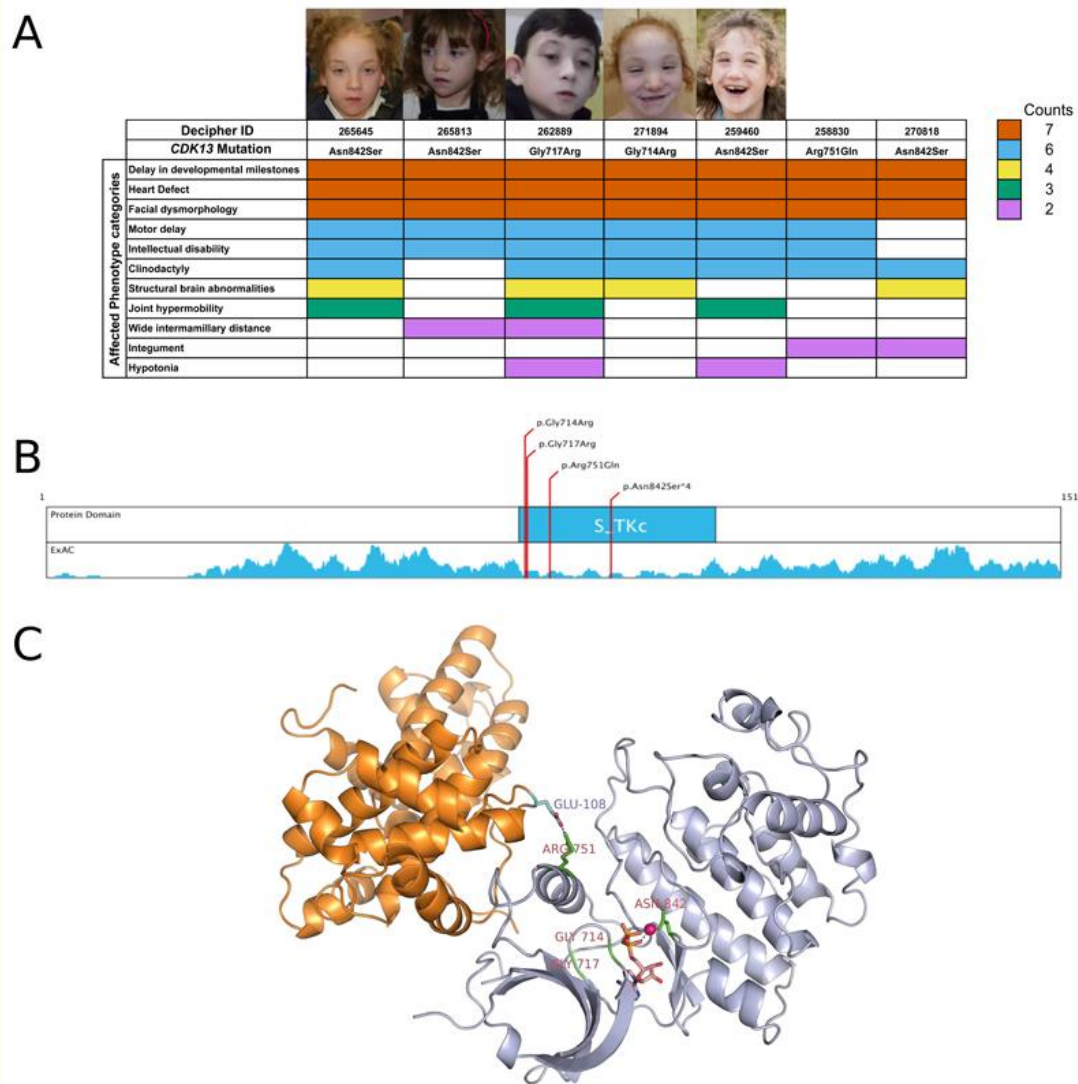
Figure 3

Overview of *CDK13* mutations in S–CHD cases

A) Phenotype summary of probands carrying missense mutations in *CDK13*. Colors indicate the number of times a certain phenotype was observed in individuals carrying a *de novo* mutation in *CDK13*. Photographs of affected probands are shown for which consent could be obtained for publication. B) clustering of DNMs in Serine-Threonine kinase domain. Density plot displays a sliding window (±10 amino acids) missense variant count in the Non-Finnish European population of the Exome Aggregation Consortium data, showing a marked reduction of missense variants in the kinase domain. C) 3D protein structure of *CDK13* by homology modelling adapted from *CDK12*. Mutated residues are marked in bright green. Catalysing Magnesium ion is highlighted in magenta, and the co-crystallized AMP ligand is portrayed in orange.

We observed five S-CHD individuals with DNMs in *CHD4* (4 missense variants and 1 in-frame deletion), which encodes a chromodomain containing protein that catalyses ATP-dependent chromatin remodelling as a core component of the nucleosome remodeling and histone deacetylase (NuRD) repressor complex[17]. Three patients manifested Tetralogy of Fallot or Fallot-like features, while the remaining two had an aortic coarctation and a septal defect (Supplementary Figure 2, Supplementary Table 11). All had significant early delay in

neurodevelopment, two had Chiari malformations and three of the four males had cryptorchidism or ambiguous genitalia. These features suggests an overlap with CHARGE syndrome (MIM #214800) caused by heterozygous loss-of-function mutations in the paralogous gene, *CHD7*, which also achieves significance in S-CHD cases ([Table 1]). Haploinsufficiency of another component of the NuRD complex, *GATAD2B*, has been identified as causing a recognisable intellectual disability syndrome, although associated CHD has not been reported[18]. More generally, several components of other ATP-dependent chromatin remodelling complexes have been associated with dominant developmental syndromes, including CHD in some patients[6,7]. A recent study showed that mice with endothelial knockdown of *CHD4*, resulting in a dysfunctional NuRD-complex, die of vascular rupture during midgestation[19]. This finding suggests NuRD-complex dysfunction as a possible mechanism for the observed human cardiac phenotype.

We identified three S-CHD individuals with *de novo* missense mutations in *PRKD1*, with two having identical DNMs, a mutational pattern suggestive of gain of function ([Supplementary Figure 3], [Supplementary Table 12]). Two out of the three individuals are affected by atrioventricular septal defects, whereas the third is affected by pulmonic stenosis. Other features included: severe developmental delay, ectodermal (dry skin, teeth and nail defects) and limb abnormalities. A homozygous PTV in *PRKD1* has recently been associated with truncus arteriosus through autozygosity mapping[20]. *PRKD1* encodes a serine/threonine protein kinase that regulates diverse cellular functions, including the transcriptional response to cardiac hypertrophy[21]. Homozygous knockout of *Prkd1* in mice is embryonic lethal and tissue-specific knockout results in abnormal cardiac remodelling[21].

The burden analyses described above clearly show enrichment for *de novo* PTVs, *de novo* missense variants and inherited PTVs within our CHD dataset. Therefore we hypothesised that some genes might be enriched for both *de novo* and rare inherited variants and that integrating both classes of variation, in trios and in singletons, using a previously described hierarchical Bayesian model[22] ([Online Methods]), may improve power to detect novel CHD-associated genes. We analysed PTVs and missense variants separately and considered candidate CHD-associated genes at strong (FDR < 1%), intermediate (1%< FDR < 5%) and weak (5%< FDR < 10%) levels of confidence ([Figure 4], [Supplementary Tables 13-14]). We found 16 genes at the strongest level of confidence, 12 were known developmental disorder-associated genes, 1 gene was only associated with CHD but not with developmental disorders (*MYH6*), and 3 are novel candidate genes (*CHD4*, *CDK13*, *DIAPH3*). Most high confidence genes, exhibited enrichment for either DNMs or inherited variants, only two genes, *NOTCH1* and *KAT6A* exhibited appreciable enrichment for both. *NOTCH1* was notable as being the only high confidence gene for which the evidence from inherited PTVs exceeds that from DNMs ([Figure 4B]). Due to the likely concentration of false discovery signals in novel gene associations, we believe this analysis alone to be insufficient to conclusively assert novel CHD associations. Additional functional evidence can prioritise genes for future follow-up studies ([Supplementary Table 15]). We evaluated the over-representation of particular gene functions and pathways among the top 374 genes with an FDR < 50% ([Online Methods]). We observed a significant (FDR < 10%) over-representation of genes associated with Gene Ontology terms relating to chromatin modification, protein phosphorylation, neural tube and cardiac development ([Supplementary Table 16]). Over-represented pathways included: *NOTCH1*-, *IGF1*-, *HDAC Class II*-, *ERBB*- and *NFKB*- signalling ([Supplementary Table 17]). In addition, the 374 top-ranking genes exhibited considerable functional coherence, with many genes forming a single large inter-connected subnetwork of high-confidence (STRING Score > 0.9) protein-protein interactions

([Supplementary Figure 4](#)), the degree of interconnection of which was significantly higher than expected by chance (p=5.84x10^{-3}). Key hubs in this subnetwork were *NOTCH1*, *SOS1*, *EP300* and *SMAD4*.
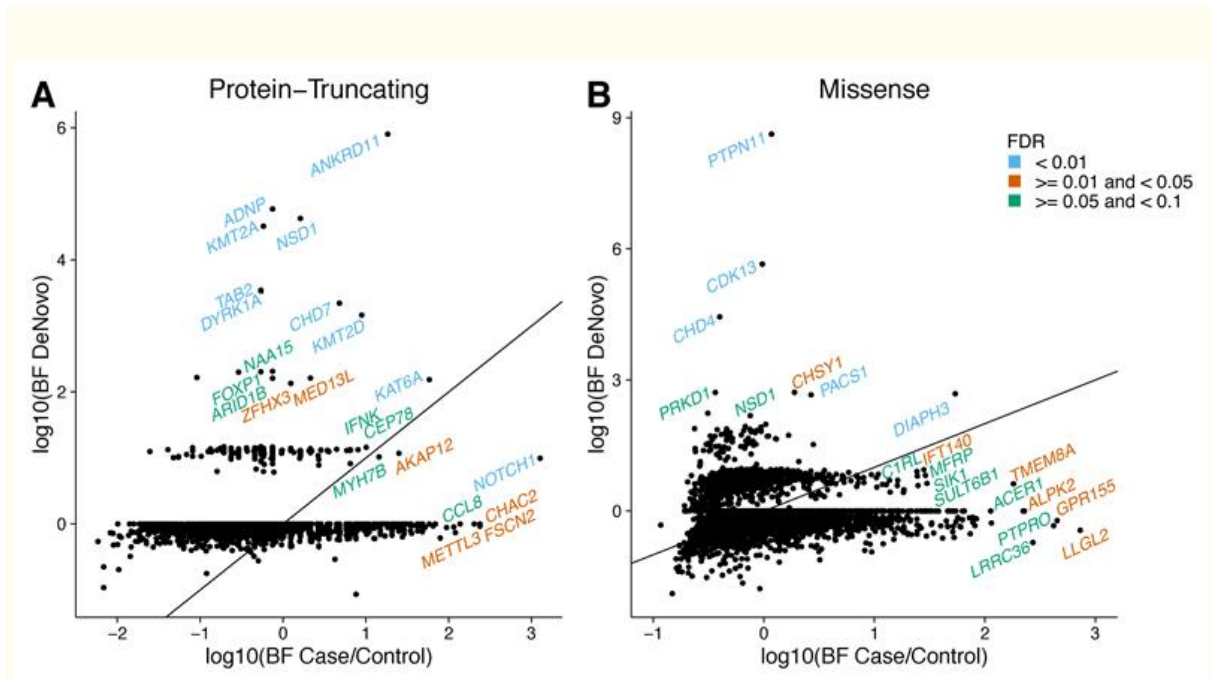
[Figure 4](#)

Integrated analysis of *de novo* and inherited variant enrichment using Hierarchical Bayesian modelling

Scatter plots representing Bayes factors (ratio of the evidence given the alternative model of the gene being associated with CHD over the evidence given the null model of the gene not being associated with CHD) for the *de novo* and inherited components of the model for PTVs and missense variants. The diagonal solid line represents the identity line, where equal signal is obtained from *de novo* variation compared to inherited variation. Genes at an FDR < 10% are labelled and colors represent different confidence thresholds.

Several mechanisms have been proposed to explain the low sibling recurrence risk of CHD, ranging from a major role for DNMs[7], incomplete penetrance of variants with large effect sizes, and a polygenic and/or multifactorial aetiology[23]. Our analyses (see [Supplementary Table 18](#) for an overview) show that the relative contributions of DNMs and incomplete penetrance differ markedly between NS-CHD and S-CHD, with a major role for *de novo* mutations in the latter, and inherited high-risk variants in the former. By focusing on unresolved S-CHD cases, we discovered three novel S-CHD disorders caused by mutations in genes not previously associated with S-CHD (*PRKD1*, *CHD4* and *CDK13*). CHD is often not fully penetrant in syndromic CHD disorders (e.g. *KMT2D*[24], *NSD1*[25]), and as all patients in our study were ascertained for CHD, further studies are necessary to quantify the penetrance of CHD in these three new syndromes. These three new genes increase the percentage of S-CHD probands with a putatively pathogenic DNM from 23% to 26% of patients, effectively increasing the diagnostic yield of this class of variation by 13%.

Current sample sizes provide limited statistical power to detect novel S-CHD disorders, and given the observed burden of *de novo* PTVs in S-CHD we estimate that data sets at least 20-fold larger will be needed to discover most dominant CHD-associated genes (Supplementary Figure 5). This challenge is likely to be even greater for identifying most genes harbouring incompletely penetrant variation in NS-CHD[26]. Our data motivate different study design strategies for S-CHD (trios) and NS-CHD (case/control), nonetheless international collaboration and data sharing will be essential to achieve a deeper understanding of the genetic architecture of CHD.

Acknowledgements

Abbreviations

CHD   Congenital Heart Defect

S-CHD   Syndromic CHD

NS-CHD Non-Syndromic CHD

PTV   Protein-truncating variant

DNM   *De Novo* Mutation

FDR   False Discovery Rate

T.S., G.J.S., C.F.W., H.V.F., J.C.B., K.D., D.R.F., J.D.B., M.E.H. interpreted the results. A.S., M-P.H., A.W., M.E.H. wrote the manuscript. M.E.H. supervised the project.

Competing financial interests

M.E.H. is a co-founder of, and holds shares in, Congenica Ltd, a genetics diagnostic company.

References

1. Hoffman JIE, Kaplan S. The incidence of congenital heart disease. *J Am Coll Cardiol.* 2002;39:1890–900. [PubMed] [Google Scholar]

2. Øyen N, et al. Recurrence of congenital heart defects in families. *Circulation.* 2009;120:295–301. [PubMed] [Google Scholar]

3. Gill HK, Splitt M, Sharland GK, Simpson JM. Patterns of recurrence of congenital heart disease: an analysis of 6,640 consecutive pregnancies evaluated by detailed fetal echocardiography. *J Am Coll Cardiol.* 2003;42:923–9. [PubMed] [Google Scholar]

4. Schulkey CE, et al. The maternal-age-associated risk of congenital heart disease is modifiable. *Nature.* 2015;520:230–3. [PMC free article] [PubMed] [Google Scholar]

5. Li Y, et al. Global genetic analysis in mice unveils central role for cilia in congenital heart disease. *Nature.* 2015;521:520–524. [PMC free article] [PubMed] [Google Scholar]

6. Fitzgerald TW, et al. Large-scale discovery of novel genetic causes of developmental disorders. *Nature.* 2014;519:223–228. [PMC free article] [PubMed] [Google Scholar]

7. Zaidi S, et al. De novo mutations in histone-modifying genes in congenital heart disease. *Nature.* 2013;498:220–3. [PMC free article] [PubMed] [Google Scholar]

8. Homsy J, et al. De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science (80-.)* 2015;350:1262–1266. [PMC free article] [PubMed] [Google Scholar]

9. Samocha KE, et al. A framework for the interpretation of de novo mutation in human disease. *Nat Genet.* 2014;46:944–950. [PMC free article] [PubMed] [Google Scholar]

10. Iossifov I, et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature.* 2014;515:216–21. [PMC free article] [PubMed] [Google Scholar]

11. Bauer RC, et al. Jagged1 (JAG1) mutations in patients with tetralogy of Fallot or pulmonic stenosis. *Hum Mutat.* 2010;31:594–601. [PMC free article] [PubMed] [Google Scholar]

12. Jia Y, et al. The diagnostic value of next generation sequencing in familial nonsyndromic congenital heart defects. *Am J Med Genet A.* 2015;167A:1822–9. [PubMed] [Google Scholar]

13. Thienpont B, et al. Haploinsufficiency of TAB2 causes congenital heart defects in humans. *Am J Hum Genet.* 2010;86:839–849. [PMC free article] [PubMed] [Google Scholar]

14. Liang K, et al. Characterization of human cyclin-dependent kinase 12 (CDK12) and CDK13 complexes in C-terminal domain phosphorylation, gene transcription, and RNA processing. *Mol Cell Biol.* 2015;35:928–38. [PMC free article] [PubMed] [Google Scholar]

15. Blazek D, et al. The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev.* 2011;25:2158–72. [PMC free article] [PubMed] [Google Scholar]

16. Chen H-H, Wang Y-C, Fann M-J. Identification and characterization of the CDK12/cyclin L1 complex involved in alternative splicing regulation. *Mol Cell Biol.* 2006;26:2736–45. [PMC free article] [PubMed] [Google Scholar]

17. Polo SE, Kaidi A, Baskcomb L, Galanty Y, Jackson SP. Regulation of DNA-damage responses and cell-cycle progression by the chromatin remodelling factor CHD4. *EMBO J.* 2010;29:3130–3139. [PMC free article] [PubMed] [Google Scholar]

18. de Ligt J, et al. Diagnostic Exome Sequencing in Persons with Severe Intellectual Disability. *N Engl J Med.* 2012;367:1921–1929. [PubMed] [Google Scholar]

19. Ingram KG, Curtis CD, Silasi-Mansat R, Lupu F, Griffin CT. The NuRD chromatin-remodeling enzyme CHD4 promotes embryonic vascular integrity by transcriptionally regulating extracellular matrix proteolysis. *PLoS Genet.* 2013;9:e1004031. [PMC free article] [PubMed] [Google Scholar]

20. Shaheen R, et al. Positional mapping of PRKD1, NRP1 and PRDM1 as novel candidate disease genes in truncus arteriosus. *J Med Genet.* 2015;52:322–9. [PubMed] [Google Scholar]

21. Fielitz J, et al. Requirement of protein kinase D1 for pathological cardiac remodeling. *Proc Natl Acad Sci U S A.* 2008;105:3059–63. [PMC free article] [PubMed] [Google Scholar]

22. He X, et al. Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. *PLoS Genet.* 2013;9:e1003671. [PMC free article] [PubMed] [Google Scholar]

23. Pierpont ME, et al. Genetic Basis for Congenital Heart Defects: Current Knowledge A Scientific Statement From the American Heart Association Congenital Cardiac Defects Committee, Council on Cardiovascular. *Circulation.* 2007;115:3015–3038. [PubMed] [Google Scholar]

24. Miyake N, et al. MLL2 and KDM6A mutations in patients with Kabuki syndrome. *Am J Med Genet A.* 2013;161A:2234–43. [PubMed] [Google Scholar]

25. Tatton-Brown K, et al. Genotype-phenotype associations in Sotos syndrome: an analysis of 266 individuals with NSD1 aberrations. *Am J Hum Genet.* 2005;77:193–204. [PMC free article] [PubMed] [Google Scholar]

26. Kiezun A, et al. Exome sequencing and the genetic basis of complex traits. *Nat Genet.* 2012;44:623–30. [PMC free article] [PubMed] [Google Scholar]

27. Giroud JM, et al. Report from the international society for nomenclature of paediatric and congenital heart disease: creation of a visual encyclopedia illustrating the terms and definitions of the international pediatric and congenital cardiac code. *World J Pediatr Congenit Heart Surg.* 2010;1:300–13. [PubMed] [Google Scholar]

28. Robinson PN, et al. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *Am J Hum Genet.* 2008;83:610–5. [PMC free article] [PubMed] [Google Scholar]

29. Moore C, et al. The INTERVAL trial to determine whether intervals between blood donations can be safely and acceptably decreased to optimise blood supply: study protocol for a randomised controlled trial. *Trials.* 2014;15:363. [PMC free article] [PubMed] [Google Scholar]

30. Wright CF, et al. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet.* 2014;385:1305–14. [PMC free article] [PubMed] [Google Scholar]

31. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2010;26:589–595. [PMC free article] [PubMed] [Google Scholar]

32. McKenna AH, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20:1297–303. [PMC free article] [PubMed] [Google Scholar]

33. Li H, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25:2078–2079. [PMC free article] [PubMed] [Google Scholar]

34. Ramu A, et al. DeNovoGear: de novo indel and point mutation discovery and phasing. *Nat Methods.* 2013;10:985–7. [PMC free article] [PubMed] [Google Scholar]

35. Akawi N, et al. Discovery of four recessive developmental disorders using probabilistic genotype and phenotype matching among 4,125 families. *Nat Genet.* 2015;47:1363–1369. [PMC free article] [PubMed] [Google Scholar]

36. Breuer K, et al. InnateDB: systems biology of innate immunity and beyond--recent updates and continuing curation. *Nucleic Acids Res.* 2013;41:D1228–33. [PMC free article] [PubMed] [Google Scholar]

37. Szklarczyk D, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015;43:D447–52. [PMC free article] [PubMed] [Google Scholar]

38. Bordoli L, et al. Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc.* 2009;4:1–13. [PubMed] [Google Scholar]

39. The Exome Aggregation Consortium et al. Analysis of protein-coding genetic variation in 60,706 humans. *bioRxiv.* 2015 doi: 10.1101/030338. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

Appendix G: Syndromic and Non-Syndromic Cohort Composition

		S-CHD			NS-CHD			TOTAL
		Male	Female	Total	Male	Female	Total	
ALL	Trios	286	232	518	508	339	847	1365
	Singletons	42	44	86	215	157	372	458
	Multisib	4	2	6 (3)	31	31	62 (29)	68 (32)
	Total	332	278	610	754	527	1281	1891
EU	Trios	261	210	471	412	251	663	1134
	Singletons	38	35	73	184	131	315	338
	Multisib	4	2	6 (3)	25	16	41 (19)	47 (22)
	Total	303	247	550	621	398	1019	1569

Table G.1 Cohort composition: Syndromic and non-syndromic

Supplementary Table 1 from Sifrim et al [12]: Proband counts stratified by syndromic status, gender, European ancestry and availability of parental exome sequencing data. For probands coming from families with multiple affected individuals the numbers represent the number of probands sequenced, whereas the numbers between brackets show the number of individual pedigrees.

Appendix H: Cohort Composition

Cohort name	Nr. Trios	Nr. Non-Trios	EGA / dbGAP access	PMID
DDD	501	92	EGAS00001000775	26437029, 25533962, 25529582
PCGC (dbGAP)	326	13	phs000571.v1.p1	23665959, 23410879
Nottingham	276	32	EGAS00001000762, EGAS00001000808	
Germany	93	105	EGAS00001000368	
Leuven	53	3	EGAD00001000796	
Saudi-Arabia	41	0	EGAS00001000544	
Newcastle	29	0	EGAD00001000344	
Royal Brompton Hospital	12	6	EGAD00001000797	
Toronto	0	86	EGAD00001000799	24702954, 25996639
UK10K	0	123	EGAS00001000125	26367797

Table H.1 Cohort composition: Participating centres.

Supplementary Table 2 from Sifrim et al [12]: Samples by cohort: Samples are stratified as being part of a parent-child trio or not (i.e. duo or singletons). If the dataset was previously published, as part of another study, the respective Pubmed IDs are given.

Appendix I: Cardiac Phenotypes

Description	HPO Term	Count TOTAL	Count S-CHD	Count NSCHD	% TOTAL	% S-CHD	% NS-CHD
Ventricular septal defect	HP:0001629	360	190	170	19.04%	31.15%	13.27%
Abnormality of the aortic valve	HP:0001646	335	59	276	17.72%	9.67%	21.55%
Defect in the atrial septum	HP:0001631	281	140	141	14.86%	22.95%	11.01%
Coarctation of aorta	HP:0001680	263	39	224	13.91%	6.39%	17.49%
Tetralogy of Fallot	HP:0001636	201	42	159	10.63%	6.89%	12.41%
Atrioventricular canal defect	HP:0006695	194	21	173	10.26%	3.44%	13.51%
Abnormality of the pulmonary valve	HP:0001641	169	68	101	8.94%	11.15%	7.88%
Hypoplastic left heart	HP:0004383	155	8	147	8.20%	1.31%	11.48%
Patent ductus arteriosus	HP:0001643	151	86	65	7.99%	14.10%	5.07%
Transposition of the great arteries with ventricular septal defect	HP:0011607	84	17	67	4.44%	2.79%	5.23%
Abnormal branching pattern of the aortic arch	HP:0011587	84	19	65	4.44%	3.11%	5.07%
Double outlet right ventricle	HP:0001719	68	11	57	3.60%	1.80%	4.45%
Situs inversus totalis	HP:0001696	66	13	53	3.49%	2.13%	4.14%
Abnormality of the mitral valve	HP:0001633	63	14	49	3.33%	2.30%	3.83%
Abnormality of the vena cava	HP:0005345	60	3	57	3.17%	0.49%	4.45%
Hypoplastic aortic arch	HP:0012304	59	3	56	3.12%	0.49%	4.37%
Transposition of the great arteries with intact ventricular septum	HP:0011606	51	3	48	2.70%	0.49%	3.75%
Pulmonary valve atresia	HP:0010882	43	1	42	2.27%	0.16%	3.28%
Abnormality of cardiac morphology	HP:0001627	37	36	1	1.96%	5.90%	0.08%
Cardiomyopathy	HP:0001638	35	28	7	1.85%	4.59%	0.55%

Abnormality of the coronary arteries	HP:0006704	35	1	34	1.85%	0.16%	2.65%
Abnormality of the left ventricular outflow tract	HP:0011103	33	5	28	1.75%	0.82%	2.19%
Tricuspid atresia	HP:0011662	29	7	22	1.53%	1.15%	1.72%
Abnormality of the tricuspid valve	HP:0001702	22	4	18	1.16%	0.66%	1.41%
Abnormality of the pulmonary artery	HP:0004414	22	3	19	1.16%	0.49%	1.48%
Total anomalous pulmonary venous return	HP:0005160	21	7	14	1.11%	1.15%	1.09%
Mitral atresia	HP:0011560	19	0	19	1.00%	0.00%	1.48%
Hypoplastic right heart	HP:0010954	17	0	17	0.90%	0.00%	1.33%
Double inlet left ventricle	HP:0011555	16	1	15	0.85%	0.16%	1.17%
Partial anomalous pulmonary venous return	HP:0010773	16	2	14	0.85%	0.33%	1.09%
Pulmonary artery atresia	HP:0004935	16	6	10	0.85%	0.98%	0.78%
Ebstein's anomaly of the tricuspid valve	HP:0010316	15	1	14	0.79%	0.16%	1.09%
Left atrial isomerism	HP:0011537	15	0	15	0.79%	0.00%	1.17%
Congenitally corrected transposition of the great arteries	HP:0011540	14	1	13	0.74%	0.16%	1.01%
Truncus arteriosus	HP:0001660	13	6	7	0.69%	0.98%	0.55%
Right atrial isomerism	HP:0011536	12	2	10	0.63%	0.33%	0.78%
Abnormality of the left ventricle	HP:0001711	10	8	2	0.53%	1.31%	0.16%
Hypoplasia of right ventricle	HP:0004762	9	2	7	0.48%	0.33%	0.55%
Arrhythmia	HP:0011675	5	5	0	0.26%	0.82%	0.00%
Abnormality of cardiac atrium	HP:0005120	4	3	1	0.21%	0.49%	0.08%
Interrupted aortic arch	HP:0011611	4	0	4	0.21%	0.00%	0.31%
Abnormality of the right ventricle	HP:0001707	3	3	0	0.16%	0.49%	0.00%

Table I.1 Composition of cohort: CHD phenotype.

Supplementary Table 3 reproduced from Sifrim et al. [12]: Counts of phenotype annotations across all probands (n=1891), the S-CHD cohort (n=610) and the NS-CHD cohort (n=1281). Percentages were computed as the number of probands annotated with that particular phenotype divided by the number of samples in the respective cohorts (all, S-CHD, NS-CHD).

Appendix J: CHD gene list criteria

Confirmed CHD Gene (Tier 1)	1	Plausible disease-causing mutations* within, affecting or encompassing an interpretable functional region** of a single gene identified in multiple (>3) unrelated cases/families with a cardiac phenotype***
	2	Plausible disease-causing mutations within, affecting or encompassing cis-regulatory elements convincingly affecting the expression of a single gene identified in multiple (>3) unrelated cases/families
	3	Point 1 and 2 of a Probable Cardiac Gene with convincing bioinformatic or functional evidence for causation e.g. animal model which recapitulates the human phenotype
Probable CHD Gene (Tier 2)	1	Plausible disease-causing mutations within, affecting or encompassing an interpretable functional region of a single gene identified in more than one (2 or 3) unrelated cases/families or segregation within multiple individuals within a single large family
	2	Plausible disease-causing mutations within, affecting or encompassing cis-regulatory elements convincingly affecting the expression of a single gene identified in in more than one (2 or 3) unrelated cases/families
	3	Possible Cardiac Gene (see below) with addition of convincing bioinformatic or functional evidence of causation e.g. animal model which recapitulates the human phenotype
Possible CHD Gene (Tier 2)	4	Plausible disease-causing mutations within, affecting or encompassing an interpretable functional region of a single gene identified in one case or segregation within multiple individuals within a small family
	5	Plausible disease-causing mutations within, affecting or encompassing cis-regulatory elements convincingly affecting the expression of a single gene identified in one case/family
	6	Possible disease-causing mutations within, affecting or encompassing an interpretable functional region of a single gene identified in more than one unrelated cases/families or segregation within multiple individuals

Table J.1 Criteria for CHD associated genes.

Supplementary Table 19 from Sifrim et al [12]: Criteria used for manual curation of genes associated with CHD.
* Recurrent *de novo* mutations convincingly affecting gene function, rare, fully-penetrant mutations - relevant genotype never seen in controls
** ORF in protein coding genes, miRNA stem or loop
*** All structural congenital heart defects (also PDA and ASD), not restricted to childhood. Including BAV and abnormalities of large vessels (Aorta root dilation and aortic aneurysm).

Appendix K: High Confidence CHD Gene List

symbol	ensg_id	description	tier	inheritance
GATA4	ENSG00000136574	CHD - Non-syndromic	1	Monoallelic
GATA6	ENSG00000141448	CHD - Non-syndromic	1	Monoallelic
GJA1	ENSG00000152661	CHD - Non-syndromic	1	Monoallelic/Biallelic
ACTA2	ENSG00000107796	CHD - Non-syndromic	1	Monoallelic
ACVR2B	ENSG00000114739	CHD - Non-syndromic	1	Monoallelic
CCDC103	ENSG00000167131	CHD - Non-syndromic	1	Biallelic
CCDC114	ENSG00000105479	CHD - Non-syndromic	1	Biallelic
CCDC151	ENSG00000198003	CHD - Non-syndromic	1	Biallelic
CCDC39	ENSG00000145075	CHD - Non-syndromic	1	Biallelic
CCDC40	ENSG00000141519	CHD - Non-syndromic	1	Biallelic
CFC1	ENSG00000136698	CHD - Non-syndromic	1	Monoallelic
CITED2	ENSG00000164442	CHD - Non-syndromic	1	Monoallelic
CRELD1	ENSG00000163703	CHD - Non-syndromic	1	Monoallelic
DNAAF3	ENSG00000167646	CHD - Non-syndromic	1	Biallelic
DYX1C1	ENSG00000256061	CHD - Non-syndromic	1	Biallelic
LEFTY2	ENSG00000143768	CHD - Non-syndromic	1	Monoallelic
MYH11	ENSG00000133392	CHD - Non-syndromic	1	Monoallelic
MYH7	ENSG00000092054	CHD - Non-syndromic	1	Monoallelic
NKX2-5	ENSG00000183072	CHD - Non-syndromic	1	Monoallelic
NODAL	ENSG00000156574	CHD - Non-syndromic	1	Monoallelic
NR2F2	ENSG00000185551	CHD - Non-syndromic	1	Monoallelic
NOTCH1	ENSG00000148400	CHD - Non-syndromic/Syndromic	1	Monoallelic
FLNA	ENSG00000269329	CHD - Non-syndromic/Syndromic	1	X-linked

ARID1A	ENSG00000117713	CHD - Syndromic	1	Monoallelic
NOTCH2	ENSG00000134250	CHD - Syndromic	1	Monoallelic
NSD1	ENSG00000165671	CHD - Syndromic	1	Monoallelic
ACTB	ENSG00000075624	CHD - Syndromic	1	Monoallelic
ADNP	ENSG00000101126	CHD - Syndromic	1	Monoallelic
AMER1	ENSG00000184675	CHD - Syndromic	1	X-linked
ASXL1	ENSG00000171456	CHD - Syndromic	1	Monoallelic
BCOR	ENSG00000183337	CHD - Syndromic	1	X-linked
CCBE1	ENSG00000183287	CHD - Syndromic	1	Biallelic
CD96	ENSG00000153283	CHD - Syndromic	1	Biallelic
CHD7	ENSG00000171316	CHD - Syndromic	1	Monoallelic
CHST14	ENSG00000169105	CHD - Syndromic	1	Biallelic
CHST3	ENSG00000122863	CHD - Syndromic	1	Biallelic
COL1A1	ENSG00000108821	CHD - Syndromic	1	Monoallelic
COL3A1	ENSG00000168542	CHD - Syndromic	1	Monoallelic
EFTUD2	ENSG00000108883	CHD - Syndromic	1	Monoallelic
EHMT1	ENSG00000181090	CHD - Syndromic	1	Monoallelic
ELN	ENSG00000262184	CHD - Syndromic	1	Monoallelic
EVC	ENSG00000072840	CHD - Syndromic	1	Biallelic
EVC2	ENSG00000173040	CHD - Syndromic	1	Biallelic
FBN2	ENSG00000138829	CHD - Syndromic	1	Monoallelic
FGFR2	ENSG00000066468	CHD - Syndromic	1	Monoallelic
FLNB	ENSG00000136068	CHD - Syndromic	1	Monoallelic/Biallelic
FOXC1	ENSG00000054598	CHD - Syndromic	1	Monoallelic
FOXC2	ENSG00000176692	CHD - Syndromic	1	Monoallelic
FOXF1	ENSG00000103241	CHD - Syndromic	1	Monoallelic

GLI3	ENSG00000106571	CHD - Syndromic	1	Monoallelic
GPC3	ENSG00000147257	CHD - Syndromic	1	X-linked
HDAC4	ENSG00000068024	CHD - Syndromic	1	Monoallelic
HRAS	ENSG00000174775	CHD - Syndromic	1	Monoallelic
IRX5	ENSG00000176842	CHD - Syndromic	1	Biallelic
JAG1	ENSG00000101384	CHD - Syndromic	1	Monoallelic
KANSL1	ENSG00000120071	CHD - Syndromic	1	Monoallelic
KDM6A	ENSG00000147050	CHD - Syndromic	1	Monoallelic
KMT2A	ENSG00000267910	CHD - Syndromic	1	Monoallelic
KMT2D	ENSG00000167548	CHD - Syndromic	1	Monoallelic
LTBP3	ENSG00000168056	CHD - Syndromic	1	Biallelic
MAP2K1	ENSG00000169032	CHD - Syndromic	1	Monoallelic
MAP2K2	ENSG00000126934	CHD - Syndromic	1	Monoallelic
MED12	ENSG00000184634	CHD - Syndromic	1	X-linked
MED13L	ENSG00000123066	CHD - Syndromic	1	Monoallelic
MEGF8	ENSG00000105429	CHD - Syndromic	1	Biallelic
MGP	ENSG00000111341	CHD - Syndromic	1	Biallelic
MID1	ENSG00000101871	CHD - Syndromic	1	X-linked
MKKS	ENSG00000125863	CHD - Syndromic	1	Biallelic
NEK1	ENSG00000137601	CHD - Syndromic	1	Biallelic
NF1	ENSG00000196712	CHD - Syndromic	1	Monoallelic
NIPBL	ENSG00000164190	CHD - Syndromic	1	Monoallelic
PEX1	ENSG00000127980	CHD - Syndromic	1	Biallelic
PEX10	ENSG00000157911	CHD - Syndromic	1	Biallelic
PEX12	ENSG00000108733	CHD - Syndromic	1	Biallelic
PEX2	ENSG00000164751	CHD - Syndromic	1	Biallelic

PEX26	ENSG00000215193	CHD - Syndromic	1	Biallelic
PEX5	ENSG00000139197	CHD - Syndromic	1	Biallelic
PTPN11	ENSG00000179295	CHD - Syndromic	1	Monoallelic
RAB23	ENSG00000112210	CHD - Syndromic	1	Biallelic
RAI1	ENSG00000108557	CHD - Syndromic	1	Monoallelic
RBM8A	ENSG00000265241	CHD - Syndromic	1	Biallelic
SALL1	ENSG00000103449	CHD - Syndromic	1	Monoallelic
SALL4	ENSG00000101115	CHD - Syndromic	1	Monoallelic
SH3PXD2B	ENSG00000174705	CHD - Syndromic	1	Biallelic
SHOC2	ENSG00000108061	CHD - Syndromic	1	Monoallelic
SKI	ENSG00000157933	CHD - Syndromic	1	Monoallelic
SLC2A10	ENSG00000197496	CHD - Syndromic	1	Biallelic
SMAD3	ENSG00000166949	CHD - Syndromic	1	Monoallelic
SMAD4	ENSG00000141646	CHD - Syndromic	1	Monoallelic
SMARCA2	ENSG00000080503	CHD - Syndromic	1	Monoallelic
SMARCA4	ENSG00000127616	CHD - Syndromic	1	Monoallelic
SMARCB1	ENSG00000099956	CHD - Syndromic	1	Monoallelic
SMARCE1	ENSG00000073584	CHD - Syndromic	1	Monoallelic
SMC1A	ENSG00000269384	CHD - Syndromic	1	Monoallelic
SOS1	ENSG00000115904	CHD - Syndromic	1	Monoallelic
SOX2	ENSG00000181449	CHD - Syndromic	1	Monoallelic
SOX9	ENSG00000125398	CHD - Syndromic	1	Monoallelic
STRA6	ENSG00000137868	CHD - Syndromic	1	Biallelic
TBX1	ENSG00000184058	CHD - Syndromic	1	Monoallelic
TBX20	ENSG00000164532	CHD - Syndromic	1	Monoallelic
TBX3	ENSG00000135111	CHD - Syndromic	1	Monoallelic

TBX5	ENSG00000089225	CHD - Syndromic	1	Monoallelic
TFAP2B	ENSG00000008196	CHD - Syndromic	1	Monoallelic
TGFBR1	ENSG00000106799	CHD - Syndromic	1	Monoallelic
TGFBR2	ENSG00000163513	CHD - Syndromic	1	Monoallelic
UBR1	ENSG00000159459	CHD - Syndromic	1	Biallelic
ZEB2	ENSG00000169554	CHD - Syndromic	1	Monoallelic
ZIC3	ENSG00000156925	CHD - Syndromic	1	X-linked
DHCR7	ENSG00000172893	CHD - Syndromic	1	Biallelic
GDF1	ENSG00000130283	CHD - Non-syndromic	2	Monoallelic/Biallelic
NAA15	ENSG00000164134	CHD - Non-syndromic	2	
NFATC1	ENSG00000131196	CHD - Non-syndromic	2	
NKX2-6	ENSG00000180053	CHD - Non-syndromic	2	Biallelic
PLXND1	ENSG00000004399	CHD - Non-syndromic	2	
PQBP1	ENSG00000268142	CHD - Non-syndromic	2	
SMAD2	ENSG00000175387	CHD - Non-syndromic	2	
TAB2	ENSG00000055208	CHD - Non-syndromic	2	Monoallelic
ATIC	ENSG00000138363	CHD - Syndromic	2	Biallelic
B3GALT6	ENSG00000176022	CHD - Syndromic	2	Biallelic
CEP290	ENSG00000198707	CHD - Syndromic	2	Biallelic
CEP57	ENSG00000166037	CHD - Syndromic	2	Biallelic
CHRNA1	ENSG00000138435	CHD - Syndromic	2	Monoallelic/Biallelic
CHUK	ENSG00000213341	CHD - Syndromic	2	Biallelic
COL2A1	ENSG00000139219	CHD - Syndromic	2	Monoallelic
COX7B	ENSG00000269083	CHD - Syndromic	2	
CREBBP	ENSG00000005339	CHD - Syndromic	2	Monoallelic
CTCF	ENSG00000102974	CHD - Syndromic	2	Monoallelic

DDX11	ENSG00000013573	CHD - Syndromic	2	Biallelic
DDX59	ENSG00000118197	CHD - Syndromic	2	Monoallelic
DIS3L2	ENSG00000144535	CHD - Syndromic	2	Biallelic
EOGT	ENSG00000163378	CHD - Syndromic	2	Biallelic
EP300	ENSG00000100393	CHD - Syndromic	2	Monoallelic
ESCO2	ENSG00000171320	CHD - Syndromic	2	Biallelic
FBLN5	ENSG00000140092	CHD - Syndromic	2	Biallelic
FIG4	ENSG00000112367	CHD - Syndromic	2	Biallelic
FKBP14	ENSG00000106080	CHD - Syndromic	2	Biallelic
FKTN	ENSG00000106692	CHD - Syndromic	2	Biallelic
GPC6	ENSG00000183098	CHD - Syndromic	2	Biallelic
IGFBP7	ENSG00000163453	CHD - Syndromic	2	Biallelic
KAT6B	ENSG00000156650	CHD - Syndromic	2	Monoallelic
KIF7	ENSG00000166813	CHD - Syndromic	2	Biallelic
LRP2	ENSG00000081479	CHD - Syndromic	2	Biallelic
LRP5	ENSG00000162337	CHD - Syndromic	2	Monoallelic/Biallelic
MGAT2	ENSG00000168282	CHD - Syndromic	2	Biallelic
MKS1	ENSG00000011143	CHD - Syndromic	2	Biallelic
MPLKIP	ENSG00000168303	CHD - Syndromic	2	Biallelic
NFIX	ENSG00000008441	CHD - Syndromic	2	Monoallelic
NPHP3	ENSG00000113971	CHD - Syndromic	2	Biallelic
NRXN1	ENSG00000179915	CHD - Syndromic	2	Monoallelic/Biallelic
NSDHL	ENSG00000269336	CHD - Syndromic	2	
PHGDH	ENSG00000092621	CHD - Syndromic	2	Biallelic
PKD1	ENSG00000008710	CHD - Syndromic	2	
PKD2	ENSG00000118762	CHD - Syndromic	2	Monoallelic

PTF1A	ENSG00000168267	CHD - Syndromic	2	Biallelic
RARB	ENSG00000077092	CHD - Syndromic	2	Monoallelic/Biallelic
RBM10	ENSG00000271352	CHD - Syndromic	2	
RNF135	ENSG00000181481	CHD - Syndromic	2	Monoallelic
RNU4ATAC	ENSG00000264229	CHD - Syndromic	2	Biallelic
ROR2	ENSG00000169071	CHD - Syndromic	2	Monoallelic/Biallelic
RPS19	ENSG00000272852	CHD - Syndromic	2	Monoallelic
RRAS	ENSG00000126458	CHD - Syndromic	2	
SETBP1	ENSG00000152217	CHD - Syndromic	2	Monoallelic
SF3B4	ENSG00000263977	CHD - Syndromic	2	Monoallelic
SHANK3	ENSG00000251322	CHD - Syndromic	2	Monoallelic
SNRPB	ENSG00000125835	CHD - Syndromic	2	Monoallelic
SRCAP	ENSG00000080603	CHD - Syndromic	2	Monoallelic
STAMBP	ENSG00000124356	CHD - Syndromic	2	Biallelic
TBC1D24	ENSG00000162065	CHD - Syndromic	2	Biallelic
TLL1	ENSG00000038295	CHD - Syndromic	2	Monoallelic
TTC37	ENSG00000198677	CHD - Syndromic	2	Biallelic
TWIST1	ENSG00000122691	CHD - Syndromic	2	Monoallelic
WDR60	ENSG00000126870	CHD - Syndromic	2	Biallelic
WNT5A	ENSG00000114251	CHD - Syndromic	2	Monoallelic
ZFPM2	ENSG00000169946	CHD - Syndromic	2	Monoallelic
ZMPSTE24	ENSG00000084073	CHD - Syndromic	2	Biallelic
ACTC1	ENSG00000159251	CHD/CM - Non-syndromic	1	Monoallelic
MYH6	ENSG00000197616	CHD/CM - Non-syndromic	1	Monoallelic
NRAS	ENSG00000213281	CHD/CM - Syndromic	1	Monoallelic

BRAF	ENSG00000157764	CHD/CM - Syndromic	1	Monoallelic
CBL	ENSG00000110395	CHD/CM - Syndromic	1	Monoallelic
HCCS	ENSG00000004961	CHD/CM - Syndromic	1	X-linked
KRAS	ENSG00000133703	CHD/CM - Syndromic	1	Monoallelic
BRAF1	NA	CHD/CM - Syndromic	1	Monoallelic
RIT1	ENSG00000143622	CHD/CM - Syndromic	1	Monoallelic
FTO	ENSG00000140718	CHD/CM - Syndromic	2	Biallelic
ABCC9	ENSG00000069431	CHD/CM - Syndromic	1	Monoallelic

Table K.1 CHD associated genes.
Supplementary table 20 from Sifrim et al [12]. Tier 1 and tier 2 genes that we are highly confident cause CHD in humans.

Appendix L. Clinical features reported with mutations in *CDK13*
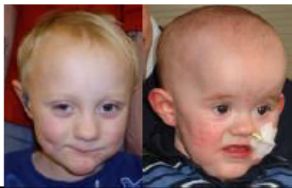
	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7
Decipher_ID/Study_ID	265645/DDDP101673	265813/DDDP104368	259460/DDDP111320	262889/DDDP111394	271894/DDDP112485	258830/DDDP102232	270818/DDDP112274
Mutations	NC_000007.13:g.40085606A>G; NM_031267.3:c.2525A>G; NP_112557.2:p.Asn842Ser	NC_000007.13:g.40085606A>G; NM_031267.3:c.2525A>G; NP_112557.2:p.Asn842Ser	NC_000007.13:g.40085606A>G; NM_031267.3:c.2525A>G; NP_112557.2:p.Asn842Ser	NC_000007.13:g.40039066G>A; NM_031267.3:c.2149G>A; NP_112557.2:p.Gly717Arg	NC_000007.13:g.40039057G>C; NM_031267.3:c.2140G>C; NP_112557.2:p.Gly714Arg	NC_000007.13:g.40041529G>A; NM_031267.3:c.2252G>A; NP_112557.2:p.Arg751Gln	NC_000007.13:g.40085606A>G; NM_031267.3:c.2525A>G; NP_112557.2:p.Asn842Ser
Family history	Singleton	Singleton	Singleton	Singleton	Singleton	Singleton	Unknown, only mother tested negative thus far
Sex	Female	Female	Female	Male	Female	Female	Male
Age at last assesment	7.18 years	0.19 years	3.21 years	7.83 years	4.98 years	11.87 years	1.25 years
Prenatal or Perinatal development							
Gestation	40 weeks	38 weeks	41 weeks	38 weeks	38 weeks	35 weeks	39 weeks
Birth weight	3640 grams (SD 0.51)	3500 grams (SD 1.16)	3310 grams (SD -0.20)	2000 grams (SD -2.65)	2636 grams (SD -0.89)	1700 grams (SD -1.93)	
Birth head circumference (BHC)		-	34 cm (SD -0.44)	35cm (0.4th - 2nd)	30.8 cm (SD -2.08)		
Complications	Contractions from 28/40 Polyhydramnion			IUGR			
Developmental milestones							
Social smile	16 weeks	-	6 weeks	Unknown	6 weeks	Unknown	Unknown
Sat Independently	3-4 years	-	11 months	12 months	12 months	13 months	11 months
Walked Independently	Not yet achieved (walking with a frame)	-	23 months	3-4 years	2-2.5 years	3-4 years	Unknown
First words	20 months	-	14 months	2-2.5 years	2.5-3 years	5 years and over	Unknown
Postnatal Growth							
Height/length	106.8 cm (SD -3.13)		81 cm (SD -1.05)	82.4 cm (SD -3.84)	108 cm (SD -0.06)	136 cm (SD -1.74)	
Weight	22.5 kg (SD -0.41)	4.39 kg (SD -1.16)	10.06 kg (SD -1.34)	11.96 kg (SD -2.08)	18 kg (SD -0.05)	41 kg (SD 0.24)	7.84 kg (SD -3.08)
Occipital Frontal Circumference (OFC)	51.2 cm (SD -1.27)	38 cm (SD -0.66)	47 cm (SD -1.42)	46.4 cm (SD -2.81)	49 cm (SD -2.21)	53 cm (SD -0.99)	44 cm (SD -3.71)
Clinical features							
Cardiovascular system							
Vasculature							
Cardiac Morphology	Ventricular septal defect (VSD)	Interatrial communication (ASD) and Pulmonary valvar abnormality	Interatrial communication (ASD) within oval fossa (secundum)	Ventricular septal defect (VSD) and Pulmonary valvar abnormality	Interatrial communication (ASD)	Interatrial communication (ASD) and Ventricular septal defect (VSD)	Interatrial communication (ASD) within oval fossa (secundum)
Cardiovascular system Physiology							
Fetal Cardiovascular System							
Head and Neck	Low posterior hairline	Redundant neck skin, Prominent occiput,		Postnatal microcephaly	Postnatal microcephaly	Plagiocephaly	Microcephaly
Facial	Hypertelorism, Upslanted palpebral fissure, Epicanthus, Thick eyebrow, Highly arched eyebrow, Prominent nasal bridge	Hypertelorism, Downturned corners of mouth, Short lingual frenulum, Nevus flammeus of the forehead, Telecanthus	Curly hair, Upslanted palpebral fissure, Hypertelorism, Narrow mouth, Thin upper lip vermilion, Thick upper lip vermilion, Bulbous nose	Plagiocephaly with mild facial assymetry. Epicanthus with wide nasal bridge with upslanting palpebral fissures, thin upper lip, long smooth philtrum and small mouth	Hypertelorism, Microdontia, Upslanted palpebral fissure, Wide nasal bridge, Short philtrum	Upslanted palpebral fissure, Downturned corners of mouth, narrow mouth, Blepharophimosis, Ptosis	Thin lips, Abnormality of the pinna

Nervous system							
Global developmental delay	Moderate	Yes	Yes 'spiky profile'	Yes	Yes	Yes	Moderate
Motor delay	Yes		Yes	Yes	Yes	Yes	
Intellictual disability	Yes	Yes	moderate	Yes	Moderate		
Behavioral/Psychiatric Manifestations	Autism		Pica				
Seizure	Absence seizures						
Structural	Periventricular leukomalacia		Spina bifida occulta	Underdevelopment of the front part of the brain and corpus callosum	Agenesis of corpus callosum	Aplasia of the inferior half of the cerebellar vermis, small cerebral cortex	Agenesis of corpus callosum
Neuromuscular	Spastic diplegia						
Eye/Vision	Strabismus			Strabismus	disassociated vertical deviation	Myopia, Strabismus	
Ear/Hearing	Low-set, posteriorly rotated ears, Preauricular pit			Dysplastic ears with cupped and posteriorly roatated ears.	Grommets		
Thoracic cavity							
Respiratory system			Central sleep apnoea requiring overnight oxygen. Obstructive element surgically cured (adenotonsillectomy)				
Breast		Wide intermamillary distance		Wide spaced nipples			
Abdomen	Gastroesophageal reflux, Constipation, Episodic vomiting						
Genitourinary system							
Limbs	Clinodactyly of the 5th finger		Second toe clinodactyly, Third toe clinodactyly, Clinodactyly of the 5th finger	Short fingers and toes with 2nd toe clinodactyly. Symmetrical skin folds like	Clinodactyly of the 5th finger	Camptodactyly (2nd-5th fingers), Carpal tunnel release	5th finger camptodactyly
Integument						Dry skin	Hypertrichosis
Skeletal system	Scoliosis, Hyperextensibility of the finger joints		Joint hypermobility	Hypotonia with joint hypermobility		Hyperlordosis	
Musculature	Lower limb hypertonia		Muscular hypotonia	Hypotonia			
Growth abnormality							Short stature
Connective tissue							
Endocrine system							
Immune system							
Blood and Blood-forming Tissues							
Metabolism/homeostasis						Mucopolysaccharides found on synovial biopsy but urine MPS screen and lysosomal enzyme screen normal, Abnormality of lysosomal metabolism	
Neoplasm							
Intial Clinical diagnosis		CHARGE syndrome		Ohdo considred but felt not to be typical. Charge also raised but did not have typical ear anomalies.	Feeding difficulties. PEG fed for some years.		Rubinstein-taybi syndrome
Additional Comments	Feeding difficulties		Velopharyngeal dysfunction causing unsafe swallow	Feeding difficulties requiring PEG insertion	Facially wondered about Ohdo but KAT6B negative	Truncal obesity, also has de novo TCF12 missense change	Feeding difficulties

Figure M.1 Clinical phenotypes of probands with *de novo PRKD1* mutations.

Supplementary Figure 3 from Sifrim et al [12].

  A. Clinical synopsis of the observed phenotypes across patients carrying *PRKD1* mutations. Columns represent single probands, shades of cells represent the number of probands sharing a phenotype in the given phenotypic categories. Photographs of affected probands are shown for which consent could be obtained for publication.

  B. Protein plot showing *PRKD1* protein domains and the distribution of *de novo* mutations. Two probands share identical missense mutations (Gly592Arg).
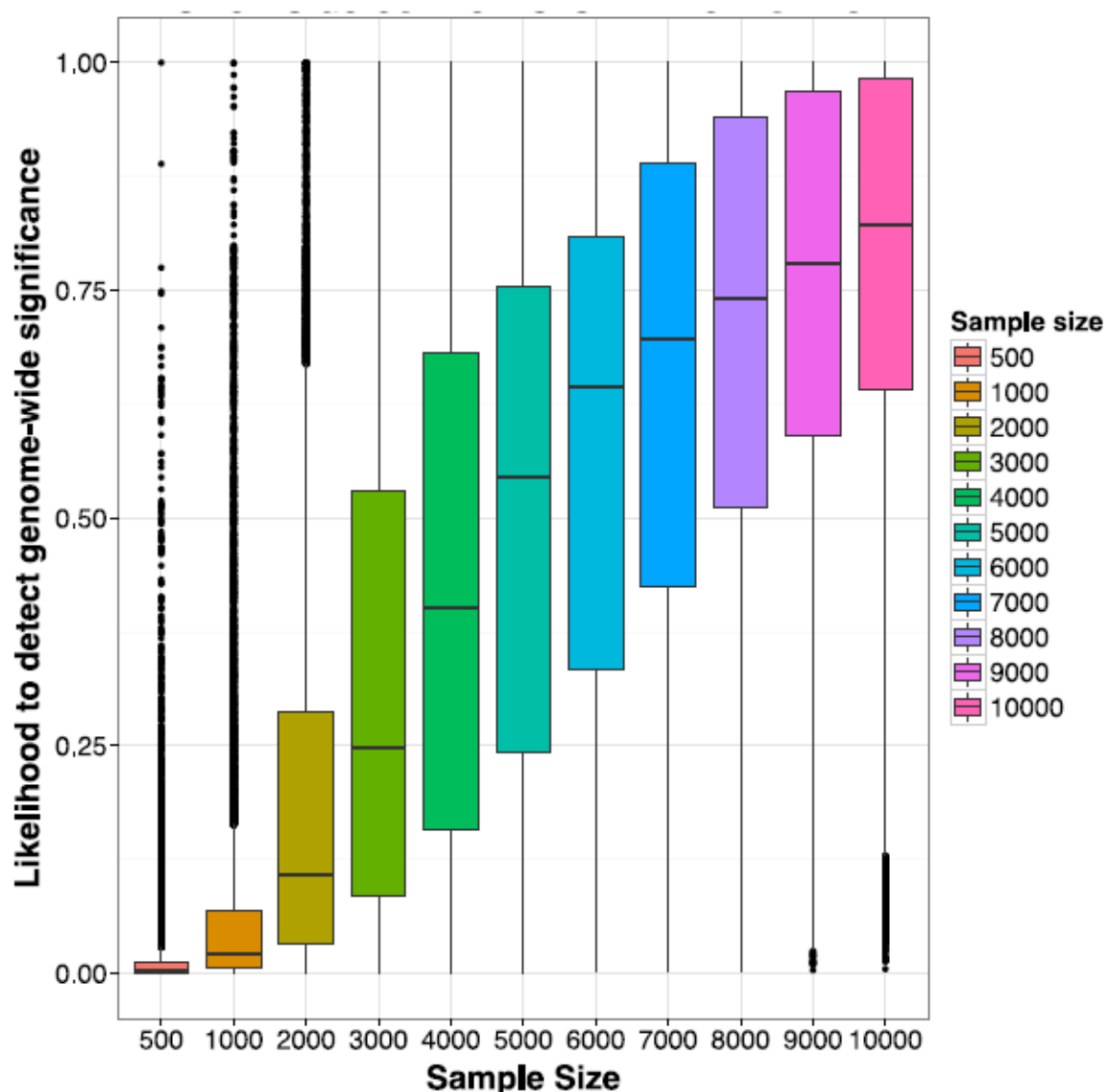
Figure N.1 Saturation analysis for detecting haploinsufficient S-CHD-associated genes.

Figure 5. Supplementary Figure 5 from Sifrim et al [12].

A box plot showing the distribution of statistical power to detect a significant enrichment of PTV mutations across 19252 genes in the genome, for different numbers of trios studied, from 500 trios to 10,000 trios. Line within the box shows the median, box corresponds to the first and third quartiles (the 25th and 75th percentiles) and whiskers correspond to most extreme values within 1.5 times the interquartile range from the box.

References

1. Hoffman, J.I. and S. Kaplan, *The incidence of congenital heart disease.* J Am Coll Cardiol, 2002. **39**(12): p. 1890-900.
2. van der Linde, D., et al., *Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis.* J Am Coll Cardiol, 2011. **58**(21): p. 2241-7.
3. Oyen, N., et al., *Recurrence of congenital heart defects in families.* Circulation, 2009. **120**(4): p. 295-301.
4. Gelb, B.D., *Genetic Discovery for Congenital Heart Defects*, in *Etiology and Morphogenesis of Congenital Heart Disease: From Gene Function and Cellular Interaction to Morphology*, T. Nakanishi, et al., Editors. 2016, Springer

Copyright 2016, The Author(s). Tokyo. p. 355-60.
5. Marelli, A.J., et al., *Lifetime prevalence of congenital heart disease in the general population from 2000 to 2010.* Circulation, 2014. **130**(9): p. 749-56.
6. Liu, Y., et al., *Global birth prevalence of congenital heart defects 1970-2017: updated systematic review and meta-analysis of 260 studies.* Int J Epidemiol, 2019. **48**(2): p. 455-463.
7. Ferencz, C., et al., *Congenital heart disease: prevalence at livebirth. The Baltimore-Washington Infant Study.* Am J Epidemiol, 1985. **121**(1): p. 31-6.
8. Mandalenakis, Z., et al., *Survivorship in Children and Young Adults With Congenital Heart Disease in Sweden.* JAMA Intern Med, 2017. **177**(2): p. 224-230.
9. Best, K.E. and J. Rankin, *Long-Term Survival of Individuals Born With Congenital Heart Disease: A Systematic Review and Meta-Analysis.* J Am Heart Assoc, 2016. **5**(6).
10. De Backer, J., et al., *Genetic counselling and testing in adults with congenital heart disease: A consensus document of the ESC Working Group of Grown-Up Congenital Heart Disease, the ESC Working Group on Aorta and Peripheral Vascular Disease and the European Society of Human Genetics.* Eur J Prev Cardiol, 2019: p. 2047487319854552.
11. Fahed, A.C., et al., *Genetics of congenital heart disease: the glass half empty.* Circ Res, 2013. **112**(4): p. 707-20.
12. Sifrim, A., et al., *Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing.* Nat Genet, 2016. **48**(9): p. 1060-5.
13. Schott, J.J., et al., *Congenital heart disease caused by mutations in the transcription factor NKX2-5.* Science, 1998. **281**(5373): p. 108-11.
14. Homsy, J., et al., *De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies.* Science, 2015. **350**(6265): p. 1262-6.
15. Jin, S.C., et al., *Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands.* Nat Genet, 2017. **49**(11): p. 1593-1601.
16. Zaidi, S., et al., *De novo mutations in histone-modifying genes in congenital heart disease.* Nature, 2013. **498**(7453): p. 220-3.
17. Soemedi, R., et al., *Contribution of global rare copy-number variants to the risk of sporadic congenital heart disease.* Am J Hum Genet, 2012. **91**(3): p. 489-501.
18. Greenway, S.C., et al., *De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot.* Nat Genet, 2009. **41**(8): p. 931-5.
19. Page, D.J., et al., *Whole Exome Sequencing Reveals the Major Genetic Contributors to Nonsyndromic Tetralogy of Fallot.* Circ Res, 2019. **124**(4): p. 553-563.

20. Kloesel, B., J.A. DiNardo, and S.C. Body, *Cardiac Embryology and Molecular Mechanisms of Congenital Heart Disease: A Primer for Anesthesiologists.* Anesth Analg, 2016. **123**(3): p. 551-69.

21. Tan, C.M.J. and A.J. Lewandowski, *The Transitional Heart: From Early Embryonic and Fetal Development to Neonatal Life.* Fetal Diagn Ther, 2020. **47**(5): p. 373-386.

22. Carmona, R., et al., *Comparative developmental biology of the cardiac inflow tract.* J Mol Cell Cardiol, 2018. **116**: p. 155-164.

23. M, K., *Cardiac Development*. First ed. 2007, New York, NY: Oxford University Press.

24. Kelly, R.G., *The second heart field.* Curr Top Dev Biol, 2012. **100**: p. 33-65.

25. Waldo, K.L., et al., *Secondary heart field contributes myocardium and smooth muscle to the arterial pole of the developing heart.* Dev Biol, 2005. **281**(1): p. 78-90.

26. Scherptong, R.W., et al., *Morphogenesis of outflow tract rotation during cardiac development: the pulmonary push concept.* Dev Dyn, 2012. **241**(9): p. 1413-22.

27. Dominguez, J.N., et al., *Asymmetric fate of the posterior part of the second heart field results in unexpected left/right contributions to both poles of the heart.* Circ Res, 2012. **111**(10): p. 1323-35.

28. Vincent, S.D. and M.E. Buckingham, *How to make a heart: the origin and regulation of cardiac progenitor cells.* Curr Top Dev Biol, 2010. **90**: p. 1-41.

29. Briggs, L.E., J. Kakarla, and A. Wessels, *The pathogenesis of atrial and atrioventricular septal defects with special emphasis on the role of the dorsal mesenchymal protrusion.* Differentiation, 2012. **84**(1): p. 117-30.

30. Creazzo, T.L., et al., *Role of cardiac neural crest cells in cardiovascular development.* Annu Rev Physiol, 1998. **60**: p. 267-86.

31. Plein, A., A. Fantin, and C. Ruhrberg, *Neural crest cells in cardiovascular development.* Curr Top Dev Biol, 2015. **111**: p. 183-200.

32. Hutson, M.R. and M.L. Kirby, *Model systems for the study of heart development and disease. Cardiac neural crest and conotruncal malformations.* Semin Cell Dev Biol, 2007. **18**(1): p. 101-10.

33. Nahirney, P.C., T. Mikawa, and D.A. Fischman, *Evidence for an extracellular matrix bridge guiding proepicardial cell migration to the myocardium of chick embryos.* Dev Dyn, 2003. **227**(4): p. 511-23.

34. Ratajska, A., E. Czarnowska, and B. Ciszek, *Embryonic development of the proepicardium and coronary vessels.* Int J Dev Biol, 2008. **52**(2-3): p. 229-36.

35. Mikawa, T. and R.G. Gourdie, *Pericardial mesoderm generates a population of coronary smooth muscle cells migrating into the heart along with ingrowth of the epicardial organ.* Dev Biol, 1996. **174**(2): p. 221-32.

36. Dettman, R.W., et al., *Common epicardial origin of coronary vascular smooth muscle, perivascular fibroblasts, and intermyocardial fibroblasts in the avian heart.* Dev Biol, 1998. **193**(2): p. 169-81.

37. Tomanek, R.J., H.K. Hansen, and E.I. Dedkov, *Vascular patterning of the quail coronary system during development.* Anat Rec A Discov Mol Cell Evol Biol, 2006. **288**(9): p. 989-99.

38. Schoenwolf G, B.S., Brauer P, Francis-West P. , *Development of the Heart Larsen's Human Embryology*. 2015, Philadelphia, PA: Churchill Livingstone.

39. Brade, T., et al., *Embryonic heart progenitors and cardiogenesis.* Cold Spring Harb Perspect Med, 2013. **3**(10): p. a013847.

40. Harvey, R.P., *Patterning the vertebrate heart.* Nat Rev Genet, 2002. **3**(7): p. 544-56.

41. Srivastava, D. and E.N. Olson, *A genetic blueprint for cardiac development.* Nature, 2000. **407**(6801): p. 221-6.

42. Kavka, A.I. and J.B. Green, *Tales of tails: Brachyury and the T-box genes.* Biochim Biophys Acta, 1997. **1333**(2): p. F73-84.

43. Arnold, S.J., et al., *Pivotal roles for eomesodermin during axis formation, epithelium-to-mesenchyme transition and endoderm specification in the mouse.* Development, 2008. **135**(3): p. 501-11.
44. Teo, A.K., et al., *Pluripotency factors regulate definitive endoderm specification through eomesodermin.* Genes Dev, 2011. **25**(3): p. 238-50.
45. Bondue, A., et al., *Mesp1 acts as a master regulator of multipotent cardiovascular progenitor specification.* Cell Stem Cell, 2008. **3**(1): p. 69-84.
46. Lindsley, R.C., et al., *Mesp1 coordinately regulates cardiovascular fate restriction and epithelial-mesenchymal transition in differentiating ESCs.* Cell Stem Cell, 2008. **3**(1): p. 55-68.
47. van den Ameele, J., et al., *Eomesodermin induces Mesp1 expression and cardiac differentiation from embryonic stem cells in the absence of Activin.* EMBO Rep, 2012. **13**(4): p. 355-62.
48. Costello, I., et al., *The T-box transcription factor Eomesodermin acts upstream of Mesp1 to specify cardiac mesoderm during mouse gastrulation.* Nat Cell Biol, 2011. **13**(9): p. 1084-91.
49. Liu, Y. and R.J. Schwartz, *Transient Mesp1 expression: a driver of cardiac cell fate determination.* Transcription, 2013. **4**(3): p. 92-6.
50. Bruneau, B.G., *Signaling and transcriptional networks in heart development and regeneration.* Cold Spring Harb Perspect Biol, 2013. **5**(3): p. a008292.
51. Zhu, H., *Forkhead box transcription factors in embryonic heart development and congenital heart disease.* Life Sci, 2016. **144**: p. 194-201.
52. McCulley, D.J. and B.L. Black, *Transcription factor pathways and congenital heart disease.* Curr Top Dev Biol, 2012. **100**: p. 253-77.
53. Zaffran, S. and M. Frasch, *Early signals in cardiac development.* Circ Res, 2002. **91**(6): p. 457-69.
54. Madabhushi, M. and E. Lacy, *Anterior visceral endoderm directs ventral morphogenesis and placement of head and heart via BMP2 expression.* Dev Cell, 2011. **21**(5): p. 907-19.
55. H.Q. Liu, W.Q.C., *Medical Developmental Biology* Third ed. 2012, Beijing: Science Press.
56. Yamagishi, H., et al., *Molecular embryology for an understanding of congenital heart diseases.* Anat Sci Int, 2009. **84**(3): p. 88-94.
57. Anderson, R.H., et al., *Insights from cardiac development relevant to congenital defects and adult clinical anatomy.* J Cardiovasc Transl Res, 2013. **6**(2): p. 107-17.
58. Kelly, R.G., M.E. Buckingham, and A.F. Moorman, *Heart fields and cardiac morphogenesis.* Cold Spring Harb Perspect Med, 2014. **4**(10).
59. Schlange, T., et al., *BMP2 is required for early heart development during a distinct time period.* Mech Dev, 2000. **91**(1-2): p. 259-70.
60. Ilagan, R., et al., *Fgf8 is required for anterior heart field development.* Development, 2006. **133**(12): p. 2435-45.
61. Saga, Y., et al., *MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube.* Development, 1999. **126**(15): p. 3437-47.
62. Kuo, C.T., et al., *GATA4 transcription factor is required for ventral morphogenesis and heart tube formation.* Genes Dev, 1997. **11**(8): p. 1048-60.
63. Molkentin, J.D., et al., *Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis.* Genes Dev, 1997. **11**(8): p. 1061-72.
64. Roebroek, A.J., et al., *Failure of ventral closure and axial rotation in embryos lacking the proprotein convertase Furin.* Development, 1998. **125**(24): p. 4863-76.
65. Constam, D.B. and E.J. Robertson, *Tissue-specific requirements for the proprotein convertase furin/SPC1 during embryonic turning and heart looping.* Development, 2000. **127**(2): p. 245-54.
66. Farraj KL, Z.R., *Embryology, Heart Tube*. Stat Pearls. 2018, Treasure Island FL: StatPearls Publishing.

67. Tyser, R.C., et al., *Calcium handling precedes cardiac differentiation to initiate the first heartbeat.* Elife, 2016. **5**.

68. Sylva, M., M.J. van den Hoff, and A.F. Moorman, *Development of the human heart.* Am J Med Genet A, 2014. **164A**(6): p. 1347-71.

69. Pandur, P., et al., *Islet1-expressing cardiac progenitor cells: a comparison across species.* Dev Genes Evol, 2013. **223**(1-2): p. 117-29.

70. Yoshiba, S. and H. Hamada, *Roles of cilia, fluid flow, and Ca2+ signaling in breaking of left-right symmetry.* Trends Genet, 2014. **30**(1): p. 10-7.

71. Hirokawa, N., et al., *Nodal flow and the generation of left-right asymmetry.* Cell, 2006. **125**(1): p. 33-45.

72. Amack, J.D., *Salient features of the ciliated organ of asymmetry.* Bioarchitecture, 2014. **4**(1): p. 6-15.

73. Shiraishi, I. and H. Ichikawa, *Human heterotaxy syndrome - from molecular genetics to clinical features, management, and prognosis.* Circ J, 2012. **76**(9): p. 2066-75.

74. Ryan, A.K., et al., *Pitx2 determines left-right asymmetry of internal organs in vertebrates.* Nature, 1998. **394**(6693): p. 545-51.

75. Shiratori, H. and H. Hamada, *TGFbeta signaling in establishing left-right asymmetry.* Semin Cell Dev Biol, 2014. **32**: p. 80-4.

76. Bamforth, S.D., et al., *Cited2 controls left-right patterning and heart development through a Nodal-Pitx2c pathway.* Nat Genet, 2004. **36**(11): p. 1189-96.

77. Weninger, W.J., et al., *Cited2 is required both for heart morphogenesis and establishment of the left-right axis in mouse development.* Development, 2005. **132**(6): p. 1337-48.

78. Weiss, A. and L. Attisano, *The TGFbeta superfamily signaling pathway.* Wiley Interdiscip Rev Dev Biol, 2013. **2**(1): p. 47-63.

79. Wrana, J.L., *Signaling by the TGFbeta superfamily.* Cold Spring Harb Perspect Biol, 2013. **5**(10): p. a011197.

80. Attisano, L. and J.L. Wrana, *Signal transduction by the TGF-beta superfamily.* Science, 2002. **296**(5573): p. 1646-7.

81. Waardenberg, A.J., et al., *Genetic networks governing heart development.* Cold Spring Harb Perspect Med, 2014. **4**(11): p. a013839.

82. Breckenridge, R.A., T.J. Mohun, and E. Amaya, *A role for BMP signalling in heart looping morphogenesis in Xenopus.* Dev Biol, 2001. **232**(1): p. 191-203.

83. Bisgrove, B.W., et al., *Polaris and Polycystin-2 in dorsal forerunner cells and Kupffer's vesicle are required for specification of the zebrafish left-right axis.* Dev Biol, 2005. **287**(2): p. 274-88.

84. Ramsdell, A.F., *Left-right asymmetry and congenital cardiac defects: getting to the heart of the matter in vertebrate left-right axis determination.* Dev Biol, 2005. **288**(1): p. 1-20.

85. Simard, A., et al., *The Pitx2c N-terminal domain is a critical interaction domain required for asymmetric morphogenesis.* Dev Dyn, 2009. **238**(10): p. 2459-70.

86. Noseda, M., et al., *Cardiopoietic factors: extracellular signals for cardiac lineage commitment.* Circ Res, 2011. **108**(1): p. 129-52.

87. Pereira, F.A., et al., *The orphan nuclear receptor COUP-TFII is required for angiogenesis and heart development.* Genes Dev, 1999. **13**(8): p. 1037-49.

88. Xavier-Neto, J., et al., *A retinoic acid-inducible transgenic marker of sino-atrial development in the mouse heart.* Development, 1999. **126**(12): p. 2677-87.

89. Bruneau, B.G., et al., *Cardiac expression of the ventricle-specific homeobox gene Irx4 is modulated by Nkx2-5 and dHand.* Dev Biol, 2000. **217**(2): p. 266-77.

90. Mommersteeg, M.T., et al., *Molecular pathway for the localized formation of the sinoatrial node.* Circ Res, 2007. **100**(3): p. 354-62.

91. Wang, J., et al., *Pitx2 prevents susceptibility to atrial arrhythmias by inhibiting left-sided pacemaker specification.* Proc Natl Acad Sci U S A, 2010. **107**(21): p. 9753-8.

92.   Takeuchi, J.K., et al., *Tbx5 specifies the left/right ventricles and ventricular septum position during cardiogenesis.* Development, 2003. **130**(24): p. 5953-64.

93.   McFadden, D.G., et al., *The Hand1 and Hand2 transcription factors regulate expansion of the embryonic cardiac ventricles in a gene dosage-dependent manner.* Development, 2005. **132**(1): p. 189-201.

94.   Lin, C.J., et al., *Partitioning the heart: mechanisms of cardiac septation and valve development.* Development, 2012. **139**(18): p. 3277-99.

95.   Wessels, A., et al., *Atrial development in the human heart: an immunohistochemical study with emphasis on the role of mesenchymal tissues.* Anat Rec, 2000. **259**(3): p. 288-300.

96.   Anderson, R.H., et al., *Development of the heart: (2) Septation of the atriums and ventricles.* Heart, 2003. **89**(8): p. 949-58.

97.   Schleich, J.M., et al., *An overview of cardiac morphogenesis.* Arch Cardiovasc Dis, 2013. **106**(11): p. 612-23.

98.   Bosada, F.M., et al., *Wnt/β-catenin signaling enables developmental transitions during valvulogenesis.* Development, 2016. **143**(6): p. 1041-54.

99.   Steimle, J.D. and I.P. Moskowitz, *TBX5: A Key Regulator of Heart Development.* Curr Top Dev Biol, 2017. **122**: p. 195-221.

100.   Zhang, K.K., et al., *Gene network and familial analyses uncover a gene network involving Tbx5/Osr1/Pcsk6 interaction in the second heart field for atrial septation.* Hum Mol Genet, 2016. **25**(6): p. 1140-51.

101.   Zhou, L., et al., *Gata4 potentiates second heart field proliferation and Hedgehog signaling for cardiac septation.* Proc Natl Acad Sci U S A, 2017. **114**(8): p. E1422-e1431.

102.   Xie, L., et al., *Tbx5-hedgehog molecular networks are essential in the second heart field for atrial septation.* Dev Cell, 2012. **23**(2): p. 280-91.

103.   Bruneau, B.G., et al., *A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease.* Cell, 2001. **106**(6): p. 709-21.

104.   Allen, B.L., T. Tenzen, and A.P. McMahon, *The Hedgehog-binding proteins Gas1 and Cdo cooperate to positively regulate Shh signaling during mouse development.* Genes Dev, 2007. **21**(10): p. 1244-57.

105.   Posch, M.G., et al., *Molecular genetics of congenital atrial septal defects.* Clin Res Cardiol, 2010. **99**(3): p. 137-47.

106.   Matsson, H., et al., *Alpha-cardiac actin mutations produce atrial septal defects.* Hum Mol Genet, 2008. **17**(2): p. 256-65.

107.   Ching, Y.H., et al., *Mutation in myosin heavy chain 6 causes atrial septal defect.* Nat Genet, 2005. **37**(4): p. 423-8.

108.   Posch, M.G., et al., *Cardiac alpha-myosin (MYH6) is the predominant sarcomeric disease gene for familial atrial septal defects.* PLoS One, 2011. **6**(12): p. e28872.

109.   Snarr, B.S., et al., *Isl1 expression at the venous pole identifies a novel role for the second heart field in cardiac development.* Circ Res, 2007. **101**(10): p. 971-4.

110.   Briggs, L.E., et al., *Expression of the BMP receptor Alk3 in the second heart field is essential for development of the dorsal mesenchymal protrusion and atrioventricular septation.* Circ Res, 2013. **112**(11): p. 1420-32.

111.   Annabi MR, M.A., *Embryology, Atrioventricular Septum.* StatPearls. 2019, Treasure Island (FL): StatPearls Publishing.

112.   Gittenberger-de Groot, A.C., et al., *Basics of cardiac development for the understanding of congenital heart malformations.* Pediatr Res, 2005. **57**(2): p. 169-76.

113.   Saremi, F., et al., *Fibrous Skeleton of the Heart: Anatomic Overview and Evaluation of Pathologic Conditions with CT and MR Imaging.* Radiographics, 2017. **37**(5): p. 1330-1351.

114.   Anderson, R.H., et al., *The development of septation in the four-chambered heart.* Anat Rec (Hoboken), 2014. **297**(8): p. 1414-29.

115. Franco, D., et al., *Left and right ventricular contributions to the formation of the interventricular septum in the mouse heart.* Dev Biol, 2006. **294**(2): p. 366-75.
116. Koshiba-Takeuchi, K., et al., *Reptilian heart development and the molecular basis of cardiac chamber evolution.* Nature, 2009. **461**(7260): p. 95-8.
117. Rana, M.S., V.M. Christoffels, and A.F. Moorman, *A molecular and genetic outline of cardiac morphogenesis.* Acta Physiol (Oxf), 2013. **207**(4): p. 588-615.
118. Wu, M., *Mechanisms of Trabecular Formation and Specification During Cardiogenesis.* Pediatr Cardiol, 2018. **39**(6): p. 1082-1089.
119. Manasek, F.J., *Embryonic development of the heart. I. A light and electron microscopic study of myocardial development in the early chick embryo.* J Morphol, 1968. **125**(3): p. 329-65.
120. Van Mierop, L.H., *Embryology of the univentricular heart.* Herz, 1979. **4**(2): p. 78-85.
121. Sedmera, D., et al., *Developmental patterning of the myocardium.* Anat Rec, 2000. **258**(4): p. 319-37.
122. Icardo, J.M. and A. Fernandez-Terán, *Morphologic study of ventricular trabeculation in the embryonic chick heart.* Acta Anat (Basel), 1987. **130**(3): p. 264-74.
123. Chen, H., et al., *BMP10 is essential for maintaining cardiac growth during murine cardiogenesis.* Development, 2004. **131**(9): p. 2219-31.
124. Gassmann, M., et al., *Aberrant neural and cardiac development in mice lacking the ErbB4 neuregulin receptor.* Nature, 1995. **378**(6555): p. 390-4.
125. Meyer, D. and C. Birchmeier, *Multiple essential functions of neuregulin in development.* Nature, 1995. **378**(6555): p. 386-90.
126. Grego-Bessa, J., et al., *Notch signaling is essential for ventricular chamber development.* Dev Cell, 2007. **12**(3): p. 415-29.
127. Wang, H.U., Z.F. Chen, and D.J. Anderson, *Molecular distinction and angiogenic interaction between embryonic arteries and veins revealed by ephrin-B2 and its receptor Eph-B4.* Cell, 1998. **93**(5): p. 741-53.
128. VanDusen, N.J., et al., *Hand2 is an essential regulator for two Notch-dependent functions within the embryonic endocardium.* Cell Rep, 2014. **9**(6): p. 2071-83.
129. Xin, M., et al., *Regulation of insulin-like growth factor signaling by Yap governs cardiomyocyte proliferation and embryonic heart size.* Sci Signal, 2011. **4**(196): p. ra70.
130. Zhao, C., et al., *Numb family proteins are essential for cardiac morphogenesis and progenitor differentiation.* Development, 2014. **141**(2): p. 281-95.
131. Ma, L., et al., *Bmp2 is essential for cardiac cushion epithelial-mesenchymal transition and myocardial patterning.* Development, 2005. **132**(24): p. 5601-11.
132. Hurle, J.M., E. Colveé, and A.M. Blanco, *Development of mouse semilunar valves.* Anat Embryol (Berl), 1980. **160**(1): p. 83-91.
133. Restivo, A., et al., *Cardiac outflow tract: a review of some embryogenetic aspects of the conotruncal region of the heart.* Anat Rec A Discov Mol Cell Evol Biol, 2006. **288**(9): p. 936-43.
134. Okamoto, N., et al., *Formal genesis of the outflow tracts of the heart revisited: previous works in the light of recent observations.* Congenit Anom (Kyoto), 2010. **50**(3): p. 141-58.
135. de Lange, F.J., et al., *Lineage and morphogenetic analysis of the cardiac valves.* Circ Res, 2004. **95**(6): p. 645-54.
136. Stankunas, K., et al., *VEGF signaling has distinct spatiotemporal roles during heart valve development.* Dev Biol, 2010. **347**(2): p. 325-36.
137. LaHaye, S., J. Lincoln, and V. Garg, *Genetics of valvular heart disease.* Curr Cardiol Rep, 2014. **16**(6): p. 487.
138. Rivera-Feliciano, J., et al., *Development of heart valves requires Gata4 expression in endothelial-derived cells.* Development, 2006. **133**(18): p. 3607-18.

139. Tao, G., et al., *Mmp15 is a direct target of Snai1 during endothelial to mesenchymal transformation and endocardial cushion development.* Dev Biol, 2011. **359**(2): p. 209-21.
140. Luna-Zurita, L., et al., *Integration of a Notch-dependent mesenchymal gene program and Bmp2-driven cell invasiveness regulates murine cardiac valve formation.* J Clin Invest, 2010. **120**(10): p. 3493-507.
141. Chang, C.P., et al., *A field of myocardial-endocardial NFAT signaling underlies heart valve morphogenesis.* Cell, 2004. **118**(5): p. 649-63.
142. Lincoln, J., et al., *Sox9 is required for precursor cell expansion and extracellular matrix organization during mouse heart valve development.* Dev Biol, 2007. **305**(1): p. 120-32.
143. Levay, A.K., et al., *Scleraxis is required for cell lineage differentiation and extracellular matrix remodeling during murine heart valve formation in vivo.* Circ Res, 2008. **103**(9): p. 948-56.
144. Schleich, J.M., *Images in cardiology. Development of the human heart: days 15-21.* Heart, 2002. **87**(5): p. 487.
145. Kau, T., et al., *Aortic development and anomalies.* Semin Intervent Radiol, 2007. **24**(2): p. 141-52.
146. Rosen RD, B.B., *Embryology, Aortic Arch*, ed. StatPearls. 2020, Treasure Island FL: StatPearls Publishing.
147. S, K., *Regional anatomy of the thoracic aorta.* Atlas of Normal and Variant Angiographic Anatomy. , ed. K. S. 1991, Philadelphia: WB Saunders.
148. Khalid, N. and B. Bordoni, *Embryology, Great Vessel*, in *StatPearls*. 2020, StatPearls Publishing

Copyright © 2020, StatPearls Publishing LLC.: Treasure Island (FL).
149. Tomanek, R. and P. Angelini, *Embryology of coronary arteries and anatomy/pathophysiology of coronary anomalies. A comprehensive update.* Int J Cardiol, 2019. **281**: p. 28-34.
150. Nanka, O., et al., *Experimental hypoxia and embryonic angiogenesis.* Dev Dyn, 2006. **235**(3): p. 723-33.
151. Yue, X. and R.J. Tomanek, *Stimulation of coronary vasculogenesis/angiogenesis by hypoxia in cultured embryonic hearts.* Dev Dyn, 1999. **216**(1): p. 28-36.
152. Yue, X. and R.J. Tomanek, *Effects of VEGF(165) and VEGF(121) on vasculogenesis and angiogenesis in cultured embryonic quail hearts.* Am J Physiol Heart Circ Physiol, 2001. **280**(5): p. H2240-7.
153. Lavine, K.J. and D.M. Ornitz, *Shared circuitry: developmental signaling cascades regulate both embryonic and adult coronary vasculature.* Circ Res, 2009. **104**(2): p. 159-69.
154. Lavine, K.J. and D.M. Ornitz, *Rebuilding the coronary vasculature: hedgehog as a new candidate for pharmacologic revascularization.* Trends Cardiovasc Med, 2007. **17**(3): p. 77-83.
155. Ando, K., et al., *Development of proximal coronary arteries in quail embryonic heart: multiple capillaries penetrating the aortic sinus fuse to form main coronary trunk.* Circ Res, 2004. **94**(3): p. 346-52.
156. Velkey, J.M. and D.H. Bernanke, *Apoptosis during coronary artery orifice development in the chick embryo.* Anat Rec, 2001. **262**(3): p. 310-7.
157. Tian, X., et al., *Peritruncal coronary endothelial cells contribute to proximal coronary artery stems and their aortic orifices in the mouse heart.* PLoS One, 2013. **8**(11): p. e80857.
158. Wothe, D., et al., *Increased coronary blood flow signals growth of coronary resistance vessels in near-term ovine fetuses.* Am J Physiol Regul Integr Comp Physiol, 2002. **282**(1): p. R295-302.
159. Tomanek, R.J., *Developmental Progression of the Coronary Vasculature in Human Embryos and Fetuses.* Anat Rec (Hoboken), 2016. **299**(1): p. 25-41.

160. Malek, A.M., et al., *Fluid shear stress differentially modulates expression of genes encoding basic fibroblast growth factor and platelet-derived growth factor B chain in vascular endothelium.* J Clin Invest, 1993. **92**(4): p. 2013-21.

161. Hood, L.C. and T.H. Rosenquist, *Coronary artery development in the chick: origin and deployment of smooth muscle cells, and the effects of neural crest ablation.* Anat Rec, 1992. **234**(2): p. 291-300.

162. Waldo, K.L., D.H. Kumiski, and M.L. Kirby, *Association of the cardiac neural crest with development of the coronary arteries in the chick embryo.* Anat Rec, 1994. **239**(3): p. 315-31.

163. del Monte, G., et al., *Differential Notch signaling in the epicardium is required for cardiac inflow development and coronary vessel morphogenesis.* Circ Res, 2011. **108**(7): p. 824-36.

164. Grieskamp, T., et al., *Notch signaling regulates smooth muscle differentiation of epicardium-derived cells.* Circ Res, 2011. **108**(7): p. 813-23.

165. de la Pompa, J.L. and J.A. Epstein, *Coordinating tissue interactions: Notch signaling in cardiac development and disease.* Dev Cell, 2012. **22**(2): p. 244-54.

166. Luxán, G., et al., *Mutations in the NOTCH pathway regulator MIB1 cause left ventricular noncompaction cardiomyopathy.* Nat Med, 2013. **19**(2): p. 193-201.

167. D'Amato, G., G. Luxán, and J.L. de la Pompa, *Notch signalling in ventricular chamber development and cardiomyopathy.* Febs j, 2016. **283**(23): p. 4223-4237.

168. D'Amato, G., et al., *Sequential Notch activation regulates ventricular chamber development.* Nat Cell Biol, 2016. **18**(1): p. 7-20.

169. van Weerd, J.H. and V.M. Christoffels, *The formation and function of the cardiac conduction system.* Development, 2016. **143**(2): p. 197-210.

170. DeHaan, R.L., *Development of pacemaker tissue in the embryonic heart.* Ann N Y Acad Sci, 1965. **127**(1): p. 7-18.

171. Moorman, A.F. and V.M. Christoffels, *Cardiac chamber formation: development, genes, and evolution.* Physiol Rev, 2003. **83**(4): p. 1223-67.

172. Virágh, S. and C.E. Challice, *The development of the conduction system in the mouse embryo heart.* Dev Biol, 1980. **80**(1): p. 28-45.

173. Hoogaars, W.M., et al., *Tbx3 controls the sinoatrial node gene program and imposes pacemaker function on the atria.* Genes Dev, 2007. **21**(9): p. 1098-112.

174. Vicente-Steijn, R., et al., *Electrical activation of sinus venosus myocardium and expression patterns of RhoA and Isl-1 in the chick embryo.* J Cardiovasc Electrophysiol, 2010. **21**(11): p. 1284-92.

175. Hoogaars, W.M., et al., *The transcriptional repressor Tbx3 delineates the developing central conduction system of the heart.* Cardiovasc Res, 2004. **62**(3): p. 489-99.

176. Frank, D.U., et al., *Lethal arrhythmias in Tbx3-deficient mice reveal extreme dosage sensitivity of cardiac conduction system function and homeostasis.* Proc Natl Acad Sci U S A, 2012. **109**(3): p. E154-63.

177. Aanhaanen, W.T., et al., *The Tbx2+ primary myocardium of the atrioventricular canal forms the atrioventricular node and the base of the left ventricle.* Circ Res, 2009. **104**(11): p. 1267-74.

178. Vicente-Steijn, R., et al., *Funny current channel HCN4 delineates the developing cardiac conduction system in chicken heart.* Heart Rhythm, 2011. **8**(8): p. 1254-63.

179. Singh, R., et al., *Tbx2 and Tbx3 induce atrioventricular myocardial development and endocardial cushion formation.* Cell Mol Life Sci, 2012. **69**(8): p. 1377-89.

180. Bressan, M., et al., *Reciprocal myocardial-endocardial interactions pattern the delay in atrioventricular junction conduction.* Development, 2014. **141**(21): p. 4149-57.

181. Wessels, A., et al., *The development of the atrioventricular junction in the human heart.* Circ Res, 1996. **78**(1): p. 110-7.

182. Zhou, B., et al., *Genetic fate mapping demonstrates contribution of epicardium-derived cells to the annulus fibrosis of the mammalian heart.* Dev Biol, 2010. **338**(2): p. 251-61.

183. Lockhart, M.M., et al., *Alk3 mediated Bmp signaling controls the contribution of epicardially derived cells to the tissues of the atrioventricular junction.* Dev Biol, 2014. **396**(1): p. 8-18.

184. Aanhaanen, W.T., et al., *Defective Tbx2-dependent patterning of the atrioventricular canal myocardium causes accessory pathway formation in mice.* J Clin Invest, 2011. **121**(2): p. 534-44.

185. Christoffels, V.M., et al., *T-box transcription factor Tbx2 represses differentiation and formation of the cardiac chambers.* Dev Dyn, 2004. **229**(4): p. 763-70.

186. Harrelson, Z., et al., *Tbx2 is essential for patterning the atrioventricular canal and for morphogenesis of the outflow tract during heart development.* Development, 2004. **131**(20): p. 5041-52.

187. Zhou, P., A. He, and W.T. Pu, *Regulation of GATA4 transcriptional activity in cardiovascular development and disease.* Curr Top Dev Biol, 2012. **100**: p. 143-69.

188. Stefanovic, S., et al., *GATA-dependent regulatory switches establish atrioventricular canal specificity during heart development.* Nat Commun, 2014. **5**: p. 3680.

189. Stefanovic, S. and V.M. Christoffels, *GATA-dependent transcriptional and epigenetic control of cardiac lineage specification and differentiation.* Cell Mol Life Sci, 2015. **72**(20): p. 3871-81.

190. Gourdie, R.G., et al., *Terminal diversification of the myocyte lineage generates Purkinje fibers of the cardiac conduction system.* Development, 1995. **121**(5): p. 1423-31.

191. Cheng, G., et al., *Development of the cardiac conduction system involves recruitment within a multipotent cardiomyogenic lineage.* Development, 1999. **126**(22): p. 5041-9.

192. Wessels, A., et al., *Spatial distribution of "tissue-specific" antigens in the developing human heart and skeletal muscle. III. An immunohistochemical analysis of the distribution of the neural tissue antigen G1N2 in the embryonic heart; implications for the development of the atrioventricular conduction system.* Anat Rec, 1992. **232**(1): p. 97-111.

193. Rentschler, S., et al., *Visualization and functional characterization of the developing murine cardiac conduction system.* Development, 2001. **128**(10): p. 1785-92.

194. Miquerol, L., et al., *Biphasic development of the mammalian ventricular conduction system.* Circ Res, 2010. **107**(1): p. 153-61.

195. Zhang, S.S., et al., *Iroquois homeobox gene 3 establishes fast conduction in the cardiac His-Purkinje network.* Proc Natl Acad Sci U S A, 2011. **108**(33): p. 13576-81.

196. Moskowitz, I.P., et al., *A molecular pathway including Id2, Tbx5, and Nkx2-5 required for cardiac conduction system development.* Cell, 2007. **129**(7): p. 1365-76.

197. Costantini, D.L., et al., *The homeodomain transcription factor Irx5 establishes the mouse cardiac ventricular repolarization gradient.* Cell, 2005. **123**(2): p. 347-58.

198. Finnemore, A. and A. Groves, *Physiology of the fetal and transitional circulation.* Semin Fetal Neonatal Med, 2015. **20**(4): p. 210-6.

199. Morton, S.U. and D. Brodsky, *Fetal Physiology and the Transition to Extrauterine Life.* Clin Perinatol, 2016. **43**(3): p. 395-407.

200. Chatfield, B.A., et al., *Hemodynamic effects of endothelin-1 on ovine fetal pulmonary circulation.* Am J Physiol, 1991. **261**(1 Pt 2): p. R182-7.

201. Paradis, A. and L. Zhang, *Role of endothelin in uteroplacental circulation and fetal vascular function.* Curr Vasc Pharmacol, 2013. **11**(5): p. 594-605.

202. S, B., *Maternal, Fetal & Neonatal Physiology E-Book*. Elsevir Health Sciences. 2014.
203. Nagasawa, H., et al., *Time to spontaneous ductus arteriosus closure in full-term neonates.* Open Heart, 2016. **3**(1): p. e000413.
204. Tibby, S.M., et al., *Clinicians' abilities to estimate cardiac index in ventilated children and infants.* Arch Dis Child, 1997. **77**(6): p. 516-8.
205. Baschat, A.A., *The fetal circulation and essential organs-a new twist to an old tale.* Ultrasound Obstet Gynecol, 2006. **27**(4): p. 349-54.
206. Anderson, R.H., *Teratogenecity in the setting of cardiac development and maldevelopment.* Embryology, 2016.
207. Blom, N.A., et al., *Deficiency of the vestibular spine in atrioventricular septal defects in human fetuses with down syndrome.* Am J Cardiol, 2003. **91**(2): p. 180-4.
208. Manner, J., W. Seidl, and G. Steding, *Embryological observations on the morphogenesis of double-outlet right ventricle with subaortic ventricular septal defect and normal arrangement of the great arteries.* Thorac Cardiovasc Surg, 1995. **43**(6): p. 307-12.
209. Jarrell, D.K., M.L. Lennon, and J.G. Jacot, *Epigenetics and Mechanobiology in Heart Development and Congenital Heart Disease.* Diseases, 2019. **7**(3).
210. Dietrich, A.C., et al., *Blood flow and Bmp signaling control endocardial chamber morphogenesis.* Dev Cell, 2014. **30**(4): p. 367-77.
211. Rasouli, S.J., et al., *The flow responsive transcription factor Klf2 is required for myocardial wall integrity by modulating Fgf signaling.* Elife, 2018. **7**.
212. Messerschmidt, V., et al., *Light-sheet Fluorescence Microscopy to Capture 4-Dimensional Images of the Effects of Modulating Shear Stress on the Developing Zebrafish Heart.* J Vis Exp, 2018(138).
213. Kennedy, M.P., et al., *Congenital heart disease and other heterotaxic defects in a large cohort of patients with primary ciliary dyskinesia.* Circulation, 2007. **115**(22): p. 2814-21.
214. Dickinson, M.E., et al., *High-throughput discovery of novel developmental phenotypes.* Nature, 2016. **537**(7621): p. 508-514.
215. Chen, C.M., et al., *Detecting cardiac contractile activity in the early mouse embryo using multiple modalities.* Front Physiol, 2014. **5**: p. 508.
216. Conway, S.J., et al., *What cardiovascular defect does my prenatal mouse mutant have, and why?* Genesis, 2003. **35**(1): p. 1-21.
217. Savolainen, S.M., J.F. Foley, and S.A. Elmore, *Histology atlas of the developing mouse heart with emphasis on E11.5 to E18.5.* Toxicol Pathol, 2009. **37**(4): p. 395-414.
218. Krishnan, A., et al., *A detailed comparison of mouse and human cardiac development.* Pediatr Res, 2014. **76**(6): p. 500-7.
219. Captur, G., et al., *Morphogenesis of myocardial trabeculae in the mouse embryo.* J Anat, 2016. **229**(2): p. 314-25.
220. Shen, Y., et al., *Cardiovascular phenotyping of fetal mice by noninvasive high-frequency ultrasound facilitates recovery of ENU-induced mutations causing congenital cardiac and extracardiac defects.* Physiol Genomics, 2005. **24**(1): p. 23-36.
221. Brunetti-Pierri, N., et al., *Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities.* Nat Genet, 2008. **40**(12): p. 1466-71.
222. Mohun, T.J. and W.J. Weninger, *Episcopic three-dimensional imaging of embryos.* Cold Spring Harb Protoc, 2012. **2012**(6): p. 641-6.
223. Weninger, W.J., et al., *High-resolution episcopic microscopy: a rapid technique for high detailed 3D analysis of gene activity in the context of tissue architecture and morphology.* Anat Embryol (Berl), 2006. **211**(3): p. 213-21.

224. Geyer, S.H., et al., *Morphology, topology and dimensions of the heart and arteries of genetically normal and mutant mouse embryos at stages S21-S23.* J Anat, 2017. **231**(4): p. 600-614.

225. Andersen, T.A., L. Troelsen Kde, and L.A. Larsen, *Of mice and men: molecular genetics of congenital heart disease.* Cell Mol Life Sci, 2014. **71**(8): p. 1327-52.

226. Nora, J.J. and A.H. Nora, *Update on counseling the family with a first-degree relative with a congenital heart defect.* Am J Med Genet, 1988. **29**(1): p. 137-42.

227. Hanna, E.J., N.C. Nevin, and J. Nelson, *Genetic study of congenital heart defects in Northern Ireland (1974-1978).* J Med Genet, 1994. **31**(11): p. 858-63.

228. Calcagni, G., et al., *Familial recurrence of congenital heart disease: an overview and review of the literature.* Eur J Pediatr, 2007. **166**(2): p. 111-6.

229. Brodwall, K., et al., *Recurrence of congenital heart defects among siblings-a nationwide study.* Am J Med Genet A, 2017. **173**(6): p. 1575-1585.

230. Loffredo, C.A., et al., *Prevalence of congenital cardiovascular malformations among relatives of infants with hypoplastic left heart, coarctation of the aorta, and d-transposition of the great arteries.* Am J Med Genet A, 2004. **124A**(3): p. 225-30.

231. Egbe, A., et al., *Incidences and sociodemographics of specific congenital heart diseases in the United States of America: an evaluation of hospital discharge diagnoses.* Pediatr Cardiol, 2014. **35**(6): p. 975-82.

232. Fixler, D.E., et al., *Trends in congenital heart disease in Dallas County births. 1971-1984.* Circulation, 1990. **81**(1): p. 137-42.

233. Burn, J., et al., *Recurrence risks in offspring of adults with major heart defects: results from first cohort of British collaborative study.* Lancet, 1998. **351**(9099): p. 311-6.

234. Oyen, N., et al., *Recurrence of discordant congenital heart defects in families.* Circ Cardiovasc Genet, 2010. **3**(2): p. 122-8.

235. Nees, S.N. and W.K. Chung, *Genetic Basis of Human Congenital Heart Disease.* Cold Spring Harb Perspect Biol, 2019.

236. Øyen, N., et al., *Prepregnancy Diabetes and Offspring Risk of Congenital Heart Disease: A Nationwide Cohort Study.* Circulation, 2016. **133**(23): p. 2243-53.

237. Lisowski, L.A., et al., *Congenital heart disease in pregnancies complicated by maternal diabetes mellitus. An international clinical collaboration, literature review, and meta-analysis.* Herz, 2010. **35**(1): p. 19-26.

238. Wilson, P.D., et al., *Attributable fraction for cardiac malformations.* Am J Epidemiol, 1998. **148**(5): p. 414-23.

239. Thienpont, B., et al., *Submicroscopic chromosomal imbalances detected by array-CGH are a frequent cause of congenital heart defects in selected patients.* Eur Heart J, 2007. **28**(22): p. 2778-84.

240. Sheng, W., et al., *Association of promoter methylation statuses of congenital heart defect candidate genes with Tetralogy of Fallot.* J Transl Med, 2014. **12**: p. 31.

241. Gelb, B.D., *Molecular genetics of congenital heart disease.* Curr Opin Cardiol, 1997. **12**(3): p. 321-8.

242. Cowan, J.R., et al., *Copy number variation as a genetic basis for heterotaxy and heterotaxy-spectrum congenital heart defects.* Philos Trans R Soc Lond B Biol Sci, 2016. **371**(1710).

243. Patel, A., et al., *Prevalence of Noncardiac and Genetic Abnormalities in Neonates Undergoing Cardiac Operations: Analysis of The Society of Thoracic Surgeons Congenital Heart Surgery Database.* Ann Thorac Surg, 2016. **102**(5): p. 1607-1614.

244. Zaidi, S. and M. Brueckner, *Genetics and Genomics of Congenital Heart Disease.* Circ Res, 2017. **120**(6): p. 923-940.

245. Goldenberg, P.C., et al., *High burden of genetic conditions diagnosed in a cardiac neurodevelopmental clinic.* Cardiol Young, 2016: p. 1-8.

246. Ahrens-Nicklas, R.C., et al., *Utility of genetic evaluation in infants with congenital heart defects admitted to the cardiac intensive care unit.* Am J Med Genet A, 2016.

247. Washington Smoak, I., et al., *Sonic hedgehog is required for cardiac outflow tract and neural crest cell development.* Dev Biol, 2005. **283**(2): p. 357-72.

248. Zhang, Z., et al., *Massively parallel sequencing identifies the gene Megf8 with ENU-induced mutation causing heterotaxy.* Proc Natl Acad Sci U S A, 2009. **106**(9): p. 3219-24.

249. Liu, C., et al., *Rare copy number variants analysis identifies novel candidate genes in heterotaxy syndrome patients with congenital heart defects.* Genome Med, 2018. **10**(1): p. 40.

250. Breckpot, J., et al., *Array comparative genomic hybridization as a diagnostic tool for syndromic heart defects.* J Pediatr, 2010. **156**(5): p. 810-7, 817 e1-817 e4.

251. Baker, K., et al., *Critical congenital heart disease--utility of routine screening for chromosomal and other extracardiac malformations.* Congenit Heart Dis, 2012. **7**(2): p. 145-50.

252. Al Turki, S., et al., *Rare variants in NR2F2 cause congenital heart defects in humans.* Am J Hum Genet, 2014. **94**(4): p. 574-85.

253. Connor, J.A., et al., *Genetic testing practices in infants with congenital heart disease.* Congenit Heart Dis, 2014. **9**(2): p. 158-67.

254. Retterer, K., et al., *Clinical application of whole-exome sequencing across clinical indications.* Genet Med, 2016. **18**(7): p. 696-704.

255. Allen HD, D.D., Shaddy RE, Feltes TF, *Moss and Adams' heart disease in infants, children, and adolescents: including the fetus and young adult,*. 8th ed. 2013, Philadelphia: Wolters

Kluwer.

256. Bull, M.J., *Health supervision for children with Down syndrome.* Pediatrics, 2011. **128**(2): p. 393-406.

257. de Graaf, G., et al., *Estimation of live birth and population prevalence of Down syndrome in nine U.S. states.* Am J Med Genet A, 2017. **173**(10): p. 2710-2719.

258. Van Praagh, S., et al., *Cardiac malformations in trisomy-18: a study of 41 postmortem cases.* J Am Coll Cardiol, 1989. **13**(7): p. 1586-97.

259. Musewe, N.N., et al., *Echocardiographic evaluation of the spectrum of cardiac anomalies associated with trisomy 13 and trisomy 18.* J Am Coll Cardiol, 1990. **15**(3): p. 673-7.

260. Embleton, N.D., et al., *Natural history of trisomy 18.* Arch Dis Child Fetal Neonatal Ed, 1996. **75**(1): p. F38-41.

261. Springett, A., et al., *Congenital anomalies associated with trisomy 18 or trisomy 13: A registry-based study in 16 European countries, 2000-2011.* Am J Med Genet A, 2015. **167a**(12): p. 3062-9.

262. Wyllie, J.P., et al., *Natural history of trisomy 13.* Arch Dis Child, 1994. **71**(4): p. 343-5.

263. Lin, H.Y., et al., *Clinical characteristics and survival of trisomy 13 in a medical center in Taiwan, 1985-2004.* Pediatr Int, 2007. **49**(3): p. 380-6.

264. Goldstein, H. and K.G. Nielsen, *Rates and survival of individuals with trisomy 13 and 18. Data from a 10-year period in Denmark.* Clin Genet, 1988. **34**(6): p. 366-72.

265. Carlson, M., et al., *Moderate aortic enlargement and bicuspid aortic valve are associated with aortic dissection in Turner syndrome: report of the international turner syndrome aortic dissection registry.* Circulation, 2012. **126**(18): p. 2220-6.

266. Gravholt, C.H., et al., *Clinical practice guidelines for the care of girls and women with Turner syndrome: proceedings from the 2016 Cincinnati International Turner Syndrome Meeting.* Eur J Endocrinol, 2017. **177**(3): p. G1-g70.

267. Sybert, V.P. and E. McCauley, *Turner's syndrome.* N Engl J Med, 2004. **351**(12): p. 1227-38.

268. Mercer-Rosa, L., et al., *22q11.2 deletion status and disease burden in children and adolescents with tetralogy of Fallot.* Circ Cardiovasc Genet, 2015. **8**(1): p. 74-81.

269. van Engelen, K., et al., *22q11.2 Deletion Syndrome is under-recognised in adult patients with tetralogy of Fallot and pulmonary atresia.* Heart, 2010. **96**(8): p. 621-4.

270. Botto, L.D., et al., *A population-based study of the 22q11.2 deletion: phenotype, incidence, and contribution to major birth defects in the population.* Pediatrics, 2003. **112**(1 Pt 1): p. 101-7.

271. Digilio, M.C., et al., *Spectrum of clinical variability in familial deletion 22q11.2: from full manifestation to extremely mild clinical anomalies.* Clin Genet, 2003. **63**(4): p. 308-13.

272. Kececioglu, D., S. Kotthoff, and J. Vogt, *Williams-Beuren syndrome: a 30-year follow-up of natural and postoperative course.* Eur Heart J, 1993. **14**(11): p. 1458-64.

273. Morris, C.A. and C.B. Mervis, *Williams syndrome and related disorders.* Annu Rev Genomics Hum Genet, 2000. **1**: p. 461-84.

274. Smith AC, G.A., *Smith-Magenis syndrome*. 3rd ed. Management of Genetic Syndromes, ed. C.S.a.A. J. 2010, New York NY: Wiley-Blackwell.

275. Battaglia, A., T. Filippi, and J.C. Carey, *Update on the clinical features and natural history of Wolf-Hirschhorn (4p-) syndrome: experience with 87 patients and recommendations for routine health supervision.* Am J Med Genet C Semin Med Genet, 2008. **148c**(4): p. 246-51.

276. Silversides, C.K., et al., *Rare copy number variations in adults with tetralogy of Fallot implicate novel risk gene pathways.* PLoS Genet, 2012. **8**(8): p. e1002843.

277. Soemedi, R., et al., *Phenotype-specific effect of chromosome 1q21.1 rearrangements and GJA5 duplications in 2436 congenital heart disease patients and 6760 controls.* Hum Mol Genet, 2012. **21**(7): p. 1513-20.

278. Battaglia, A., et al., *Further delineation of deletion 1p36 syndrome in 60 patients: a recognizable phenotype and common cause of developmental delay and mental retardation.* Pediatrics, 2008. **121**(2): p. 404-10.

279. Rosenfeld, J.A., et al., *New cases and refinement of the critical region in the 1q41q42 microdeletion syndrome.* Eur J Med Genet, 2011. **54**(1): p. 42-9.

280. Wat, M.J., et al., *Chromosome 8p23.1 deletions as a cause of complex congenital heart defects and diaphragmatic hernia.* Am J Med Genet A, 2009. **149a**(8): p. 1661-77.

281. Reamon-Buettner, S.M. and J. Borlak, *NKX2-5: an update on this hypermutable homeodomain protein and its role in human congenital heart disease (CHD).* Hum Mutat, 2010. **31**(11): p. 1185-94.

282. Harvey, R.P., et al., *Homeodomain factor Nkx2-5 in heart development and disease.* Cold Spring Harb Symp Quant Biol, 2002. **67**: p. 107-14.

283. Stallmeyer, B., et al., *Mutational spectrum in the cardiac transcription factor gene NKX2.5 (CSX) associated with congenital heart disease.* Clin Genet, 2010. **78**(6): p. 533-40.

284. Perera, J.L., et al., *Novel and highly lethal NKX2.5 missense mutation in a family with sudden death and ventricular arrhythmia.* Pediatr Cardiol, 2014. **35**(7): p. 1206-12.

285. Zakariyah, A.F., et al., *In Vitro Modeling of Congenital Heart Defects Associated with an NKX2-5 Mutation Revealed a Dysregulation in BMP/Notch-Mediated Signaling.* Stem Cells, 2018. **36**(4): p. 514-526.

286. Sletten, L.J. and M.E. Pierpont, *Variation in severity of cardiac disease in Holt-Oram syndrome.* Am J Med Genet, 1996. **65**(2): p. 128-32.

287. Baban, A., et al., *Holt-Oram syndrome with intermediate atrioventricular canal defect, and aortic coarctation: functional characterization of a de novo TBX5 mutation.* Am J Med Genet A, 2014. **164a**(6): p. 1419-24.

288. Barisic, I., et al., *Holt Oram syndrome: a registry-based study in Europe.* Orphanet J Rare Dis, 2014. **9**: p. 156.

289. Basson, C.T., et al., *The clinical and genetic spectrum of the Holt-Oram syndrome (heart-hand syndrome).* N Engl J Med, 1994. **330**(13): p. 885-91.

290. McDermott DA, F.J., Basson CT, *Holt–Oram syndrome*. 1993, Seattle: University of Washington.

291. Holt, M. and S. Oram, *Familial heart disease with skeletal malformations.* Br Heart J, 1960. **22**(2): p. 236-42.

292. Basson, C.T., et al., *Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome.* Nat Genet, 1997. **15**(1): p. 30-5.

293. Böhm, J., et al., *Functional analysis of the novel TBX5 c.1333delC mutation resulting in an extended TBX5 protein.* BMC Med Genet, 2008. **9**: p. 88.

294. Fan, C., M. Liu, and Q. Wang, *Functional analysis of TBX5 missense mutations associated with Holt-Oram syndrome.* J Biol Chem, 2003. **278**(10): p. 8780-5.

295. Muru, K., et al., *A Boy with Holt-Oram Syndrome Caused by Novel Mutation c.1304delT in the TBX5 Gene.* Mol Syndromol, 2011. **1**(6): p. 307-310.

296. Garg, V., et al., *GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5.* Nature, 2003. **424**(6947): p. 443-7.

297. Zhang, Y., et al., *Associations of GATA4 genetic mutations with the risk of congenital heart disease: A meta-analysis.* Medicine (Baltimore), 2017. **96**(18): p. e6857.

298. Yang, B., et al., *Protein-altering and regulatory genetic variants near GATA4 implicated in bicuspid aortic valve.* Nat Commun, 2017. **8**: p. 15481.

299. Zhao, R., et al., *Loss of both GATA4 and GATA6 blocks cardiac myocyte differentiation and results in acardia in mice.* Dev Biol, 2008. **317**(2): p. 614-9.

300. Shan, J.P., et al., *Novel and functional DNA sequence variants within the GATA5 gene promoter in ventricular septal defects.* World J Pediatr, 2014. **10**(4): p. 348-53.

301. Jiang, J.Q., et al., *Prevalence and spectrum of GATA5 mutations associated with congenital heart disease.* Int J Cardiol, 2013. **165**(3): p. 570-3.

302. Lin, X., et al., *A novel GATA6 mutation in patients with tetralogy of Fallot or atrial septal defect.* J Hum Genet, 2010. **55**(10): p. 662-7.

303. Kodo, K., et al., *GATA6 mutations cause human cardiac outflow tract defects by disrupting semaphorin-plexin signaling.* Proc Natl Acad Sci U S A, 2009. **106**(33): p. 13933-8.

304. Xu, Y.J., et al., *GATA6 loss-of-function mutation contributes to congenital bicuspid aortic valve.* Gene, 2018. **663**: p. 115-120.

305. Yagi, H., et al., *Role of TBX1 in human del22q11.2 syndrome.* Lancet, 2003. **362**(9393): p. 1366-73.

306. Rauch, R., et al., *Comprehensive genotype-phenotype analysis in 230 patients with tetralogy of Fallot.* J Med Genet, 2010. **47**(5): p. 321-31.

307. Kirk, E.P., et al., *Mutations in cardiac T-box factor gene TBX20 are associated with diverse cardiac pathologies, including defects of septation and valvulogenesis and cardiomyopathy.* Am J Hum Genet, 2007. **81**(2): p. 280-91.

308. Posch, M.G., et al., *A gain-of-function TBX20 mutation causes congenital atrial septal defects, patent foramen ovale and cardiac valve defects.* J Med Genet, 2010. **47**(4): p. 230-5.

309. Reamon-Buettner, S.M., et al., *A loss-of-function mutation in the binding domain of HAND1 predicts hypoplasia of the human hearts.* Hum Mol Genet, 2008. **17**(10): p. 1397-405.

310. Reamon-Buettner, S.M., et al., *A functional genetic study identifies HAND1 mutations in septation defects of the human heart.* Hum Mol Genet, 2009. **18**(19): p. 3567-78.

311. Shen, L., et al., *Transcription factor HAND2 mutations in sporadic Chinese patients with congenital heart disease.* Chin Med J (Engl), 2010. **123**(13): p. 1623-7.
312. Sun, Y.M., et al., *A HAND2 Loss-of-Function Mutation Causes Familial Ventricular Septal Defect and Pulmonary Stenosis.* G3 (Bethesda), 2016. **6**(4): p. 987-92.
313. Topf, A., et al., *Functionally significant, rare transcription factor variants in tetralogy of Fallot.* PLoS One, 2014. **9**(8): p. e95453.
314. Zhu, X., et al., *A novel mutation of Hyaluronan synthase 2 gene in Chinese children with ventricular septal defect.* PLoS One, 2014. **9**(2): p. e87437.
315. Granados-Riveron, J.T., et al., *Alpha-cardiac myosin heavy chain (MYH6) mutations affecting myofibril formation are associated with congenital heart defects.* Hum Mol Genet, 2010. **19**(20): p. 4007-16.
316. Theis, J.L., et al., *Recessive MYH6 Mutations in Hypoplastic Left Heart With Reduced Ejection Fraction.* Circ Cardiovasc Genet, 2015. **8**(4): p. 564-71.
317. Postma, A.V., et al., *Mutations in the sarcomere gene MYH7 in Ebstein anomaly.* Circ Cardiovasc Genet, 2011. **4**(1): p. 43-50.
318. Wells, Q.S., et al., *Familial dilated cardiomyopathy associated with congenital defects in the setting of a novel VCL mutation (Lys815Arg) in conjunction with a known MYPBC3 variant.* Cardiogenetics, 2011. **1**(1).
319. Wessels, M.W., et al., *Compound heterozygous or homozygous truncating MYBPC3 mutations cause lethal cardiomyopathy with features of noncompaction and septal defects.* Eur J Hum Genet, 2015. **23**(7): p. 922-8.
320. Lederer, D., et al., *Deletion of KDM6A, a histone demethylase interacting with MLL2, in three patients with Kabuki syndrome.* Am J Hum Genet, 2012. **90**(1): p. 119-24.
321. Digilio, M.C., et al., *Congenital heart defects in Kabuki syndrome.* Am J Med Genet, 2001. **100**(4): p. 269-74.
322. Faralli, H., et al., *UTX demethylase activity is required for satellite cell-mediated muscle regeneration.* J Clin Invest, 2016. **126**(4): p. 1555-65.
323. Cyran, S.E., et al., *Spectrum of congenital heart disease in CHARGE association.* J Pediatr, 1987. **110**(4): p. 576-8.
324. Vissers, L.E., et al., *Mutations in a new member of the chromodomain gene family cause CHARGE syndrome.* Nat Genet, 2004. **36**(9): p. 955-7.
325. Tatton-Brown, K., et al., *Genotype-phenotype associations in Sotos syndrome: an analysis of 266 individuals with NSD1 aberrations.* Am J Hum Genet, 2005. **77**(2): p. 193-204.
326. Leventopoulos, G., et al., *A clinical study of Sotos syndrome patients with review of the literature.* Pediatr Neurol, 2009. **40**(5): p. 357-64.
327. Huang, R.T., et al., *CASZ1 loss-of-function mutation associated with congenital heart disease.* Gene, 2016. **595**(1): p. 62-68.
328. Xiao, D., et al., *The roles of SMYD4 in epigenetic regulation of cardiac development in zebrafish.* PLoS Genet, 2018. **14**(8): p. e1007578.
329. Tartaglia, M., et al., *Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome.* Nat Genet, 2001. **29**(4): p. 465-8.
330. Tartaglia, M., et al., *PTPN11 mutations in Noonan syndrome: molecular spectrum, genotype-phenotype correlation, and phenotypic heterogeneity.* Am J Hum Genet, 2002. **70**(6): p. 1555-63.
331. Roberts, A.E., et al., *Germline gain-of-function mutations in SOS1 cause Noonan syndrome.* Nat Genet, 2007. **39**(1): p. 70-4.
332. Kouz, K., et al., *Genotype and phenotype in patients with Noonan syndrome and a RIT1 mutation.* Genet Med, 2016. **18**(12): p. 1226-1234.
333. Aoki, Y., et al., *Recent advances in RASopathies.* J Hum Genet, 2016. **61**(1): p. 33-9.
334. Carta, C., et al., *Germline missense mutations affecting KRAS Isoform B are associated with a severe Noonan syndrome phenotype.* Am J Hum Genet, 2006. **79**(1): p. 129-35.

335. Cordeddu, V., et al., *Mutation of SHOC2 promotes aberrant protein N-myristoylation and causes Noonan-like syndrome with loose anagen hair.* Nat Genet, 2009. **41**(9): p. 1022-6.
336. El Bouchikhi, I., et al., *Noonan syndrome-causing genes: Molecular update and an assessment of the mutation rate.* Int J Pediatr Adolesc Med, 2016. **3**(4): p. 133-142.
337. Pandit, B., et al., *Gain-of-function RAF1 mutations cause Noonan and LEOPARD syndromes with hypertrophic cardiomyopathy.* Nat Genet, 2007. **39**(8): p. 1007-12.
338. Marino, B., et al., *Congenital heart diseases in children with Noonan syndrome: An expanded cardiac spectrum with high prevalence of atrioventricular canal.* J Pediatr, 1999. **135**(6): p. 703-6.
339. Jhang, W.K., et al., *Cardiac Manifestations and Associations with Gene Mutations in Patients Diagnosed with RASopathies.* Pediatr Cardiol, 2016. **37**(8): p. 1539-1547.
340. Pierpont, M.E., et al., *Cardio-facio-cutaneous syndrome: clinical features, diagnosis, and management guidelines.* Pediatrics, 2014. **134**(4): p. e1149-62.
341. MacGrogan, D., J. Münch, and J.L. de la Pompa, *Notch and interacting signalling pathways in cardiac development, disease, and regeneration.* Nat Rev Cardiol, 2018. **15**(11): p. 685-704.
342. Timmerman, L.A., et al., *Notch promotes epithelial-mesenchymal transition during cardiac development and oncogenic transformation.* Genes Dev, 2004. **18**(1): p. 99-115.
343. Garg, V., et al., *Mutations in NOTCH1 cause aortic valve disease.* Nature, 2005. **437**(7056): p. 270-4.
344. Preuss, C., et al., *Family Based Whole Exome Sequencing Reveals the Multifaceted Role of Notch Signaling in Congenital Heart Disease.* PLoS Genet, 2016. **12**(10): p. e1006335.
345. Li, B., et al., *MIB1 mutations reduce Notch signaling activation and contribute to congenital heart disease.* Clin Sci (Lond), 2018. **132**(23): p. 2483-2491.
346. Warthen, D.M., et al., *Jagged1 (JAG1) mutations in Alagille syndrome: increasing the mutation detection rate.* Hum Mutat, 2006. **27**(5): p. 436-43.
347. Spinner NB, L.L., Krantz ID, *Alagille Syndrome*, in *Genereviews*. 1993, University of Washington Seattle.
348. McDaniell, R., et al., *NOTCH2 mutations cause Alagille syndrome, a heterogeneous disorder of the notch signaling pathway.* Am J Hum Genet, 2006. **79**(1): p. 169-73.
349. Li, L., et al., *Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1.* Nat Genet, 1997. **16**(3): p. 243-51.
350. Oda, T., et al., *Mutations in the human Jagged1 gene are responsible for Alagille syndrome.* Nat Genet, 1997. **16**(3): p. 235-42.
351. Emerick, K.M., et al., *Features of Alagille syndrome in 92 patients: frequency and relation to prognosis.* Hepatology, 1999. **29**(3): p. 822-9.
352. McElhinney, D.B., et al., *Analysis of cardiovascular phenotype and genotype-phenotype correlation in individuals with a JAG1 mutation and/or Alagille syndrome.* Circulation, 2002. **106**(20): p. 2567-74.
353. Priest, J.R., et al., *De Novo and Rare Variants at Multiple Loci Support the Oligogenic Origins of Atrioventricular Septal Heart Defects.* PLoS Genet, 2016. **12**(4): p. e1005963.
354. Neeb, Z., et al., *Cardiac outflow tract anomalies.* Wiley Interdiscip Rev Dev Biol, 2013. **2**(4): p. 499-530.
355. MacGrogan, D., G. Luxán, and J.L. de la Pompa, *Genetic and functional genomics approaches targeting the Notch pathway in cardiac development and congenital heart disease.* Brief Funct Genomics, 2014. **13**(1): p. 15-27.
356. Xie, H.M., et al., *Rare copy number variants in patients with congenital conotruncal heart defects.* Birth Defects Res, 2017. **109**(4): p. 271-295.

357. Reamon-Buettner, S.M. and J. Borlak, *HEY2 mutations in malformed hearts.* Hum Mutat, 2006. **27**(1): p. 118.

358. Rochais, F., K. Mesbah, and R.G. Kelly, *Signaling pathways controlling second heart field development.* Circ Res, 2009. **104**(8): p. 933-42.

359. Merks, A.M., et al., *Planar cell polarity signalling coordinates heart tube remodelling through tissue-scale polarisation of actomyosin activity.* Nat Commun, 2018. **9**(1): p. 2161.

360. Caron, A., X. Xu, and X. Lin, *Wnt/β-catenin signaling directly regulates Foxj1 expression and ciliogenesis in zebrafish Kupffer's vesicle.* Development, 2012. **139**(3): p. 514-24.

361. May-Simera, H.L. and M.W. Kelley, *Cilia, Wnt signaling, and the cytoskeleton.* Cilia, 2012. **1**(1): p. 7.

362. Ye, B., et al., *APC controls asymmetric Wnt/β-catenin signaling and cardiomyocyte proliferation gradient in the heart.* J Mol Cell Cardiol, 2015. **89**(Pt B): p. 287-96.

363. Cantù, C., et al., *Mutations in Bcl9 and Pygo genes cause congenital heart defects by tissue-specific perturbation of Wnt/β-catenin signaling.* Genes Dev, 2018. **32**(21-22): p. 1443-1458.

364. Phillips, M.D., et al., *Dkk1 and Dkk2 regulate epicardial specification during mouse heart development.* Int J Cardiol, 2011. **150**(2): p. 186-92.

365. Bonachea, E.M., et al., *Use of a targeted, combinatorial next-generation sequencing approach for the study of bicuspid aortic valve.* BMC Med Genomics, 2014. **7**: p. 56.

366. Durst, R., et al., *Mutations in DCHS1 cause mitral valve prolapse.* Nature, 2015. **525**(7567): p. 109-13.

367. Grunert, M., et al., *Rare and private variations in neural crest, apoptosis and sarcomere genes define the polygenic background of isolated Tetralogy of Fallot.* Hum Mol Genet, 2014. **23**(12): p. 3115-28.

368. Qian, B., et al., *Common variations in BMP4 confer genetic susceptibility to sporadic congenital heart disease in a Han Chinese population.* Pediatr Cardiol, 2014. **35**(8): p. 1442-7.

369. Tan, H.L., et al., *Nonsynonymous variants in the SMAD6 gene predispose to congenital cardiovascular malformation.* Hum Mutat, 2012. **33**(4): p. 720-7.

370. Granadillo, J.L., et al., *Variable cardiovascular phenotypes associated with SMAD2 pathogenic variants.* Hum Mutat, 2018. **39**(12): p. 1875-1884.

371. Tian, E., et al., *Galnt1 is required for normal heart valve development and cardiac function.* PLoS One, 2015. **10**(1): p. e0115861.

372. Roessler, E., et al., *Reduced NODAL signaling strength via mutation of several pathway members including FOXH1 is linked to human heart defects and holoprosencephaly.* Am J Hum Genet, 2008. **83**(1): p. 18-29.

373. Mohapatra, B., et al., *Identification and functional characterization of NODAL rare variants in heterotaxy and isolated cardiovascular malformations.* Hum Mol Genet, 2009. **18**(5): p. 861-71.

374. Warburton, D., et al., *The contribution of de novo and rare inherited copy number changes to congenital heart disease in an unselected sample of children with conotruncal defects or hypoplastic left heart disease.* Hum Genet, 2014. **133**(1): p. 11-27.

375. Zhao, C.M., et al., *PITX2 Loss-of-Function Mutation Contributes to Congenital Endocardial Cushion Defect and Axenfeld-Rieger Syndrome.* PLoS One, 2015. **10**(4): p. e0124409.

376. Li, X., et al., *Heterogeneity Analysis and Diagnosis of Complex Diseases Based on Deep Learning Method.* Sci Rep, 2018. **8**(1): p. 6155.

377. Ackerman, C., et al., *An excess of deleterious variants in VEGF-A pathway genes in Down-syndrome-associated atrioventricular septal defects.* Am J Hum Genet, 2012. **91**(4): p. 646-59.

378. Reuter, M.S., et al., *Haploinsufficiency of vascular endothelial growth factor related signaling genes is associated with tetralogy of Fallot.* Genet Med, 2019. **21**(4): p. 1001-1007.

379. Pierpont, M.E., et al., *Genetic Basis for Congenital Heart Disease: Revisited: A Scientific Statement From the American Heart Association.* Circulation, 2018. **138**(21): p. e653-e711.

380. Williams, K., J. Carson, and C. Lo, *Genetics of Congenital Heart Disease.* Biomolecules, 2019. **9**(12).

381. Hartman, R.J., et al., *The contribution of chromosomal abnormalities to congenital heart defects: a population-based study.* Pediatr Cardiol, 2011. **32**(8): p. 1147-57.

382. Ferencz, C., et al., *Congenital cardiovascular malformations: questions on inheritance. Baltimore-Washington Infant Study Group.* J Am Coll Cardiol, 1989. **14**(3): p. 756-63.

383. Bruns, D., *Presenting physical characteristics, medical conditions, and developmental status of long-term survivors with trisomy 9 mosaicism.* Am J Med Genet A, 2011. **155a**(5): p. 1033-9.

384. Schinzel, A., *A Catalogue of Unbalanced Chromosome Aberrations in Man.* Chromosome Res, 2002. **10**: p. 1-4.

385. Redon, R., et al., *Global variation in copy number in the human genome.* Nature, 2006. **444**(7118): p. 444-54.

386. Goodship, J., et al., *Monozygotic twins with chromosome 22q11 deletion and discordant phenotype.* J Med Genet, 1995. **32**(9): p. 746-8.

387. Hitz, M.P., et al., *Rare copy number variants contribute to congenital left-sided heart disease.* PLoS Genet, 2012. **8**(9): p. e1002903.

388. Costain, G., C.K. Silversides, and A.S. Bassett, *The importance of copy number variation in congenital heart disease.* NPJ Genom Med, 2016. **1**: p. 16031.

389. Breckpot, J., et al., *Challenges of interpreting copy number variation in syndromic and non-syndromic congenital heart defects.* Cytogenet Genome Res, 2011. **135**(3-4): p. 251-9.

390. Lalani, S.R., et al., *Rare DNA copy number variants in cardiovascular malformations with extracardiac abnormalities.* Eur J Hum Genet, 2013. **21**(2): p. 173-81.

391. Franco, D., D. Sedmera, and E. Lozano-Velasco, *Multiple Roles of Pitx2 in Cardiac Development and Disease.* J Cardiovasc Dev Dis, 2017. **4**(4).

392. Simmons, M.A. and M. Brueckner, *The genetics of congenital heart disease… understanding and improving long-term outcomes in congenital heart disease: a review for the general cardiologist and primary care physician.* Curr Opin Pediatr, 2017. **29**(5): p. 520-528.

393. Bruneau, B.G., *Chromatin remodeling in heart development.* Curr Opin Genet Dev, 2010. **20**(5): p. 505-11.

394. Bird, A., *DNA methylation patterns and epigenetic memory.* Genes Dev, 2002. **16**(1): p. 6-21.

395. Chamberlain, A.A., et al., *DNA methylation is developmentally regulated for genes essential for cardiogenesis.* J Am Heart Assoc, 2014. **3**(3): p. e000976.

396. Sheng, W., et al., *DNA methylation status of NKX2-5, GATA4 and HAND1 in patients with tetralogy of fallot.* BMC Med Genomics, 2013. **6**: p. 46.

397. Lickert, H., et al., *Baf60c is essential for function of BAF chromatin remodelling complexes in heart development.* Nature, 2004. **432**(7013): p. 107-12.

398. Takeuchi, J.K. and B.G. Bruneau, *Directed transdifferentiation of mouse mesoderm to heart tissue by defined factors.* Nature, 2009. **459**(7247): p. 708-11.

399. Takeuchi, J.K., et al., *Chromatin remodelling complex dosage modulates transcription factor function in heart development.* Nat Commun, 2011. **2**: p. 187.

400. Kakita, T., et al., *p300 protein as a coactivator of GATA-5 in the transcription of cardiac-restricted atrial natriuretic factor gene.* J Biol Chem, 1999. **274**(48): p. 34096-102.
401. Yao, T.P., et al., *Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300.* Cell, 1998. **93**(3): p. 361-72.
402. Hata, A., *Functions of microRNAs in cardiovascular biology and disease.* Annu Rev Physiol, 2013. **75**: p. 69-93.
403. Schlesinger, J., et al., *The cardiac transcription network modulated by Gata4, Mef2a, Nkx2.5, Srf, histone modifications, and microRNAs.* PLoS Genet, 2011. **7**(2): p. e1001313.
404. Liu, N., et al., *An intragenic MEF2-dependent enhancer directs muscle-specific expression of microRNAs 1 and 133.* Proc Natl Acad Sci U S A, 2007. **104**(52): p. 20844-9.
405. Zhao, Y., E. Samal, and D. Srivastava, *Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis.* Nature, 2005. **436**(7048): p. 214-20.
406. Zhao, Y., et al., *Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1-2.* Cell, 2007. **129**(2): p. 303-17.
407. Ohanian, M., et al., *A heterozygous variant in the human cardiac miR-133 gene, MIR133A2, alters miRNA duplex processing and strand abundance.* BMC Genet, 2013. **14**: p. 18.
408. Oliveira-Carvalho, V., V.O. Carvalho, and E.A. Bocchi, *The emerging role of miR-208a in the heart.* DNA Cell Biol, 2013. **32**(1): p. 8-12.
409. O'Brien, J.E., Jr., et al., *Noncoding RNA expression in myocardium from infants with tetralogy of Fallot.* Circ Cardiovasc Genet, 2012. **5**(3): p. 279-86.
410. Hsieh, A., et al., *Early post-zygotic mutations contribute to congenital heart disease.* bioRxiv, 2019: p. 733105.
411. Gollob, M.H., et al., *Somatic mutations in the connexin 40 gene (GJA5) in atrial fibrillation.* N Engl J Med, 2006. **354**(25): p. 2677-88.
412. Huang, R.T., et al., *Somatic mutations in the GATA6 gene underlie sporadic tetralogy of Fallot.* Int J Mol Med, 2013. **31**(1): p. 51-8.
413. Gutierrez-Roelens, I., et al., *Progressive AV-block and anomalous venous return among cardiac anomalies associated with two novel missense mutations in the CSX/NKX2-5 gene.* Hum Mutat, 2002. **20**(1): p. 75-6.
414. Winston, J.B., et al., *Heterogeneity of genetic modifiers ensures normal cardiac development.* Circulation, 2010. **121**(11): p. 1313-21.
415. Granados-Riveron, J.T., et al., *Combined mutation screening of NKX2-5, GATA4, and TBX5 in congenital heart disease: multiple heterozygosity and novel mutations.* Congenit Heart Dis, 2012. **7**(2): p. 151-9.
416. Casey A. Gifford, S.S.R., Ryan Samarakoon, Hazel T. Salunga, T. Yvanka de Soysa, Yu Huang, Ping Zhou, Aryé Elfenbein, Stacia K. Wyman, Yen Kim Bui, Kimberly R. Cordes Metzler, Philip Ursell, Kathryn N. Ivey, Deepak Srivastava, *Oligogenic inheritance of congenital heart disease involving a NKX2-5 modifier.* bioRxiv, 2018.
417. Yang, Q., et al., *Paternal age and birth defects: how strong is the association?* Hum Reprod, 2007. **22**(3): p. 696-701.
418. Jenkins, K.J., et al., *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, 2007. **115**(23): p. 2995-3014.
419. Srivastava, D., *Genetic assembly of the heart: implications for congenital heart disease.* Annu Rev Physiol, 2001. **63**: p. 451-69.
420. Manning, N. and N. Archer, *A study to determine the incidence of structural congenital heart disease in monochorionic twins.* Prenat Diagn, 2006. **26**(11): p. 1062-4.

421. Herskind, A.M., D. Almind Pedersen, and K. Christensen, *Increased prevalence of congenital heart defects in monozygotic and dizygotic twins.* Circulation, 2013. **128**(11): p. 1182-8.

422. Best, K.E. and J. Rankin, *Increased risk of congenital heart disease in twins in the North of England between 1998 and 2010.* Heart, 2015. **101**(22): p. 1807-12.

423. Caputo, S., et al., *Congenital heart disease in a population of dizygotic twins: an echocardiographic study.* Int J Cardiol, 2005. **102**(2): p. 293-6.

424. Luo, Y.L., et al., *Maternal age, parity and isolated birth defects: a population-based case-control study in Shenzhen, China.* PLoS One, 2013. **8**(11): p. e81369.

425. Hollier, L.M., et al., *Maternal age and malformations in singleton births.* Obstet Gynecol, 2000. **96**(5 Pt 1): p. 701-6.

426. Sullivan, P.M., et al., *Risk of congenital heart defects in the offspring of smoking mothers: a population-based study.* J Pediatr, 2015. **166**(4): p. 978-984.e2.

427. Malik, S., et al., *Maternal smoking and congenital heart defects.* Pediatrics, 2008. **121**(4): p. e810-6.

428. Sun, J., et al., *Maternal Alcohol Consumption before and during Pregnancy and the Risks of Congenital Heart Defects in Offspring: A Systematic Review and Meta-analysis.* Congenit Heart Dis, 2015. **10**(5): p. E216-24.

429. Hernández-Díaz, S., et al., *Folic acid antagonists during pregnancy and the risk of birth defects.* N Engl J Med, 2000. **343**(22): p. 1608-14.

430. Meijer, W.M., et al., *Folic acid sensitive birth defects in association with intrauterine exposure to folic acid antagonists.* Reprod Toxicol, 2005. **20**(2): p. 203-7.

431. Li, D.K., et al., *Maternal exposure to angiotensin converting enzyme inhibitors in the first trimester and risk of malformations in offspring: a retrospective cohort study.* Bmj, 2011. **343**: p. d5931.

432. Persson, M., et al., *Risk of major congenital malformations in relation to maternal overweight and obesity severity: cohort study of 1.2 million singletons.* BMJ, 2017. **357**: p. j2563.

433. Cai, G.J., et al., *Association between maternal body mass index and congenital heart defects in offspring: a systematic review.* Am J Obstet Gynecol, 2014. **211**(2): p. 91-117.

434. Zhu, Y., et al., *Association between maternal body mass index and congenital heart defects in infants: A meta-analysis.* Congenit Heart Dis, 2018.

435. Liu, S., et al., *Association between maternal chronic conditions and congenital heart defects: a population-based cohort study.* Circulation, 2013. **128**(6): p. 583-9.

436. Ionescu-Ittu, R., et al., *Prevalence of severe congenital heart disease after folic acid fortification of grain products: time trend analysis in Quebec, Canada.* Bmj, 2009. **338**: p. b1673.

437. Liu, S., et al., *Effect of Folic Acid Food Fortification in Canada on Congenital Heart Disease Subtypes.* Circulation, 2016. **134**(9): p. 647-55.

438. Tang, L.S., et al., *Developmental consequences of abnormal folate transport during murine heart morphogenesis.* Birth Defects Res A Clin Mol Teratol, 2004. **70**(7): p. 449-58.

439. van Beynum, I.M., et al., *Maternal MTHFR 677C>T is a risk factor for congenital heart defects: effect modification by periconceptional folate supplementation.* Eur Heart J, 2006. **27**(8): p. 981-7.

440. Ellesoe, S.G., et al., *Familial co-occurrence of congenital heart defects follows distinct patterns.* Eur Heart J, 2017.

441. Nora, J.J., *Multifactorial inheritance hypothesis for the etiology of congenital heart diseases. The genetic-environmental interaction.* Circulation, 1968. **38**(3): p. 604-17.

442. Nora, J.J., C.W. McGill, and D.G. McNamara, *Empiric recurrence risks in common and uncommon congenital heart lesions.* Teratology, 1970. **3**(4): p. 325-30.

443. Nora, J.J. and A.H. Nora, *Recurrence risks in children having one parent with a congenital heart disease.* Circulation, 1976. **53**(4): p. 701-2.
444. Nora, J.J. and A.H. Nora, *The evolution of specific genetic and environmental counseling in congenital heart diseases.* Circulation, 1978. **57**(2): p. 205-13.
445. Nora, J.J. and A.H. Nora, *Maternal transmission of congenital heart diseases: new recurrence risk figures and the questions of cytoplasmic inheritance and vulnerability to teratogens.* Am J Cardiol, 1987. **59**(5): p. 459-63.
446. Burn, J. and G. Corney, *Congenital heart defects and twinning.* Acta Genet Med Gemellol (Roma), 1984. **33**(1): p. 61-9.
447. Gill, H.K., et al., *Patterns of recurrence of congenital heart disease: an analysis of 6,640 consecutive pregnancies evaluated by detailed fetal echocardiography.* J Am Coll Cardiol, 2003. **42**(5): p. 923-9.
448. Rose, V., et al., *A possible increase in the incidence of congenital heart defects among the offspring of affected parents.* J Am Coll Cardiol, 1985. **6**(2): p. 376-82.
449. Chin-Yee, N.J., et al., *Reproductive fitness and genetic transmission of tetralogy of Fallot in the molecular age.* Circ Cardiovasc Genet, 2014. **7**(2): p. 102-9.
450. Peyvandi, S., et al., *Risk of congenital heart disease in relatives of probands with conotruncal cardiac defects: an evaluation of 1,620 families.* Am J Med Genet A, 2014. **164A**(6): p. 1490-5.
451. Brenner, J.I., et al., *Cardiac malformations in relatives of infants with hypoplastic left-heart syndrome.* Am J Dis Child, 1989. **143**(12): p. 1492-4.
452. Hinton, R.B., Jr., et al., *Hypoplastic left heart syndrome is heritable.* J Am Coll Cardiol, 2007. **50**(16): p. 1590-5.
453. Lewin, M.B., et al., *Echocardiographic evaluation of asymptomatic parental and sibling cardiovascular anomalies associated with congenital left ventricular outflow tract lesions.* Pediatrics, 2004. **114**(3): p. 691-6.
454. McBride, K.L., et al., *Inheritance analysis of congenital left ventricular outflow tract obstruction malformations: Segregation, multiplex relative risk, and heritability.* Am J Med Genet A, 2005. **134a**(2): p. 180-6.
455. Kerstjens-Frederikse, W.S., et al., *Left ventricular outflow tract obstruction: should cardiac screening be offered to first-degree relatives?* Heart, 2011. **97**(15): p. 1228-32.
456. Digilio, M.C., et al., *Complete transposition of the great arteries: patterns of congenital heart disease in familial precurrence.* Circulation, 2001. **104**(23): p. 2809-14.
457. Corone, P., et al., *Familial congenital heart disease: how are the various types related?* Am J Cardiol, 1983. **51**(6): p. 942-5.
458. Franklin, R.C.G., et al., *Nomenclature for congenital and paediatric cardiac disease: the International Paediatric and Congenital Cardiac Code (IPCCC) and the Eleventh Iteration of the International Classification of Diseases (ICD-11).* Cardiol Young, 2017. **27**(10): p. 1872-1938.
459. Köhler, S., et al., *Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources.* Nucleic Acids Res, 2019. **47**(D1): p. D1018-d1027.
460. Thiene, G. and C. Frescura, *Anatomical and pathophysiological classification of congenital heart disease.* Cardiovasc Pathol, 2010. **19**(5): p. 259-74.
461. Franklin, O., et al., *Prenatal diagnosis of coarctation of the aorta improves survival and reduces morbidity.* Heart, 2002. **87**(1): p. 67-9.
462. Kutty, S., et al., *Long-term (5- to 20-year) outcomes after transcatheter or surgical treatment of hemodynamically significant isolated secundum atrial septal defect.* Am J Cardiol, 2012. **109**(9): p. 1348-52.
463. van der Bom, T., et al., *Contemporary survival of adults with congenital heart disease.* Heart, 2015. **101**(24): p. 1989-95.
464. Mandalenakis, Z., et al., *Ischemic Stroke in Children and Young Adults With Congenital Heart Disease.* J Am Heart Assoc, 2016. **5**(2).

465.	Lopatin, D.E., et al., *Concentrations of fibronectin in the sera and crevicular fluid in various stages of periodontal disease.* J Clin Periodontol, 1989. **16**(6): p. 359-64.

466.	Nyboe, C., et al., *Long-term mortality in patients with atrial septal defect: a nationwide cohort-study.* Eur Heart J, 2018. **39**(12): p. 993-998.

467.	Oliver, J.M., et al., *Risk factors for excess mortality in adults with congenital heart diseases.* Eur Heart J, 2017. **38**(16): p. 1233-1241.

468.	Marino, B.S., et al., *Neurodevelopmental outcomes in children with congenital heart disease: evaluation and management: a scientific statement from the American Heart Association.* Circulation, 2012. **126**(9): p. 1143-72.

469.	Limperopoulos, C., et al., *Predictors of developmental disabilities after open heart surgery in young children with congenital heart defects.* J Pediatr, 2002. **141**(1): p. 51-8.

470.	Harden, B., et al., *Increased postoperative respiratory complications in heterotaxy congenital heart disease patients with respiratory ciliary dysfunction.* J Thorac Cardiovasc Surg, 2014. **147**(4): p. 1291-1298.e2.

471.	Alonso-Gonzalez, R., et al., *Abnormal lung function in adults with congenital heart disease: prevalence, relation to cardiac anatomy, and association with survival.* Circulation, 2013. **127**(8): p. 882-90.

472.	Norozi, K., et al., *Incidence and risk distribution of heart failure in adolescents and adults with congenital heart disease after cardiac surgery.* Am J Cardiol, 2006. **97**(8): p. 1238-43.

473.	Kouatli, A.A., et al., *Enalapril does not enhance exercise capacity in patients after Fontan procedure.* Circulation, 1997. **96**(5): p. 1507-12.

474.	Hsu, D.T., et al., *Enalapril in infants with single ventricle: results of a multicenter randomized trial.* Circulation, 2010. **122**(4): p. 333-40.

475.	Fahed, A.C., et al., *Heart failure in congenital heart disease: a confluence of acquired and congenital.* Heart Fail Clin, 2014. **10**(1): p. 219-27.

476.	Norton, N., et al., *Evaluating pathogenicity of rare variants from dilated cardiomyopathy in the exome era.* Circ Cardiovasc Genet, 2012. **5**(2): p. 167-74.

477.	Mercer-Rosa, L., et al., *22q11.2 Deletion syndrome is associated with perioperative outcome in tetralogy of Fallot.* J Thorac Cardiovasc Surg, 2013. **146**(4): p. 868-73.

478.	Stout, K.K., et al., *2018 AHA/ACC Guideline for the Management of Adults With Congenital Heart Disease: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines.* Circulation, 2019. **139**(14): p. e637-e697.

479.	Ito, S., et al., *Appropriate Use of Genetic Testing in Congenital Heart Disease Patients.* Curr Cardiol Rep, 2017. **19**(3): p. 24.

480.	Geddes, G.C., et al., *Genetic Testing Protocol Reduces Costs and Increases Rate of Genetic Diagnosis in Infants with Congenital Heart Disease.* Pediatr Cardiol, 2017. **38**(7): p. 1465-1470.

481.	Manning, M. and L. Hudgins, *Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities.* Genet Med, 2010. **12**(11): p. 742-5.

482.	Bensemlali, M., et al., *Associated genetic syndromes and extracardiac malformations strongly influence outcomes of fetuses with congenital heart diseases.* Arch Cardiovasc Dis, 2016. **109**(5): p. 330-6.

483.	Lazier, J., et al., *Prenatal Array Comparative Genomic Hybridization in Fetuses With Structural Cardiac Anomalies.* J Obstet Gynaecol Can, 2016. **38**(7): p. 619-26.

484.	Donofrio, M.T., et al., *Diagnosis and treatment of fetal cardiac disease: a scientific statement from the American Heart Association.* Circulation, 2014. **129**(21): p. 2183-242.

485.	Jia, Y., et al., *The diagnostic value of next generation sequencing in familial nonsyndromic congenital heart defects.* Am J Med Genet A, 2015. **167a**(8): p. 1822-9.

486.	Breckpot, J., *Genetic counselling and testing in congenital heart defects and hereditary thoracic aortic disease: Complex but essential.* Eur J Prev Cardiol, 2019. **26**(15): p. 1670-1672.

487.	Burchill, L., et al., *Genetic counseling in the adult with congenital heart disease: what is the role?* Curr Cardiol Rep, 2011. **13**(4): p. 347-55.

488.	van Engelen, K., et al., *The value of the clinical geneticist caring for adults with congenital heart disease: diagnostic yield and patients' perspective.* Am J Med Genet A, 2013. **161a**(7): p. 1628-37.

489.	Deciphering Developmental Disorders, S., *Prevalence and architecture of de novo mutations in developmental disorders.* Nature, 2017. **542**(7642): p. 433-438.

490.	Van der Auwera, G.A., et al., *From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline.* Curr Protoc Bioinformatics, 2013. **43**(1110): p. 11.10.1-11.10.33.

491.	Ramu, A., et al., *DeNovoGear: de novo indel and point mutation discovery and phasing.* Nat Methods, 2013. **10**(10): p. 985-7.

492.	He, X., et al., *Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes.* PLoS Genet, 2013. **9**(8): p. e1003671.

493.	McLaren, W., et al., *The Ensembl Variant Effect Predictor.* Genome Biol, 2016. **17**(1): p. 122.

494.	Samocha, K.E., et al., *A framework for the interpretation of de novo mutation in human disease.* Nat Genet, 2014. **46**(9): p. 944-50.

495.	Deciphering Developmental Disorders, S., *Large-scale discovery of novel genetic causes of developmental disorders.* Nature, 2015. **519**(7542): p. 223-8.

496.	Wright, C.F., et al., *Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data.* Lancet, 2015. **385**(9975): p. 1305-14.

497.	Moore, C., et al., *The INTERVAL trial to determine whether intervals between blood donations can be safely and acceptably decreased to optimise blood supply: study protocol for a randomised controlled trial.* Trials, 2014. **15**: p. 363.

498.	Iossifov, I., et al., *The contribution of de novo coding mutations to autism spectrum disorder.* Nature, 2014. **515**(7526): p. 216-21.

499.	Akawi, N., et al., *Discovery of four recessive developmental disorders using probabilistic genotype and phenotype matching among 4,125 families.* Nat Genet, 2015. **47**(11): p. 1363-9.

500.	Sanders, S.J., et al., *Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci.* Neuron, 2015. **87**(6): p. 1215-1233.

501.	Breuer, K., et al., *InnateDB: systems biology of innate immunity and beyond-- recent updates and continuing curation.* Nucleic Acids Res, 2013. **41**(Database issue): p. D1228-33.

502.	Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic Acids Res, 2000. **28**(1): p. 27-30.

503.	Jassal, B., et al., *The reactome pathway knowledgebase.* Nucleic Acids Res, 2020. **48**(D1): p. D498-d503.

504.	Szklarczyk, D., et al., *STRING v10: protein-protein interaction networks, integrated over the tree of life.* Nucleic Acids Res, 2015. **43**(Database issue): p. D447-52.

505.	Amarasinghe, K.C., J. Li, and S.K. Halgamuge, *CoNVEX: copy number variation estimation in exome sequencing data using HMM.* BMC Bioinformatics, 2013. **14 Suppl 2**: p. S2.

506.	Price, T.S., et al., *SW-ARRAY: a dynamic programming solution for the identification of copy-number changes in genomic DNA using array comparative genome hybridization data.* Nucleic Acids Res, 2005. **33**(11): p. 3455-64.

507.	McRae, J., *CNV of inheritance from exome read-depth*

508. Conrad, D.F., et al., *Origins and functional impact of copy number variation in the human genome.* Nature, 2010. **464**(7289): p. 704-12.
509. Craddock, N., et al., *Strong genetic evidence for a selective influence of GABAA receptors on a component of the bipolar disorder phenotype.* Mol Psychiatry, 2010. **15**(2): p. 146-53.
510. Genomes Project, C., et al., *A map of human genome variation from population-scale sequencing.* Nature, 2010. **467**(7319): p. 1061-73.
511. Raess, M., et al., *INFRAFRONTIER: a European resource for studying the functional basis of human disease.* Mamm Genome, 2016. **27**(7-8): p. 445-50.
512. Hunt, S.E., et al., *Ensembl variation resources.* Database (Oxford), 2018. **2018**.
513. Mohun, T.J. and W.J. Weninger, *Embedding embryos for high-resolution episcopic microscopy (HREM).* Cold Spring Harb Protoc, 2012. **2012**(6): p. 678-80.
514. *QuickTime player*. 1992, [Cupertino, Calif.] : Apple Computer, [1992-2001] ©1992-2001.
515. Rosset, A., L. Spadola, and O. Ratib, *OsiriX: an open-source software for navigating in multidimensional DICOM images.* J Digit Imaging, 2004. **17**(3): p. 205-16.
516. Green, E.C., et al., *EMPReSS: European mouse phenotyping resource for standardized screens.* Bioinformatics, 2005. **21**(12): p. 2930-1.
517. Wong, M.D., et al., *A novel 3D mouse embryo atlas based on micro-CT.* Development, 2012. **139**(17): p. 3248-56.
518. Geyer, S.H., et al., *A staging system for correct phenotype interpretation of mouse embryos harvested on embryonic day 14 (E14.5).* J Anat, 2017. **230**(5): p. 710-719.
519. Stenson, P.D., et al., *The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies.* Hum Genet, 2017. **136**(6): p. 665-677.
520. Fokkema, I.F., et al., *LOVD v.2.0: the next generation in gene variant databases.* Hum Mutat, 2011. **32**(5): p. 557-63.
521. Landrum, M.J., et al., *ClinVar: improving access to variant interpretations and supporting evidence.* Nucleic Acids Res, 2018. **46**(D1): p. D1062-D1067.
522. Firth, H.V., et al., *DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources.* Am J Hum Genet, 2009. **84**(4): p. 524-33.
523. Kusch, T., *Histone H3 lysine 4 methylation revisited.* Transcription, 2012. **3**(6): p. 310-4.
524. Nomura, M. and E. Li, *Smad2 role in mesoderm formation, left-right patterning and craniofacial development.* Nature, 1998. **393**(6687): p. 786-90.
525. Dong, C., et al., *Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies.* Hum Mol Genet, 2015. **24**(8): p. 2125-37.
526. Glessner, J.T., et al., *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, 2014. **115**(10): p. 884-96.
527. Gallagher, T.L., et al., *Rbfox-regulated alternative splicing is critical for zebrafish cardiac and skeletal muscle functions.* Dev Biol, 2011. **359**(2): p. 251-61.
528. Hickey, E.J., C.A. Caldarone, and B.W. McCrindle, *Left ventricular hypoplasia: a spectrum of disease involving the left ventricular outflow tract, aortic valve, and aorta.* J Am Coll Cardiol, 2012. **59**(1 Suppl): p. S43-54.
529. Yeo, G.W., et al., *An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells.* Nat Struct Mol Biol, 2009. **16**(2): p. 130-7.
530. Gehman, L.T., et al., *The splicing regulator Rbfox2 is required for both cerebellar development and mature motor function.* Genes Dev, 2012. **26**(5): p. 445-60.

531. Hu, J., et al., *RBFox2-miR-34a-Jph2 axis contributes to cardiac decompensation during heart failure.* Proc Natl Acad Sci U S A, 2019. **116**(13): p. 6172-6180.
532. Verma, S.K., et al., *Rbfox2 function in RNA metabolism is impaired in hypoplastic left heart syndrome patient hearts.* Sci Rep, 2016. **6**: p. 30896.
533. Koolen, D.A., et al., *Mutations in the chromatin modifier gene KANSL1 cause the 17q21.31 microdeletion syndrome.* Nat Genet, 2012. **44**(6): p. 639-41.
534. Ding, H., et al., *Characterization of a helicase-like transcription factor involved in the expression of the human plasminogen activator inhibitor-1 gene.* DNA Cell Biol, 1996. **15**(6): p. 429-42.
535. White, J., et al., *POGZ truncating alleles cause syndromic intellectual disability.* Genome Med, 2016. **8**(1): p. 3.
536. Stolerman, E.S., et al., *Genetic variants in the KDM6B gene are associated with neurodevelopmental delays and dysmorphic features.* Am J Med Genet A, 2019. **179**(7): p. 1276-1286.
537. Hauser, N.S., et al., *Experience with genomic sequencing in pediatric patients with congenital cardiac defects in a large community hospital.* Mol Genet Genomic Med, 2018. **6**(2): p. 200-212.
538. Kohler, S., et al., *The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data.* Nucleic Acids Res, 2014. **42**(Database issue): p. D966-74.
539. Franklin, R.C., et al., *Bidirectional crossmap of the Short Lists of the European Paediatric Cardiac Code and the International Congenital Heart Surgery Nomenclature and Database Project.* Cardiol Young, 2002. **12**(5): p. 431-5.
540. McKusick, V.A., *Mendelian Inheritance in Man, OMIM (TM).* Johns Hopkins University Press, 1994.
541. Shinebourne, E.A., S.V. Babu-Narayan, and J.S. Carvalho, *Tetralogy of Fallot: from fetus to adult.* Heart, 2006. **92**(9): p. 1353-9.
542. Ware, S.M., M.G. Aygun, and F. Hildebrandt, *Spectrum of clinical diseases caused by disorders of primary cilia.* Proc Am Thorac Soc, 2011. **8**(5): p. 444-50.
543. Hagen, P.T., D.G. Scholz, and W.D. Edwards, *Incidence and size of patent foramen ovale during the first 10 decades of life: an autopsy study of 965 normal hearts.* Mayo Clin Proc, 1984. **59**(1): p. 17-20.
544. UniProt, C., *UniProt: a hub for protein information.* Nucleic Acids Res, 2015. **43**(Database issue): p. D204-12.
545. Barriot, R., et al., *Collaboratively charting the gene-to-phenotype network of human congenital heart defects.* Genome Med, 2010. **2**(3): p. 16.
546. Bult, C.J., et al., *Mouse Genome Database (MGD) 2019.* Nucleic Acids Res, 2019. **47**(D1): p. D801-d806.
547. Lupianez, D.G., et al., *Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions.* Cell, 2015. **161**(5): p. 1012-25.
548. Spielmann, M., et al., *Homeotic arm-to-leg transformation associated with genomic rearrangements at the PITX1 locus.* Am J Hum Genet, 2012. **91**(4): p. 629-35.
549. Szklarczyk, D., et al., *STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets.* Nucleic Acids Res, 2019. **47**(D1): p. D607-D613.
550. Chang, S., et al., *Histone deacetylases 5 and 9 govern responsiveness of the heart to a subset of stress signals and play redundant roles in heart development.* Mol Cell Biol, 2004. **24**(19): p. 8467-76.
551. Harris, L.G., et al., *Evidence for a non-canonical role of HDAC5 in regulation of the cardiac Ncx1 and Bnp genes.* Nucleic Acids Res, 2016. **44**(8): p. 3610-7.
552. Kee, H.J. and H. Kook, *Roles and targets of class I and IIa histone deacetylases in cardiac hypertrophy.* J Biomed Biotechnol, 2011. **2011**: p. 928326.

553. Feng, Y., et al., *Maternal folic acid supplementation and the risk of congenital heart defects in offspring: a meta-analysis of epidemiological observational studies.* Sci Rep, 2015. **5**: p. 8506.
554. Mao, B., et al., *Maternal folic acid supplementation and dietary folate intake and congenital heart defects.* PLoS One, 2017. **12**(11): p. e0187996.
555. Øyen, N., et al., *Association Between Maternal Folic Acid Supplementation and Congenital Heart Defects in Offspring in Birth Cohorts From Denmark and Norway.* J Am Heart Assoc, 2019. **8**(6): p. e011615.
556. Kuhn, C., et al., *DYRK1A is a novel negative regulator of cardiomyocyte hypertrophy.* J Biol Chem, 2009. **284**(25): p. 17320-7.
557. Hille, S., et al., *Dyrk1a regulates the cardiomyocyte cell cycle via D-cyclin-dependent Rb/E2f-signalling.* Cardiovasc Res, 2016. **110**(3): p. 381-94.
558. Wang, X., et al., *Phenotypic expansion in DDX3X - a common cause of intellectual disability in females.* Ann Clin Transl Neurol, 2018. **5**(10): p. 1277-1285.
559. Bellinger, D.C., et al., *Developmental and neurologic status of children after heart surgery with hypothermic circulatory arrest or low-flow cardiopulmonary bypass.* N Engl J Med, 1995. **332**(9): p. 549-55.
560. Massaro, A.N., et al., *Neurobehavioral abnormalities in newborns with congenital heart disease requiring open-heart surgery.* J Pediatr, 2011. **158**(4): p. 678-681 e2.
561. Limperopoulos, C., et al., *Neurodevelopmental status of newborns and infants with congenital heart defects before and after open heart surgery.* J Pediatr, 2000. **137**(5): p. 638-45.
562. McQuillen, P.S., et al., *Temporal and anatomic risk profile of brain injury with neonatal repair of congenital heart defects.* Stroke, 2007. **38**(2 Suppl): p. 736-41.
563. Sigmon, E.R., et al., *Congenital Heart Disease and Autism: A Case-Control Study.* Pediatrics, 2019. **144**(5).
564. O'Roak, B.J., et al., *Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders.* Science, 2012. **338**(6114): p. 1619-22.
565. Huguet, G., E. Ey, and T. Bourgeron, *The genetic landscapes of autism spectrum disorders.* Annu Rev Genomics Hum Genet, 2013. **14**: p. 191-213.
566. Schoen, C.J., et al., *Increased activity of Diaphanous homolog 3 (DIAPH3)/diaphanous causes hearing defects in humans with auditory neuropathy and in Drosophila.* Proc Natl Acad Sci U S A, 2010. **107**(30): p. 13396-401.
567. Arboleda, V.A., et al., *De novo nonsense mutations in KAT6A, a lysine acetyl-transferase gene, cause a syndrome including microcephaly and global developmental delay.* Am J Hum Genet, 2015. **96**(3): p. 498-506.
568. Millan, F., et al., *Whole exome sequencing reveals de novo pathogenic variants in KAT6A as a cause of a neurodevelopmental disorder.* Am J Med Genet A, 2016. **170**(7): p. 1791-8.
569. Tham, E., et al., *Dominant mutations in KAT6A cause intellectual disability with recognizable syndromic features.* Am J Hum Genet, 2015. **96**(3): p. 507-13.
570. Konrad J. Karczewski, L.C.F., Grace Tiao, Beryl B. Cummings, Jessica Alföldi, Qingbo Wang, Ryan L. Collins, Kristen M. Laricchia, Andrea Ganna, Daniel P. Birnbaum, Laura D. Gauthier, Harrison Brand, Matthew Solomonson, Nicholas A. Watts, Daniel Rhodes, Moriel Singer-Berk, Eleina M. England, Eleanor G. Seaby, Jack A. Kosmicki, Raymond K. Walters, Katherine Tashman, Yossi Farjoun, Eric Banks, Timothy Poterba, Arcturus Wang, Cotton Seed, Nicola Whiffin, Jessica X. Chong, Kaitlin E. Samocha, Emma Pierce-Hoffman, Zachary Zappala, Anne H. O'Donnell-Luria, Eric Vallabh Minikel, Ben Weisburd, Monkol Lek, James S. Ware, Christopher Vittal, Irina M. Armean, Louis Bergelson, Kristian Cibulskis, Kristen M. Connolly, Miguel Covarrubias, Stacey Donnelly, Steven Ferriera, Stacey Gabriel, Jeff Gentry, Namrata Gupta, Thibault Jeandet, Diane Kaplan, Christopher Llanwarne, Ruchi Munshi, Sam Novod, Nikelle Petrillo, David Roazen, Valentin Ruano-Rubio, Andrea Saltzman, Molly Schleicher, Jose Soto, Kathleen Tibbetts,

Charlotte Tolonen, Gordon Wade, Michael E. Talkowski, The Genome Aggregation Database Consortium, Benjamin M. Neale, Mark J. Daly, Daniel G. MacArthur, *Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes.* BioRxiv, 2019.

571. Fujita, K., et al., *Abnormal basiocciput development in CHARGE syndrome.* AJNR Am J Neuroradiol, 2009. **30**(3): p. 629-34.

572. Milani, D., et al., *Rubinstein-Taybi syndrome: clinical features, genetic basis, diagnosis, and management.* Ital J Pediatr, 2015. **41**: p. 4.

573. Lai, H.S. and C.Y. Feng, *Cecal volvulus in a child with CHARGE syndrome.* Am Surg, 2006. **72**(4): p. 356-8.

574. Chang, J.H., et al., *Two cases of CHARGE syndrome with multiple congenital anomalies.* Int Ophthalmol, 2014. **34**(3): p. 623-7.

575. Weber, U. and H. Bernsmeier, *[Rubinstein-Taybi syndrome and juvenile glaucoma].* Klin Monbl Augenheilkd, 1983. **183**(1): p. 47-9.

576. McMain, K., et al., *Ocular features of CHARGE syndrome.* J AAPOS, 2008. **12**(5): p. 460-5.

577. Weiss, K., et al., *De Novo Mutations in CHD4, an ATP-Dependent Chromatin Remodeler Gene, Cause an Intellectual Disability Syndrome with Distinctive Dysmorphisms.* Am J Hum Genet, 2016. **99**(4): p. 934-941.

578. Bevilacqua, A., M.S. Willis, and S.J. Bultman, *SWI/SNF chromatin-remodeling complexes in cardiovascular development and disease.* Cardiovasc Pathol, 2014. **23**(2): p. 85-91.

579. Imbalanzo, C.G.A.M.A.N., *The Chd Family of Chromatin Remodelers.* Mutant Res, 2007. **618**(1-2): p. 30-40.

580. Silva, A.P., et al., *The N-terminal Region of Chromodomain Helicase DNA-binding Protein 4 (CHD4) Is Essential for Activity and Contains a High Mobility Group (HMG) Box-like-domain That Can Bind Poly(ADP-ribose).* J Biol Chem, 2016. **291**(2): p. 924-38.

581. Lucas Farnung, M.O., Patrick Cramer, *Nucleosome-CHD4 chromatin remodeller structure explains human disease manifestations.* Biorxiv, 2019: p. 1-20.

582. Hauk, G., et al., *The chromodomains of the Chd1 chromatin remodeler regulate DNA access to the ATPase motor.* Mol Cell, 2010. **39**(5): p. 711-23.

583. Gerlitz, G., et al., *The dynamics of HMG protein-chromatin interactions in living cells.* Biochem Cell Biol, 2009. **87**(1): p. 127-37.

584. Shur, I. and D. Benayahu, *Characterization and functional analysis of CReMM, a novel chromodomain helicase DNA-binding protein.* J Mol Biol, 2005. **352**(3): p. 646-55.

585. Schuster, E.F. and R. Stoger, *CHD5 defines a new subfamily of chromodomain-SWI2/SNF2-like helicases.* Mamm Genome, 2002. **13**(2): p. 117-9.

586. Chiba, H., et al., *Two human homologues of Saccharomyces cerevisiae SWI2/SNF2 and Drosophila brahma are transcriptional coactivators cooperating with the estrogen receptor and the retinoic acid receptor.* Nucleic Acids Res, 1994. **22**(10): p. 1815-20.

587. Sillibourne, J.E., et al., *Chromatin remodeling proteins interact with pericentrin to regulate centrosome integrity.* Mol Biol Cell, 2007. **18**(9): p. 3667-80.

588. Weiss, K. and K. Lachlan, *CHD4 Neurodevelopmental Disorder*, in *GeneReviews(®)*, M.P. Adam, et al., Editors. 1993, University of Washington, Seattle

589. Sims, R.J., 3rd, K. Nishioka, and D. Reinberg, *Histone lysine methylation: a signature for chromatin function.* Trends Genet, 2003. **19**(11): p. 629-39.

590. Kolla, V., et al., *The tumour suppressor CHD5 forms a NuRD-type chromatin remodelling complex.* Biochem J, 2015. **468**(2): p. 345-52.

591. Tong, J.K., et al., *Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex.* Nature, 1998. **395**(6705): p. 917-21.

592. Xue, Y., et al., *NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities.* Mol Cell, 1998. **2**(6): p. 851-61.

593. Basta, J. and M. Rauchman, *The nucleosome remodeling and deacetylase complex in development and disease.* Transl Res, 2015. **165**(1): p. 36-47.

594. Gnanapragasam, M.N., et al., *p66Alpha-MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2-NuRD complex.* Proc Natl Acad Sci U S A, 2011. **108**(18): p. 7487-92.

595. Polo, S.E., et al., *Regulation of DNA-damage responses and cell-cycle progression by the chromatin remodelling factor CHD4.* EMBO J, 2010. **29**(18): p. 3130-9.

596. Larsen, D.H., et al., *The chromatin-remodeling factor CHD4 coordinates signaling and repair after DNA damage.* J Cell Biol, 2010. **190**(5): p. 731-40.

597. Bornelov, S., et al., *The Nucleosome Remodeling and Deacetylation Complex Modulates Chromatin Structure at Sites of Active Transcription to Fine-Tune Gene Expression.* Mol Cell, 2018. **71**(1): p. 56-72 e4.

598. Spruijt, C.G., et al., *CDK2AP1/DOC-1 is a bona fide subunit of the Mi-2/NuRD complex.* Mol Biosyst, 2010. **6**(9): p. 1700-6.

599. de Ligt, J., et al., *Diagnostic exome sequencing in persons with severe intellectual disability.* N Engl J Med, 2012. **367**(20): p. 1921-9.

600. Gururaj, A.E., et al., *MTA1, a transcriptional activator of breast cancer amplified sequence 3.* Proc Natl Acad Sci U S A, 2006. **103**(17): p. 6670-5.

601. Bouazoune, K., et al., *The dMi-2 chromodomains are DNA binding modules important for ATP-dependent nucleosome mobilization.* EMBO J, 2002. **21**(10): p. 2430-40.

602. Hendrich, B. and A. Bird, *Identification and characterization of a family of mammalian methyl-CpG binding proteins.* Mol Cell Biol, 1998. **18**(11): p. 6538-47.

603. Hong, W., et al., *FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1.* EMBO J, 2005. **24**(13): p. 2367-78.

604. Roche, A.E., et al., *The zinc finger and C-terminal domains of MTA proteins are required for FOG-2-mediated transcriptional repression via the NuRD complex.* J Mol Cell Cardiol, 2008. **44**(2): p. 352-60.

605. Murzina, N.V., et al., *Structural basis for the recognition of histone H4 by the histone-chaperone RbAp46.* Structure, 2008. **16**(7): p. 1077-85.

606. Nair, S.S., D.Q. Li, and R. Kumar, *A core chromatin remodeling factor instructs global chromatin signaling through multivalent reading of nucleosome codes.* Mol Cell, 2013. **49**(4): p. 704-18.

607. Liang, Z., et al., *A high-resolution map of transcriptional repression.* Elife, 2017. **6**.

608. Beyer, T.A., et al., *Switch enhancers interpret TGF-beta and Hippo signaling to control cell fate in human embryonic stem cells.* Cell Rep, 2013. **5**(6): p. 1611-24.

609. Ng, S.Y., et al., *Genome-wide lineage-specific transcriptional networks underscore Ikaros-dependent lymphoid priming in hematopoietic stem cells.* Immunity, 2009. **30**(4): p. 493-507.

610. Yoshida, T., et al., *The role of the chromatin remodeler Mi-2beta in hematopoietic stem cell self-renewal and multilineage differentiation.* Genes Dev, 2008. **22**(9): p. 1174-89.

611. Waldron, L., et al., *The Cardiac TBX5 Interactome Reveals a Chromatin Remodeling Network Essential for Cardiac Septation.* Dev Cell, 2016. **36**(3): p. 262-75.

612. Hoffmeister, H., et al., *CHD3 and CHD4 form distinct NuRD complexes with different yet overlapping functionality.* Nucleic Acids Res, 2017. **45**(18): p. 10534-10554.

613. Gomez-Del Arco, P., et al., *The Chromatin Remodeling Complex Chd4/NuRD Controls Striated Muscle Identity and Metabolic Homeostasis.* Cell Metab, 2016. **23**(5): p. 881-92.

614. Wilczewski, C.M., et al., *CHD4 and the NuRD complex directly control cardiac sarcomere formation.* Proc Natl Acad Sci U S A, 2018.

615. Ingram, K.G., et al., *The NuRD chromatin-remodeling enzyme CHD4 promotes embryonic vascular integrity by transcriptionally regulating extracellular matrix proteolysis.* PLoS Genet, 2013. **9**(12): p. e1004031.

616. Eberharter, A., et al., *ACF1 improves the effectiveness of nucleosome mobilization by ISWI through PHD-histone contacts.* EMBO J, 2004. **23**(20): p. 4029-39.

617. Ragvin, A., et al., *Nucleosome binding by the bromodomain and PHD finger of the transcriptional cofactor p300.* J Mol Biol, 2004. **337**(4): p. 773-88.

618. Laity, J.H., B.M. Lee, and P.E. Wright, *Zinc finger proteins: new insights into structural and functional diversity.* Curr Opin Struct Biol, 2001. **11**(1): p. 39-46.

619. Brehm, A., et al., *The many colours of chromodomains.* Bioessays, 2004. **26**(2): p. 133-40.

620. Kelley, D.E., D.G. Stokes, and R.P. Perry, *CHD1 interacts with SSRP1 and depends on both its chromodomain and its ATPase/helicase-like domain for proper association with chromatin.* Chromosoma, 1999. **108**(1): p. 10-25.

621. Wang, H.B. and Y. Zhang, *Mi2, an auto-antigen for dermatomyositis, is an ATP-dependent nucleosome remodeling factor.* Nucleic Acids Res, 2001. **29**(12): p. 2517-21.

622. Sundaramoorthy, R., et al., *Structure of the chromatin remodelling enzyme Chd1 bound to a ubiquitinylated nucleosome.* Elife, 2018. **7**.

623. Tsurusaki, Y., et al., *Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome.* Nat Genet, 2012. **44**(4): p. 376-8.

624. Farnung, L., et al., *Nucleosome-Chd1 structure and implications for chromatin remodelling.* Nature, 2017. **550**(7677): p. 539-542.

625. Liu, X., et al., *Mechanism of chromatin remodelling revealed by the Snf2-nucleosome structure.* Nature, 2017. **544**(7651): p. 440-445.

626. Korbel, J.O., et al., *Paired-end mapping reveals extensive structural variation in the human genome.* Science, 2007. **318**(5849): p. 420-6.

627. Fuller, Z.L., et al., *Measuring intolerance to mutation in human genetics.* Nat Genet, 2019. **51**(5): p. 772-776.

628. Riggs, E.R., et al., *Towards a Universal Clinical Genomics Database: the 2012 International Standards for Cytogenomic Arrays Consortium Meeting.* Hum Mutat, 2013. **34**(6): p. 915-9.

629. Nitarska, J., et al., *A Functional Switch of NuRD Chromatin Remodeling Complex Subunits Regulates Mouse Cortical Development.* Cell Rep, 2016. **17**(6): p. 1683-1698.

630. Pilarowski, G.O., et al., *Missense variants in the chromatin remodeler CHD1 are associated with neurodevelopmental disability.* J Med Genet, 2018. **55**(8): p. 561-566.

631. Carvill, G.L., et al., *Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1.* Nat Genet, 2013. **45**(7): p. 825-30.

632. Snijders Blok, L., et al., *CHD3 helicase domain mutations cause a neurodevelopmental syndrome with macrocephaly and impaired speech and language.* Nat Commun, 2018. **9**(1): p. 4619.

633. Thompson, P.M., et al., *CHD5, a new member of the chromodomain gene family, is preferentially expressed in the nervous system.* Oncogene, 2003. **22**(7): p. 1002-11.

634. Zhuang, T., et al., *CHD5 is required for spermiogenesis and chromatin condensation.* Mech Dev, 2014. **131**: p. 35-46.

635. Lathrop, M.J., et al., *Deletion of the Chd6 exon 12 affects motor coordination.* Mamm Genome, 2010. **21**(3-4): p. 130-42.

636. O'Roak, B.J., et al., *Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations.* Nature, 2012. **485**(7397): p. 246-50.

637. Bernier, R., et al., *Disruptive CHD8 mutations define a subtype of autism early in development.* Cell, 2014. **158**(2): p. 263-276.

638. Lauberth, S.M. and M. Rauchman, *A conserved 12-amino acid motif in Sall1 recruits the nucleosome remodeling and deacetylase corepressor complex.* J Biol Chem, 2006. **281**(33): p. 23922-31.

639. Ostapcuk, V., et al., *Activity-dependent neuroprotective protein recruits HP1 and CHD4 to control lineage-specifying genes.* Nature, 2018. **557**(7707): p. 739-743.

640. Amaya, M., et al., *Mi2beta-mediated silencing of the fetal gamma-globin gene in adult erythroid cells.* Blood, 2013. **121**(17): p. 3493-501.

641. Kunert, N., et al., *dMec: a novel Mi-2 chromatin remodelling complex involved in transcriptional repression.* EMBO J, 2009. **28**(5): p. 533-44.

642. Williams, C.J., et al., *The chromatin remodeler Mi-2beta is required for CD4 expression and T cell development.* Immunity, 2004. **20**(6): p. 719-33.

643. Smeenk, G., et al., *The NuRD chromatin-remodeling complex regulates signaling and repair of DNA damage.* J Cell Biol, 2010. **190**(5): p. 741-9.

644. Cancer Genome Atlas Research, N., et al., *Integrated genomic characterization of endometrial carcinoma.* Nature, 2013. **497**(7447): p. 67-73.

645. Fujita, N., et al., *MTA3 and the Mi-2/NuRD complex regulate cell fate during B lymphocyte differentiation.* Cell, 2004. **119**(1): p. 75-86.

646. Burd, C.J., et al., *UV radiation regulates Mi-2 through protein translation and stability.* J Biol Chem, 2008. **283**(50): p. 34976-82.

647. Mansfield, R.E., et al., *Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9.* J Biol Chem, 2011. **286**(13): p. 11779-91.

648. Watson, A.A., et al., *The PHD and chromo domains regulate the ATPase activity of the human chromatin remodeler CHD4.* J Mol Biol, 2012. **422**(1): p. 3-17.

649. Musselman, C.A., et al., *Bivalent recognition of nucleosomes by the tandem PHD fingers of the CHD4 ATPase is required for CHD4-mediated repression.* Proc Natl Acad Sci U S A, 2012. **109**(3): p. 787-92.

650. Zhao, S., et al., *Landscape of somatic single-nucleotide and copy-number mutations in uterine serous carcinoma.* Proc Natl Acad Sci U S A, 2013. **110**(8): p. 2916-21.

651. Little, E.G., *Multiple Pigmented Haemorrhagic Sarcoma (?): a Case for Diagnosis.* Proc R Soc Med, 1920. **13**(Dermatol Sect): p. 13-5.

652. Ge, Q., et al., *Molecular analysis of a major antigenic region of the 240-kD protein of Mi-2 autoantigen.* J Clin Invest, 1995. **96**(4): p. 1730-7.

653. Seelig, H.P., et al., *The major dermatomyositis-specific Mi-2 autoantigen is a presumed helicase involved in transcriptional activation.* Arthritis Rheum, 1995. **38**(10): p. 1389-99.

654. Seelig, H.P., et al., *Two forms of the major antigenic protein of the dermatomyositis-specific Mi-2 autoantigen.* Arthritis Rheum, 1996. **39**(10): p. 1769-71.

655. Law, M.E., et al., *Molecular cytogenetic analysis of chromosomes 1 and 19 in glioma cell lines.* Cancer Genet Cytogenet, 2005. **160**(1): p. 1-14.

656. White, P.S., et al., *Definition and characterization of a region of 1p36.3 consistently deleted in neuroblastoma.* Oncogene, 2005. **24**(16): p. 2684-94.

657. Li, A.H., et al., *Whole exome sequencing in 342 congenital cardiac left sided lesion cases reveals extensive genetic heterogeneity and complex inheritance patterns.* Genome Med, 2017. **9**(1): p. 95.

658. Zhu, M.J., et al., *Novel mutations of AXIN2 identified in a Chinese Congenital Heart Disease Cohort.* J Hum Genet, 2019. **64**(5): p. 427-435.

659. van Walree, E.S., et al., *Germline variants in HEY2 functional domains lead to congenital heart defects and thoracic aortic aneurysms.* Genet Med, 2020.

660. Sevim Bayrak, C., et al., *De novo variants in exomes of congenital heart disease patients identify risk genes and pathways.* Genome Med, 2020. **12**(1): p. 9.
661. Weiss, K., et al., *The CHD4-related syndrome: a comprehensive investigation of the clinical spectrum, genotype-phenotype correlations, and molecular basis.* Genet Med, 2020. **22**(2): p. 389-397.
662. Pinard, A., et al., *The pleiotropy associated with de novo variants in CHD4, CNOT3, and SETD5 extends to moyamoya angiopathy.* Genet Med, 2020. **22**(2): p. 427-431.
663. Shieh, J.T., A.H. Bittles, and L. Hudgins, *Consanguinity and the risk of congenital heart disease.* Am J Med Genet A, 2012. **158A**(5): p. 1236-41.
664. Winston, J.B., et al., *Complex trait analysis of ventricular septal defects caused by Nkx2-5 mutation.* Circ Cardiovasc Genet, 2012. **5**(3): p. 293-300.
665. Bjornsson, T., et al., *A rare missense mutation in MYH6 associates with non-syndromic coarctation of the aorta.* Eur Heart J, 2018.
666. Duenas, A., et al., *The Role of Non-Coding RNA in Congenital Heart Diseases.* J Cardiovasc Dev Dis, 2019. **6**(2).
667. Ha, C.H., et al., *Protein kinase D-dependent phosphorylation and nuclear export of histone deacetylase 5 mediates vascular endothelial growth factor-induced gene expression and angiogenesis.* J Biol Chem, 2008. **283**(21): p. 14590-9.
668. Wang, S., et al., *Control of endothelial cell proliferation and migration by VEGF signaling to histone deacetylase 7.* Proc Natl Acad Sci U S A, 2008. **105**(22): p. 7738-43.
669. Eiseler, T., et al., *Protein kinase D1 regulates cofilin-mediated F-actin reorganization and cell motility through slingshot.* Nat Cell Biol, 2009. **11**(5): p. 545-56.
670. Storz, P. and A. Toker, *Protein kinase D mediates a stress-induced NF-kappaB activation and survival pathway.* EMBO J, 2003. **22**(1): p. 109-20.
671. Hurd, C., R.T. Waldron, and E. Rozengurt, *Protein kinase D complexes with C-Jun N-terminal kinase via activation loop phosphorylation and phosphorylates the C-Jun N-terminus.* Oncogene, 2002. **21**(14): p. 2154-60.
672. Endo, K., et al., *Proteolytic cleavage and activation of protein kinase C [micro] by caspase-3 in the apoptotic response of cells to 1-beta -D-arabinofuranosylcytosine and other genotoxic agents.* J Biol Chem, 2000. **275**(24): p. 18476-81.
673. Rozengurt, E., O. Rey, and R.T. Waldron, *Protein kinase D signaling.* J Biol Chem, 2005. **280**(14): p. 13205-8.
674. Baron, C.L. and V. Malhotra, *Role of diacylglycerol in PKD recruitment to the TGN and protein transport to the plasma membrane.* Science, 2002. **295**(5553): p. 325-8.
675. Czöndör, K., et al., *Protein kinase D controls the integrity of Golgi apparatus and the maintenance of dendritic arborization in hippocampal neurons.* Mol Biol Cell, 2009. **20**(7): p. 2108-20.
676. Sidorenko, S.P., et al., *Protein kinase C mu (PKC mu) associates with the B cell antigen receptor complex and regulates lymphocyte signaling.* Immunity, 1996. **5**(4): p. 353-63.
677. Ivison, S.M., et al., *Protein kinase D interaction with TLR5 is required for inflammatory signaling in response to bacterial flagellin.* J Immunol, 2007. **178**(9): p. 5735-43.
678. Hausser, A., et al., *Protein kinase D regulates vesicular transport by phosphorylating and activating phosphatidylinositol-4 kinase IIIbeta at the Golgi complex.* Nat Cell Biol, 2005. **7**(9): p. 880-6.
679. Yeaman, C., et al., *Protein kinase D regulates basolateral membrane protein exit from trans-Golgi network.* Nat Cell Biol, 2004. **6**(2): p. 106-12.
680. Woods, A.J., et al., *PKD1/PKCmu promotes alphavbeta3 integrin recycling and delivery to nascent focal adhesions.* Embo j, 2004. **23**(13): p. 2531-43.

681. Johannessen, M., et al., *Protein kinase D induces transcription through direct phosphorylation of the cAMP-response element-binding protein.* J Biol Chem, 2007. **282**(20): p. 14777-87.
682. Rykx, A., et al., *Protein kinase D: a family affair.* FEBS Lett, 2003. **546**(1): p. 81-6.
683. Weinreb, I., et al., *Novel PRKD gene rearrangements and variant fusions in cribriform adenocarcinoma of salivary gland origin.* Genes Chromosomes Cancer, 2014. **53**(10): p. 845-56.
684. Matthews, S.A., et al., *Unique functions for protein kinase D1 and protein kinase D2 in mammalian cells.* Biochem J, 2010. **432**(1): p. 153-63.
685. Oster, H., D. Abraham, and M. Leitges, *Expression of the protein kinase D (PKD) family during mouse embryogenesis.* Gene Expr Patterns, 2006. **6**(4): p. 400-8.
686. Sundram, V., S.C. Chauhan, and M. Jaggi, *Emerging roles of protein kinase D1 in cancer.* Mol Cancer Res, 2011. **9**(8): p. 985-96.
687. Van Lint, J., et al., *Protein kinase D: an intracellular traffic regulator on the move.* Trends Cell Biol, 2002. **12**(4): p. 193-200.
688. Valverde, A.M., et al., *Molecular cloning and characterization of protein kinase D: a target for diacylglycerol and phorbol esters with a distinctive catalytic domain.* Proc Natl Acad Sci U S A, 1994. **91**(18): p. 8572-6.
689. Cuello, F., et al., *Protein kinase D selectively targets cardiac troponin I and regulates myofilament Ca2+ sensitivity in ventricular myocytes.* Circ Res, 2007. **100**(6): p. 864-73.
690. Döppler, H. and P. Storz, *A novel tyrosine phosphorylation site in protein kinase D contributes to oxidative stress-mediated activation.* J Biol Chem, 2007. **282**(44): p. 31873-81.
691. Matthews, S.A., E. Rozengurt, and D. Cantrell, *Characterization of serine 916 as an in vivo autophosphorylation site for protein kinase D/Protein kinase Cmu.* J Biol Chem, 1999. **274**(37): p. 26543-9.
692. Storz, P., et al., *Tyrosine phosphorylation of protein kinase D in the pleckstrin homology domain leads to activation.* J Biol Chem, 2003. **278**(20): p. 17969-76.
693. Rey, O. and E. Rozengurt, *Protein kinase D interacts with Golgi via its cysteine-rich domain.* Biochem Biophys Res Commun, 2001. **287**(1): p. 21-6.
694. Iglesias, T., S. Matthews, and E. Rozengurt, *Dissimilar phorbol ester binding properties of the individual cysteine-rich motifs of protein kinase D.* FEBS Lett, 1998. **437**(1-2): p. 19-23.
695. Iglesias, T. and E. Rozengurt, *Protein kinase D activation by mutations within its pleckstrin homology domain.* J Biol Chem, 1998. **273**(1): p. 410-6.
696. Iglesias, T. and E. Rozengurt, *Protein kinase D activation by deletion of its cysteine-rich motifs.* FEBS Lett, 1999. **454**(1-2): p. 53-6.
697. Chen, J., et al., *Selective binding of phorbol esters and diacylglycerol by individual C1 domains of the PKD family.* Biochem J, 2008. **411**(2): p. 333-42.
698. Cozier, G.E., et al., *Membrane targeting by pleckstrin homology domains.* Curr Top Microbiol Immunol, 2004. **282**: p. 49-88.
699. Sánchez-Ruiloba, L., et al., *Protein kinase D intracellular localization and activity control kinase D-interacting substrate of 220-kDa traffic through a postsynaptic density-95/discs large/zonula occludens-1-binding motif.* J Biol Chem, 2006. **281**(27): p. 18888-900.
700. Rey, O., et al., *Rapid protein kinase D translocation in response to G protein-coupled receptor activation. Dependence on protein kinase C.* J Biol Chem, 2001. **276**(35): p. 32616-26.
701. Rey, O., et al., *Regulated nucleocytoplasmic transport of protein kinase D in response to G protein-coupled receptor activation.* J Biol Chem, 2001. **276**(52): p. 49228-35.
702. Matthews, S., et al., *Dynamic re-distribution of protein kinase D (PKD) as revealed by a GFP-PKD fusion protein: dissociation from PKD activation.* FEBS Lett, 1999. **457**(3): p. 515-21.

703. Matthews, S.A., et al., *Spatial and temporal regulation of protein kinase D (PKD).* Embo j, 2000. **19**(12): p. 2935-45.
704. Storz, P., et al., *Protein kinase C [micro] is regulated by the multifunctional chaperon protein p32.* J Biol Chem, 2000. **275**(32): p. 24601-7.
705. Prestle, J., et al., *Protein kinase C mu is located at the Golgi compartment.* J Cell Biol, 1996. **134**(6): p. 1401-10.
706. Liljedahl, M., et al., *Protein kinase D regulates the fission of cell surface destined transport carriers from the trans-Golgi network.* Cell, 2001. **104**(3): p. 409-20.
707. Jamora, C., et al., *Gbetagamma-mediated regulation of Golgi organization is through the direct activation of protein kinase D.* Cell, 1999. **98**(1): p. 59-68.
708. Bisbal, M., et al., *Protein kinase d regulates trafficking of dendritic membrane proteins in developing neurons.* J Neurosci, 2008. **28**(37): p. 9297-308.
709. Maeda, Y., et al., *Recruitment of protein kinase D to the trans-Golgi network via the first cysteine-rich domain.* Embo j, 2001. **20**(21): p. 5982-90.
710. Iglesias, T., R.T. Waldron, and E. Rozengurt, *Identification of in vivo phosphorylation sites required for protein kinase D activation.* J Biol Chem, 1998. **273**(42): p. 27662-7.
711. Waldron, R.T., et al., *Activation loop Ser744 and Ser748 in protein kinase D are transphosphorylated in vivo.* J Biol Chem, 2001. **276**(35): p. 32606-15.
712. Vertommen, D., et al., *Regulation of protein kinase D by multisite phosphorylation. Identification of phosphorylation sites by mass spectrometry and characterization by site-directed mutagenesis.* J Biol Chem, 2000. **275**(26): p. 19567-76.
713. Cowell, C.F., et al., *Mitochondrial diacylglycerol initiates protein-kinase D1-mediated ROS signaling.* J Cell Sci, 2009. **122**(Pt 7): p. 919-28.
714. Smith, C.M., et al., *The mouse Gene Expression Database (GXD): 2019 update.* Nucleic Acids Res, 2019. **47**(D1): p. D774-d779.
715. Vega, R.B., et al., *Protein kinases C and D mediate agonist-dependent cardiac hypertrophy through nuclear export of histone deacetylase 5.* Mol Cell Biol, 2004. **24**(19): p. 8374-85.
716. Matthews, S.A., et al., *Essential role for protein kinase D family kinases in the regulation of class II histone deacetylases in B lymphocytes.* Mol Cell Biol, 2006. **26**(4): p. 1569-77.
717. Bossard, C., et al., *Dimeric PKD regulates membrane fission to form transport carriers at the TGN.* J Cell Biol, 2007. **179**(6): p. 1123-31.
718. Li, J., et al., *The role of protein kinase D in neurotensin secretion mediated by protein kinase C-alpha/-delta and Rho/Rho kinase.* J Biol Chem, 2004. **279**(27): p. 28466-74.
719. von Wichert, G., et al., *Protein kinase D2 regulates chromogranin A secretion in human BON neuroendocrine tumour cells.* Cell Signal, 2008. **20**(5): p. 925-34.
720. Li, J., et al., *PKD1, PKD2, and their substrate Kidins220 regulate neurotensin secretion in the BON human endocrine cell line.* J Biol Chem, 2008. **283**(5): p. 2614-21.
721. Fugmann, T., et al., *Regulation of secretory transport by protein kinase D-mediated phosphorylation of the ceramide transfer protein.* J Cell Biol, 2007. **178**(1): p. 15-22.
722. Watkins, J.L., et al., *Phosphorylation of the Par-1 polarity kinase by protein kinase D regulates 14-3-3 binding and membrane association.* Proc Natl Acad Sci U S A, 2008. **105**(47): p. 18378-83.
723. Ding, G., et al., *Protein kinase D-mediated phosphorylation and nuclear export of sphingosine kinase 2.* J Biol Chem, 2007. **282**(37): p. 27493-502.
724. Liu, P., et al., *Protein kinase D enzymes are dispensable for proliferation, survival and antigen receptor-regulated NFkappaB activity in vertebrate B-cells.* FEBS Lett, 2007. **581**(7): p. 1377-82.

725.   Yuan, J. and E. Rozengurt, *PKD, PKD2, and p38 MAPK mediate Hsp27 serine-82 phosphorylation induced by neurotensin in pancreatic cancer PANC-1 cells.* J Cell Biochem, 2008. **103**(2): p. 648-62.

726.   Storz, P., H. Döppler, and A. Toker, *Protein kinase D mediates mitochondrion-to-nucleus signaling and detoxification from mitochondrial reactive oxygen species.* Mol Cell Biol, 2005. **25**(19): p. 8520-30.

727.   Auer, A., et al., *Role of the regulatory domain of protein kinase D2 in phorbol ester binding, catalytic activity, and nucleocytoplasmic shuttling.* Mol Biol Cell, 2005. **16**(9): p. 4375-85.

728.   Rey, O., et al., *Protein kinase C nu/protein kinase D3 nuclear localization, catalytic activation, and intracellular redistribution in response to G protein-coupled receptor agonists.* J Biol Chem, 2003. **278**(26): p. 23773-85.

729.   Sumara, G., et al., *Regulation of PKD by the MAPK p38delta in insulin secretion and glucose homeostasis.* Cell, 2009. **136**(2): p. 235-48.

730.   Park, J.E., Y.I. Kim, and A.K. Yi, *Protein kinase D1 is essential for MyD88-dependent TLR signaling pathway.* J Immunol, 2009. **182**(10): p. 6316-27.

731.   Jadali, A. and S. Ghazizadeh, *Protein kinase D is implicated in the reversible commitment to differentiation in primary cultures of mouse keratinocytes.* J Biol Chem, 2010. **285**(30): p. 23387-97.

732.   Fielitz, J., et al., *Requirement of protein kinase D1 for pathological cardiac remodeling.* Proc Natl Acad Sci U S A, 2008. **105**(8): p. 3059-63.

733.   Jensen, E.D., R. Gopalakrishnan, and J.J. Westendorf, *Bone morphogenic protein 2 activates protein kinase D to regulate histone deacetylase 7 localization and repression of Runx2.* J Biol Chem, 2009. **284**(4): p. 2225-34.

734.   Monovich, L., et al., *A novel kinase inhibitor establishes a predominant role for protein kinase D as a cardiac class IIa histone deacetylase kinase.* FEBS Lett, 2010. **584**(3): p. 631-7.

735.   Dequiedt, F., et al., *Phosphorylation of histone deacetylase 7 by protein kinase D mediates T cell receptor-induced Nur77 expression and apoptosis.* J Exp Med, 2005. **201**(5): p. 793-804.

736.   Zhang, C.L., et al., *Class II histone deacetylases act as signal-responsive repressors of cardiac hypertrophy.* Cell, 2002. **110**(4): p. 479-88.

737.   McKinsey, T.A., C.L. Zhang, and E.N. Olson, *Activation of the myocyte enhancer factor-2 transcription factor by calcium/calmodulin-dependent protein kinase-stimulated binding of 14-3-3 to histone deacetylase 5.* Proc Natl Acad Sci U S A, 2000. **97**(26): p. 14400-5.

738.   Edmondson, D.G., et al., *Mef2 gene expression marks the cardiac and skeletal muscle lineages during mouse embryogenesis.* Development, 1994. **120**(5): p. 1251-63.

739.   Lin, Q., et al., *Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C.* Science, 1997. **276**(5317): p. 1404-7.

740.   Bi, W., C.J. Drake, and J.J. Schwarz, *The transcription factor MEF2C-null mouse exhibits complex vascular malformations and reduced cardiac expression of angiopoietin 1 and VEGF.* Dev Biol, 1999. **211**(2): p. 255-67.

741.   Avkiran, M., et al., *Protein kinase d in the cardiovascular system: emerging roles in health and disease.* Circ Res, 2008. **102**(2): p. 157-63.

742.   Harrison, B.C., et al., *Regulation of cardiac stress signaling by protein kinase d1.* Mol Cell Biol, 2006. **26**(10): p. 3875-88.

743.   Haworth, R.S., et al., *Expression and activity of protein kinase D/protein kinase C mu in myocardium: evidence for alpha1-adrenergic receptor- and protein kinase C-mediated regulation.* J Mol Cell Cardiol, 2000. **32**(6): p. 1013-23.

744.   Haworth, R.S., et al., *Regulation of protein kinase D activity in adult myocardium: novel counter-regulatory roles for protein kinase Cepsilon and protein kinase A.* J Mol Cell Cardiol, 2007. **43**(6): p. 686-95.

745. Wong, C. and Z.G. Jin, *Protein kinase C-dependent protein kinase D activation modulates ERK signal pathway and endothelial cell proliferation by vascular endothelial growth factor.* J Biol Chem, 2005. **280**(39): p. 33262-9.

746. Abedi, H., E. Rozengurt, and I. Zachary, *Rapid activation of the novel serine/threonine protein kinase, protein kinase D by phorbol esters, angiotensin II and PDGF-BB in vascular smooth muscle cells.* FEBS Lett, 1998. **427**(2): p. 209-12.

747. Tan, M., et al., *Angiotensin II-induced protein kinase D activation is regulated by protein kinase Cdelta and mediated via the angiotensin II type 1 receptor in vascular smooth muscle cells.* Arterioscler Thromb Vasc Biol, 2004. **24**(12): p. 2271-6.

748. Tan, M., et al., *Thrombin rapidly induces protein kinase D phosphorylation, and protein kinase C delta mediates the activation.* J Biol Chem, 2003. **278**(5): p. 2824-8.

749. Carmeliet, P., *Angiogenesis in health and disease.* Nat Med, 2003. **9**(6): p. 653-60.

750. Bolli, R., *Preconditioning: a paradigm shift in the biology of myocardial ischemia.* Am J Physiol Heart Circ Physiol, 2007. **292**(1): p. H19-27.

751. Döppler, H., et al., *A phosphorylation state-specific antibody recognizes Hsp27, a novel substrate of protein kinase D.* J Biol Chem, 2005. **280**(15): p. 15013-9.

752. Wang, Y., et al., *Interaction between protein kinase Cmu and the vanilloid receptor type 1.* J Biol Chem, 2004. **279**(51): p. 53674-82.

753. Wang, L. and D.H. Wang, *TRPV1 gene knockout impairs postischemic recovery in isolated perfused heart in mice.* Circulation, 2005. **112**(23): p. 3617-23.

754. Hollander, J.M., et al., *Overexpression of wild-type heat shock protein 27 and a nonphosphorylatable heat shock protein 27 mutant protects against ischemia/reperfusion injury in a transgenic mouse model.* Circulation, 2004. **110**(23): p. 3544-52.

755. Ford, J.J., et al., *Protein kinase D1 is essential for bone acquisition during pubertal growth.* Endocrinology, 2013. **154**(11): p. 4182-91.

756. Bollag, W.B., et al., *Deletion of protein kinase D1 in osteoprogenitor cells results in decreased osteogenesis in vitro and reduced bone mineral density in vivo.* Mol Cell Endocrinol, 2018. **461**: p. 22-31.

757. Bollag, W.B., et al., *Protein kinase D1 conditional null mice show minimal bone loss following ovariectomy.* Mol Cell Endocrinol, 2018. **474**: p. 176-183.

758. Li, S., et al., *A Conditional Knockout Mouse Model Reveals a Critical Role of PKD1 in Osteoblast Differentiation and Bone Development.* Sci Rep, 2017. **7**: p. 40505.

759. Wang, L., Y. Mishina, and F. Liu, *Osterix-Cre transgene causes craniofacial bone development defect.* Calcif Tissue Int, 2015. **96**(2): p. 129-37.

760. Celil, A.B. and P.G. Campbell, *BMP-2 and insulin-like growth factor-I mediate Osterix (Osx) expression in human mesenchymal stem cells via the MAPK and protein kinase D signaling pathways.* J Biol Chem, 2005. **280**(36): p. 31353-9.

761. Postma, A.V., et al., *A gain-of-function TBX5 mutation is associated with atypical Holt-Oram syndrome and paroxysmal atrial fibrillation.* Circ Res, 2008. **102**(11): p. 1433-42.

762. Rashel, M., N. Alston, and S. Ghazizadeh, *Protein kinase D1 has a key role in wound healing and skin carcinogenesis.* J Invest Dermatol, 2014. **134**(4): p. 902-909.

763. Ernest Dodd, M., et al., *Regulation of protein kinase D during differentiation and proliferation of primary mouse keratinocytes.* J Invest Dermatol, 2005. **125**(2): p. 294-306.

764. Rennecke, J., et al., *Protein-kinase-Cmu expression correlates with enhanced keratinocyte proliferation in normal and neoplastic mouse epidermis and in cell culture.* Int J Cancer, 1999. **80**(1): p. 98-103.

765. Cen, C., et al., *PKD1 Promotes Functional Synapse Formation Coordinated with N-Cadherin in Hippocampus.* J Neurosci, 2018. **38**(1): p. 183-199.

766. Khan, S., et al., *A role for PKD1 in insulin secretion downstream of P2Y(1) receptor activation in mouse and human islets.* Physiol Rep, 2019. **7**(19): p. e14250.

767. Bergeron, V., et al., *Deletion of Protein Kinase D1 in Pancreatic β-Cells Impairs Insulin Secretion in High-Fat Diet-Fed Mice.* Diabetes, 2018. **67**(1): p. 71-77.

768. Kaitlin E. Samocha, J.A.K., Konrad J. Karczewski, Anne H. O'Donnell-Luria, Emma Pierce-Hoffman, Daniel G. MacArthur, Benjamin M. Neale, Mark J. Daly, *Regional missense constraint improves variant deleteriousness prediction.* BioRxiv, 2017.

769. Waldron, R.T., T. Iglesias, and E. Rozengurt, *Phosphorylation-dependent protein kinase D activation.* Electrophoresis, 1999. **20**(2): p. 382-90.

770. Waldron, R.T., T. Iglesias, and E. Rozengurt, *The pleckstrin homology domain of protein kinase D interacts preferentially with the eta isoform of protein kinase C.* J Biol Chem, 1999. **274**(14): p. 9224-30.

771. Shaheen, R., et al., *Positional mapping of PRKD1, NRP1 and PRDM1 as novel candidate disease genes in truncus arteriosus.* J Med Genet, 2015. **52**(5): p. 322-9.

772. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans.* Nature, 2016. **536**(7616): p. 285-91.

773. Huang, N., et al., *Characterising and predicting haploinsufficiency in the human genome.* PLoS Genet, 2010. **6**(10): p. e1001154.

774. Bardswell, S.C., et al., *Distinct sarcomeric substrates are responsible for protein kinase D-mediated regulation of cardiac myofilament Ca2+ sensitivity and cross-bridge cycling.* J Biol Chem, 2010. **285**(8): p. 5674-82.

775. Su, Y., et al., *The protein kinase D1-mediated classical protein secretory pathway regulates the Ras oncogene-induced senescence response.* J Cell Sci, 2018. **131**(6).

776. Weinreb, I., et al., *Hotspot activating PRKD1 somatic mutations in polymorphous low-grade adenocarcinomas of the salivary glands.* Nat Genet, 2014. **46**(11): p. 1166-9.

777. Varghese, R., et al., *Surgical Repair of Interrupted Aortic Arch and Interrupted Pulmonary Artery.* Ann Thorac Surg, 2015. **100**(6): p. e139-40.

778. Schreiber, C., et al., *The interrupted aortic arch: an overview after 20 years of surgical treatment.* Eur J Cardiothorac Surg, 1997. **12**(3): p. 466-9; discussion 469-70.

779. Sato, S., et al., *Interruption of the aortic arch: diagnosis with multidetector computed tomography.* Jpn J Radiol, 2011. **29**(1): p. 46-50.

780. Mendoza Díaz, P.M., M. Herrera Gomar, and J. Rojano Castillo, *Interrupted Aortic Arch in an Adult and Myocardial Infarction.* Rev Esp Cardiol (Engl Ed), 2016. **69**(2): p. 212.

781. Kellenberger, C.J., *Aortic arch malformations.* Pediatr Radiol, 2010. **40**(6): p. 876-84.

782. Vitelli, F., et al., *Tbx1 mutation causes multiple cardiovascular defects and disrupts neural crest and cranial nerve migratory pathways.* Hum Mol Genet, 2002. **11**(8): p. 915-22.

783. Merscher, S., et al., *TBX1 is responsible for cardiovascular defects in velo-cardio-facial/DiGeorge syndrome.* Cell, 2001. **104**(4): p. 619-29.

784. Zhang, Z., T. Huynh, and A. Baldini, *Mesodermal expression of Tbx1 is necessary and sufficient for pharyngeal arch and cardiac outflow tract development.* Development, 2006. **133**(18): p. 3587-95.

785. Ghosh, T.K., et al., *HDAC4 and 5 repression of TBX5 is relieved by protein kinase D1.* Sci Rep, 2019. **9**(1): p. 17992.

786. Barnes, R.M., et al., *MEF2C regulates outflow tract alignment and transcriptional control of Tdgf1.* Development, 2016. **143**(5): p. 774-9.

787. Papaioannou, V.E. and R.R. Behringer, *Early embryonic lethality in genetically engineered mice: diagnosis and phenotypic analysis.* Vet Pathol, 2012. **49**(1): p. 64-70.

788. Cross, J.C., *Formation of the placenta and extraembryonic membranes.* Ann N Y Acad Sci, 1998. **857**: p. 23-32.

789. Peters, J.M., et al., *Di(2-ethylhexyl) phthalate induces a functional zinc deficiency during pregnancy and teratogenesis that is independent of peroxisome proliferator-activated receptor-alpha.* Teratology, 1997. **56**(5): p. 311-6.

790. Ward, J.M., S.A. Elmore, and J.F. Foley, *Pathology methods for the evaluation of embryonic and perinatal developmental defects and lethality in genetically engineered mice.* Vet Pathol, 2012. **49**(1): p. 71-84.

791. Fagerberg, L., et al., *Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics.* Mol Cell Proteomics, 2014. **13**(2): p. 397-406.

792. Hisa, T., et al., *Hematopoietic, angiogenic and eye defects in Meis1 mutant animals.* Embo j, 2004. **23**(2): p. 450-9.

793. Lin, Q., et al., *Requirement of the MADS-box transcription factor MEF2C for vascular development.* Development, 1998. **125**(22): p. 4565-74.

794. Kodo, K., et al., *Genetic analysis of essential cardiac transcription factors in 256 patients with non-syndromic congenital heart defects.* Circ J, 2012. **76**(7): p. 1703-11.

795. Christiansen, J., et al., *Chromosome 1q21.1 contiguous gene deletion is associated with congenital heart disease.* Circ Res, 2004. **94**(11): p. 1429-35.

796. Kurihara, Y., et al., *Aortic arch malformations and ventricular septal defect in mice deficient in endothelin-1.* J Clin Invest, 1995. **96**(1): p. 293-300.

797. Winnier, G.E., et al., *Roles for the winged helix transcription factors MF1 and MFH1 in cardiovascular development revealed by nonallelic noncomplementation of null alleles.* Dev Biol, 1999. **213**(2): p. 418-31.

798. Iida, K., et al., *Essential roles of the winged helix transcription factor MFH-1 in aortic arch patterning and skeletogenesis.* Development, 1997. **124**(22): p. 4627-38.

799. Poelmann, R.E. and A.C. Gittenberger-de Groot, *Apoptosis as an instrument in cardiovascular development.* Birth Defects Res C Embryo Today, 2005. **75**(4): p. 305-13.

800. Häussermann, S., et al., *Proteolytic cleavage of protein kinase Cmu upon induction of apoptosis in U937 cells.* FEBS Lett, 1999. **462**(3): p. 442-6.

801. Vántus, T., et al., *Doxorubicin-induced activation of protein kinase D1 through caspase-mediated proteolytic cleavage: identification of two cleavage sites by microsequencing.* Cell Signal, 2004. **16**(6): p. 703-9.

802. McKinsey, T.A. and E.N. Olson, *Toward transcriptional therapies for the failing heart: chemical screens to modulate genes.* J Clin Invest, 2005. **115**(3): p. 538-46.

803. Olson, E.N., J. Backs, and T.A. McKinsey, *Control of cardiac hypertrophy and heart failure by histone acetylation/deacetylation.* Novartis Found Symp, 2006. **274**: p. 3-12; discussion 13-9, 152-5, 272-6.

804. Arise, K.K., et al., *Angiotensin II represses Npr1 expression and receptor function by recruitment of transcription factors CREB and HSF-4a and activation of HDACs.* Sci Rep, 2020. **10**(1): p. 4337.

805. Zhang, X., et al., *Genome-wide analysis of cAMP-response element binding protein occupancy, phosphorylation, and target gene activation in human tissues.* Proc Natl Acad Sci U S A, 2005. **102**(12): p. 4459-64.

806. Tissandier, O., et al., *[Atrial natriuretic factor and brain natriuretic peptide. Variations in elderly subjects with heart failure].* Presse Med, 1995. **24**(38): p. 1837-41.

807. Nelson, T.J., et al., *SRF-dependent gene expression in isolated cardiomyocytes: regulation of genes involved in cardiac hypertrophy.* J Mol Cell Cardiol, 2005. **39**(3): p. 479-89.

808. Morel, E., et al., *cAMP-binding protein Epac induces cardiomyocyte hypertrophy.* Circ Res, 2005. **97**(12): p. 1296-304.

809. Zhang, Y., et al., *Cardiomyocyte PKA Ablation Enhances Basal Contractility While Eliminates Cardiac β-Adrenergic Response Without Adverse Effects on the Heart.* Circ Res, 2019. **124**(12): p. 1760-1777.

810. Liu, X., et al., *Imaging techniques for visualizing and phenotyping congenital heart defects in murine models.* Birth Defects Res C Embryo Today, 2013. **99**(2): p. 93-105.

811. Bamforth, S.D., J.E. Schneider, and S. Bhattacharya, *High-throughput analysis of mouse embryos by magnetic resonance imaging.* Cold Spring Harb Protoc, 2012. **2012**(1): p. 93-101.

812. Parasoglou, P., et al., *High-resolution MRI of early-stage mouse embryos.* NMR Biomed, 2013. **26**(2): p. 224-31.

813. Deans, A.E., et al., *Mn enhancement and respiratory gating for in utero MRI of the embryonic mouse central nervous system.* Magn Reson Med, 2008. **59**(6): p. 1320-8.

814. Schneider, J.E., et al., *High-resolution imaging of normal anatomy, and neural and adrenal malformations in mouse embryos using magnetic resonance microscopy.* J Anat, 2003. **202**(2): p. 239-47.

815. Schneider, J.E., et al., *High-resolution, high-throughput magnetic paragraph sign resonance imaging of mouse embryonic paragraph sign anatomy using a fast gradient-echo sequence.* MAGMA, 2003. **16**(1): p. 43-51.

816. Schneider, J.E., et al., *Rapid identification and 3D reconstruction of complex cardiac malformations in transgenic mouse embryos using fast gradient echo sequence magnetic resonance imaging.* J Mol Cell Cardiol, 2003. **35**(2): p. 217-22.

817. Schneider, J.E., et al., *Identification of cardiac malformations in mice lacking Ptdsr using a novel high-throughput magnetic resonance imaging technique.* BMC Dev Biol, 2004. **4**: p. 16.

818. Zouagui, T., et al., *3D MRI heart segmentation of mouse embryos.* Comput Biol Med, 2010. **40**(1): p. 64-74.

819. Gabbay-Benziv, R., et al., *A step-wise approach for analysis of the mouse embryonic heart using 17.6Tesla MRI.* Magn Reson Imaging, 2017. **35**: p. 46-53.

820. Taylor, A.M., O.J. Arthurs, and N.J. Sebire, *Postmortem cardiac imaging in fetuses and children.* Pediatr Radiol, 2015. **45**(4): p. 549-55.

821. von Bohlen Und Halbach, O., M. Lotze, and J.P. Pfannmoller, *Post-mortem magnetic resonance microscopy (MRM) of the murine brain at 7 Tesla results in a gain of resolution as compared to in vivo MRM.* Front Neuroanat, 2014. **8**: p. 47.

822. Cahill, L.S., et al., *Preparation of fixed mouse brains for MRI.* Neuroimage, 2012. **60**(2): p. 933-9.

823. Crowe, L.A., et al., *4D cardiac imaging at clinical 3.0T provides accurate assessment of murine myocardial function and viability.* Magn Reson Imaging, 2017. **44**: p. 46-54.

824. Turnbull, D.H. and S. Mori, *MRI in mouse developmental biology.* NMR Biomed, 2007. **20**(3): p. 265-74.

825. Doevendans, P.A., et al., *Cardiovascular phenotyping in mice.* Cardiovasc Res, 1998. **39**(1): p. 34-49.

826. Baumann, P.Q., et al., *Gender-dependent differences in echocardiographic characteristics of murine hearts.* Echocardiography, 2008. **25**(7): p. 739-48.

827. Hinton, R.B., Jr., et al., *Mouse heart valve structure and function: echocardiographic and morphometric analyses from the fetus through the aged adult.* Am J Physiol Heart Circ Physiol, 2008. **294**(6): p. H2480-8.

828. Baraibar, M.A. and P. Schoning, *Effects of freezing and frozen storage on histological characteristics of canine tissues.* J Forensic Sci, 1985. **30**(2): p. 439-47.

829. Schäfer, A.T. and J.D. Kaufmann, *What happens in freezing bodies? Experimental study of histological tissue change caused by freezing injuries.* Forensic Sci Int, 1999. **102**(2-3): p. 149-58.

830. Sharma, A.R.S.a.N., *Effect of Freezing and Thawing on the Histology and Ultrastructure of Buffalo Muscle.* Asian Australasian Journal of Animal Sciences, 2004. **17**(9): p. 1291-1295.

831. Banker, M.C., et al., *Freezing preservation of the mammalian heart explant. III. Tissue dehydration and cryoprotection by polyethylene glycol.* J Heart Lung Transplant, 1992. **11**(4 Pt 1): p. 619-23.

832. Meng, H., et al., *Tissue triage and freezing for models of skeletal muscle disease.* J Vis Exp, 2014(89).

833. Bostwick, B.L., et al., *Phenotypic and molecular characterisation of CDK13-related congenital heart defects, dysmorphic facial features and intellectual developmental disorders.* Genome Med, 2017. **9**(1): p. 73.

834. Hamilton, M.J., et al., *Heterozygous mutations affecting the protein kinase domain of CDK13 cause a syndromic form of developmental delay and intellectual disability.* J Med Genet, 2018. **55**(1): p. 28-38.

835. Uehara, T., et al., *Redefining the phenotypic spectrum of de novo heterozygous CDK13 variants: Three patients without cardiac defects.* Eur J Med Genet, 2017.

836. van den Akker, W.M.R., et al., *De novo variants in CDK13 associated with syndromic ID/DD; molecular and clinical delineation of 15 individuals and a further review.* Clin Genet, 2018.

837. Liang, S., et al., *CDK12: A Potent Target and Biomarker for Human Cancer Therapy.* Cells, 2020. **9**(6).

838. Ketley, A., et al., *CDK12 inhibition reduces abnormalities in cells from patients with myotonic dystrophy and in a mouse model.* Sci Transl Med, 2020. **12**(541).

839. Li, N., et al., *Expression and prognostic value of transcription-associated cyclin-dependent kinases in human breast cancer.* Aging (Albany NY), 2021. **13**.

840. Grunberg, S. and S. Hahn, *Structural insights into transcription initiation by RNA polymerase II.* Trends Biochem Sci, 2013. **38**(12): p. 603-11.

841. Larochelle, S., et al., *Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II.* Nat Struct Mol Biol, 2012. **19**(11): p. 1108-15.

842. Allen, B.L. and D.J. Taatjes, *The Mediator complex: a central integrator of transcription.* Nat Rev Mol Cell Biol, 2015. **16**(3): p. 155-66.

843. Calpena, E., et al., *De Novo Missense Substitutions in the Gene Encoding CDK8, a Regulator of the Mediator Complex, Cause a Syndromic Developmental Disorder.* Am J Hum Genet, 2019. **104**(4): p. 709-720.

844. Peterlin, B.M. and D.H. Price, *Controlling the elongation phase of transcription with P-TEFb.* Mol Cell, 2006. **23**(3): p. 297-305.

845. Ghamari, A., et al., *In vivo live imaging of RNA polymerase II transcription factories in primary cells.* Genes Dev, 2013. **27**(7): p. 767-77.

846. Blazek, D., et al., *The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes.* Genes Dev, 2011. **25**(20): p. 2158-72.

847. Bartkowiak, B. and A.L. Greenleaf, *Expression, purification, and identification of associated proteins of the full-length hCDK12/CyclinK complex.* J Biol Chem, 2015. **290**(3): p. 1786-95.

848. Greifenberg, A.K., et al., *Structural and Functional Analysis of the Cdk13/Cyclin K Complex.* Cell Rep, 2016. **14**(2): p. 320-31.

849. Young, R.A., *RNA polymerase II.* Annu Rev Biochem, 1991. **60**: p. 689-715.

850. Cho, E.J., et al., *Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain.* Genes Dev, 2001. **15**(24): p. 3319-29.

851. Bartkowiak, B., et al., *CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1.* Genes Dev, 2010. **24**(20): p. 2303-16.

852. Cheng, S.W., et al., *Interaction of cyclin-dependent kinase 12/CrkRS with cyclin K1 is required for the phosphorylation of the C-terminal domain of RNA polymerase II.* Mol Cell Biol, 2012. **32**(22): p. 4691-704.

853. Cheung, A.C. and P. Cramer, *A movie of RNA polymerase II transcription.* Cell, 2012. **149**(7): p. 1431-7.

854. Saunders, A., L.J. Core, and J.T. Lis, *Breaking barriers to transcription elongation.* Nat Rev Mol Cell Biol, 2006. **7**(8): p. 557-67.

855. Wade, J.T. and K. Struhl, *The transition from transcriptional initiation to elongation.* Curr Opin Genet Dev, 2008. **18**(2): p. 130-6.

856. Hieb, A.R., et al., *An 8 nt RNA triggers a rate-limiting shift of RNA polymerase II complexes into elongation.* EMBO J, 2006. **25**(13): p. 3100-9.

857. Eick, D. and M. Geyer, *The RNA polymerase II carboxy-terminal domain (CTD) code.* Chem Rev, 2013. **113**(11): p. 8456-90.

858. Adelman, K. and J.T. Lis, *Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans.* Nat Rev Genet, 2012. **13**(10): p. 720-31.

859. Buratowski, S., *Progression through the RNA polymerase II CTD cycle.* Mol Cell, 2009. **36**(4): p. 541-6.

860. Fuda, N.J., M.B. Ardehali, and J.T. Lis, *Defining mechanisms that regulate RNA polymerase II transcription in vivo.* Nature, 2009. **461**(7261): p. 186-92.

861. Munoz, M.J., M. de la Mata, and A.R. Kornblihtt, *The carboxy terminal domain of RNA polymerase II and alternative splicing.* Trends Biochem Sci, 2010. **35**(9): p. 497-504.

862. Czudnochowski, N., C.A. Bosken, and M. Geyer, *Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition.* Nat Commun, 2012. **3**: p. 842.

863. Davidson, L., L. Muniz, and S. West, *3' end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells.* Genes Dev, 2014. **28**(4): p. 342-56.

864. Ko, T.K., E. Kelly, and J. Pines, *CrkRS: a novel conserved Cdc2-related protein kinase that colocalises with SC35 speckles.* J Cell Sci, 2001. **114**(Pt 14): p. 2591-603.

865. Even, Y., et al., *CDK13, a Kinase Involved in Pre-mRNA Splicing, Is a Component of the Perinucleolar Compartment.* PLoS One, 2016. **11**(2): p. e0149184.

866. Bosken, C.A., et al., *The structure and substrate specificity of human Cdk12/Cyclin K.* Nat Commun, 2014. **5**: p. 3505.

867. Baumli, S., et al., *The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation.* EMBO J, 2008. **27**(13): p. 1907-18.

868. Chen, C., et al., *Identification of a major determinant for serine-threonine kinase phosphoacceptor specificity.* Mol Cell, 2014. **53**(1): p. 140-7.

869. Fan, Y., et al., *De Novo Mutations of CCNK Cause a Syndromic Neurodevelopmental Disorder with Distinctive Facial Dysmorphism.* Am J Hum Genet, 2018. **103**(3): p. 448-455.

870. Ghosh, G. and J.A. Adams, *Phosphorylation mechanism and structure of serine-arginine protein kinases.* FEBS J, 2011. **278**(4): p. 587-97.

871. Berro, R., et al., *CDK13, a new potential human immunodeficiency virus type 1 inhibitory factor regulating viral mRNA splicing.* J Virol, 2008. **82**(14): p. 7155-66.

872. Zhou, Z. and X.D. Fu, *Regulation of splicing by SR proteins and SR protein-specific kinases.* Chromosoma, 2013. **122**(3): p. 191-207.

873. Even, Y., et al., *CDC2L5, a Cdk-like kinase with RS domain, interacts with the ASF/SF2-associated protein p32 and affects splicing in vivo.* J Cell Biochem, 2006. **99**(3): p. 890-904.

874. Johnson, L.N. and R.J. Lewis, *Structural basis for control by phosphorylation.* Chem Rev, 2001. **101**(8): p. 2209-42.

875. Schulze-Gahmen, U., et al., *The AFF4 scaffold binds human P-TEFb adjacent to HIV Tat.* Elife, 2013. **2**: p. e00327.

876. Russo, A.A., P.D. Jeffrey, and N.P. Pavletich, *Structural basis of cyclin-dependent kinase activation by phosphorylation.* Nat Struct Biol, 1996. **3**(8): p. 696-700.

877. Liang, K., et al., *Characterization of human cyclin-dependent kinase 12 (CDK12) and CDK13 complexes in C-terminal domain phosphorylation, gene transcription, and RNA processing.* Mol Cell Biol, 2015. **35**(6): p. 928-38.
878. Pan, J., et al., *Establishment and characterization of a new human acute myelomonocytic leukemia cell line JIH-3.* Leuk Res, 2012. **36**(7): p. 889-94.
879. Winsor, T.S., et al., *A DNA damage response system associated with the phosphoCTD of elongating RNA polymerase II.* PLoS One, 2013. **8**(4): p. e60909.
880. Bennett, C.B., et al., *Yeast screens identify the RNA polymerase II CTD and SPT5 as relevant targets of BRCA1 interaction.* PLoS One, 2008. **3**(1): p. e1448.
881. Ekumi, K.M., et al., *Ovarian carcinoma CDK12 mutations misregulate expression of DNA repair genes via deficient formation and function of the Cdk12/CycK complex.* Nucleic Acids Res, 2015. **43**(5): p. 2575-89.
882. Bajrami, I., et al., *Genome-wide profiling of genetic synthetic lethality identifies CDK12 as a novel determinant of PARP1/2 inhibitor sensitivity.* Cancer Res, 2014. **74**(1): p. 287-97.
883. Joshi, P.M., et al., *Ovarian cancer-associated mutations disable catalytic activity of CDK12, a kinase that promotes homologous recombination repair and resistance to cisplatin and poly(ADP-ribose) polymerase inhibitors.* J Biol Chem, 2014. **289**(13): p. 9247-53.
884. Juan, H.C., et al., *Cdk12 is essential for embryonic development and the maintenance of genomic stability.* Cell Death Differ, 2016. **23**(6): p. 1038-48.
885. Novakova, M., et al., *Mouse Model of Congenital Heart Defects, Dysmorphic Facial Features and Intellectual Developmental Disorders as a Result of Non-functional CDK13.* Front Cell Dev Biol, 2019. **7**: p. 155.
886. Lizio, M., et al., *Gateways to the FANTOM5 promoter level mammalian expression atlas.* Genome Biol, 2015. **16**: p. 22.
887. Consortium, G.T., *Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans.* Science, 2015. **348**(6235): p. 648-60.
888. Cardoso-Moreira, M., et al., *Gene expression across mammalian organ development.* Nature, 2019. **571**(7766): p. 505-509.
889. Kaufmann, M.H., *The Atlas of Mouse Development.* 2001.
890. Carneiro, T.N., et al., *Utility of trio-based exome sequencing in the elucidation of the genetic basis of isolated syndromic intellectual disability: illustrative cases.* Appl Clin Genet, 2018. **11**: p. 93-98.
891. Hamilton, M.J. and M. Suri, *CDK13-related disorder.* Adv Genet, 2019. **103**: p. 163-182.
892. Basso, C., et al., *An echocardiographic survey of primary school children for bicuspid aortic valve.* Am J Cardiol, 2004. **93**(5): p. 661-3.
893. Larson, E.W. and W.D. Edwards, *Risk factors for aortic dissection: a necropsy study of 161 cases.* Am J Cardiol, 1984. **53**(6): p. 849-55.
894. Greenleaf, A.L., *Human CDK12 and CDK13, multi-tasking CTD kinases for the new millenium.* Transcription, 2019. **10**(2): p. 91-110.
895. Hornbeck, P.V., et al., *PhosphoSitePlus, 2014: mutations, PTMs and recalibrations.* Nucleic Acids Res, 2015. **43**(Database issue): p. D512-20.
896. Exome Aggregation Consortium, M.L., Konrad Karczewski, Eric Minikel, Kaitlin Samocha, Eric Banks, Timothy Fennell, Anne O'Donnell-Luria, James Ware, Andrew Hill, Beryl Cummings, Taru Tukiainen, Daniel Birnbaum, Jack Kosmicki, Laramie Duncan, Karol Estrada, Fengmei Zhao, James Zou, Emma Pierce-Hoffman, David Cooper, Mark DePristo, Ron Do, Jason Flannick, Menachem Fromer, Laura Gauthier, Jackie Goldstein, Namrata Gupta, Daniel Howrigan, Adam Kiezun, Mitja Kurki, Ami Levy Moonshine, Pradeep Natarajan, Lorena Orozco, Gina Peloso, Ryan Poplin, Manuel Rivas, Valentin Ruano-Rubio, Douglas Ruderfer, Khalid Shakir, Peter Stenson, Christine Stevens, Brett Thomas, Grace Tiao, Maria Tusie-Luna, Ben Weisburd, Hong-Hee Won, Dongmei Yu, David Altshuler, Diego Ardissino, Michael Boehnke, John Danesh, Elosua Roberto, Jose

Florez, Stacey Gabriel, Gad Getz, Christina Hultman, Sekar Kathiresan, Markku Laakso, Steven McCarroll, Mark McCarthy, Dermot McGovern, Ruth McPherson, Benjamin Neale, Aarno Palotie, Shaun Purcell, Danish Saleheen, Jeremiah Scharf, Pamela Sklar, Sullivan Patrick, Jaakko Tuomilehto, Hugh Watkins, James Wilson, Mark Daly, Daniel MacArthur, *Analysis of protein-coding genetic variation in 60,706 humans.* BioRxiv, 2015.

897. MacDonald, J.R., et al., *The Database of Genomic Variants: a curated collection of structural variation in the human genome.* Nucleic Acids Res, 2014. **42**(Database issue): p. D986-92.

898. Larivière, L., M. Seizl, and P. Cramer, *A structural perspective on Mediator function.* Curr Opin Cell Biol, 2012. **24**(3): p. 305-13.

899. Asadollahi, R., et al., *Genotype-phenotype evaluation of MED13L defects in the light of a novel truncating and a recurrent missense mutation.* Eur J Med Genet, 2017. **60**(9): p. 451-464.

900. Snijders Blok, L., et al., *De novo mutations in MED13, a component of the Mediator complex, are associated with a novel neurodevelopmental disorder.* Hum Genet, 2018. **137**(5): p. 375-388.

901. Dickinson, L.A., et al., *Cyclin L is an RS domain protein involved in pre-mRNA splicing.* J Biol Chem, 2002. **277**(28): p. 25465-73.

902. Loyer, P., et al., *Characterization of cyclin L1 and L2 interactions with CDK11 and splicing factors: influence of cyclin L isoforms on splice site selection.* J Biol Chem, 2008. **283**(12): p. 7721-32.

903. Bastian F PG, R.J., Moretti S, Laudet V, Robinson-Recavi M, *BGEE: Integrating and Comparing Hetrogeneous Transcriptome Data Amoung Species.* 2008.

904. !!! INVALID CITATION !!! {}.

905. Dai, Q., et al., *Cyclin K-containing kinase complexes maintain self-renewal in murine embryonic stem cells.* J Biol Chem, 2012. **287**(30): p. 25344-52.

906. Karl, T.R., *Atrioventricular septal defect with tetralogy of Fallot or double-outlet right ventricle: surgical considerations.* Semin Thorac Cardiovasc Surg, 1997. **9**(1): p. 26-34.

907. Raya, A., et al., *Notch activity acts as a sensor for extracellular calcium during vertebrate left-right determination.* Nature, 2004. **427**(6970): p. 121-8.

908. Fernandez, B., et al., *The coronary arteries of the C57BL/6 mouse strains: implications for comparison with mutant models.* J Anat, 2008. **212**(1): p. 12-8.

909. Al-Balool, H.H., et al., *Post-transcriptional exon shuffling events in humans can be evolutionarily conserved and abundant.* Genome Res, 2011. **21**(11): p. 1788-99.

910. Wang, Y., et al., *Dysregulated expression of microRNAs and mRNAs in myocardial infarction.* Am J Transl Res, 2015. **7**(11): p. 2291-304.

911. England, G., *The National Genomics Research and Healthcare Knowledgebase v5*. 2019.

912. Government, U., *Genome UK: the future of healthcare*, D.o.H.a.S. Care, Editor. 2020.

913. Alison Hall, T.F., Corinna Alberg, *Realising Genomics in Clinical Practice*, P. Foundation, Editor. 2014, PHG Foundation: Cambridge UK.

914. Cooke, S., Crawford, Gillian, Parker, Michael, Lucassen, Anneke and Hallowell, Nina, *Clinical Ethics, Recall of participation in research projects in cancer genetics: some implications for research ethics.* Clinical Ethics, 2008. **3**(4): p. 180-184.

915. Wolf, S.M., et al., *Managing incidental findings in human subjects research: analysis and recommendations.* J Law Med Ethics, 2008. **36**(2): p. 219-48, 211.

916. *The Creutzfeldt-Jakob Litigation* 1997.

917. Melham, J.K.P.B.J.D.V.H.G.N.H.C.H.K., *Ethical legal and social issues arrising from the use of GWAS in medical research.* 2009.

918. *Halushka v. University of Saskatchewan* 1965.

919. Gilissen, C., et al., *Genome sequencing identifies major causes of severe intellectual disability.* Nature, 2014. **511**(7509): p. 344-7.

920. Taylor, J.C., et al., *Factors influencing success of clinical genome sequencing across a broad spectrum of disorders.* Nat Genet, 2015. **47**(7): p. 717-726.

921. Schuol, S., et al., *So rare we need to hunt for them: reframing the ethical debate on incidental findings.* Genome Med, 2015. **7**(1): p. 83.

922. Hehir-Kwa, J.Y., et al., *Towards a European consensus for reporting incidental findings during clinical NGS testing.* Eur J Hum Genet, 2015. **23**(12): p. 1601-6.

923. Hu, P., et al., *Clinical application of targeted next-generation sequencing in fetuses with congenital heart defect.* Ultrasound Obstet Gynecol, 2018. **52**(2): p. 205-211.

924. Mayer, A.N., et al., *A timely arrival for genomic medicine.* Genet Med, 2011. **13**(3): p. 195-6.

925. Ormond, K.E., et al., *Challenges in the clinical application of whole-genome sequencing.* Lancet, 2010. **375**(9727): p. 1749-51.

926. Berg, J.S., M.J. Khoury, and J.P. Evans, *Deploying whole genome sequencing in clinical practice and public health: meeting the challenge one bin at a time.* Genet Med, 2011. **13**(6): p. 499-504.

927. Green, R.C., et al., *ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing.* Genet Med, 2013. **15**(7): p. 565-74.

928. *ACMG policy statement: updated recommendations regarding analysis and reporting of secondary findings in clinical genome-scale sequencing.* Genet Med, 2015. **17**(1): p. 68-9.

929. van El, C.G., et al., *Whole-genome sequencing in health care: recommendations of the European Society of Human Genetics.* Eur J Hum Genet, 2013. **21**(6): p. 580-4.

930. Rigter, T., et al., *Reflecting on earlier experiences with unsolicited findings: points to consider for next-generation sequencing and informed consent in diagnostics.* Hum Mutat, 2013. **34**(10): p. 1322-8.

931. Alison Hall, N.H., Ron Zimmern, *Managing Incidental and pertinent findings from WGS in the 100,000 Genome Project, PHG foundation.* 2013.

932. Lango Allen, H., et al., *GATA6 haploinsufficiency causes pancreatic agenesis in humans.* Nat Genet, 2012. **44**(1): p. 20-2.

933. Maheshwari, M., et al., *PTPN11 mutations in Noonan syndrome type I: detection of recurrent mutations in exons 3 and 13.* Hum Mutat, 2002. **20**(4): p. 298-304.

934. Musante, L., et al., *Spectrum of mutations in PTPN11 and genotype-phenotype correlation in 96 patients with Noonan syndrome and five patients with cardio-facio-cutaneous syndrome.* Eur J Hum Genet, 2003. **11**(2): p. 201-6.

935. Dorschner, M.O., et al., *Actionable, pathogenic incidental findings in 1,000 participants' exomes.* Am J Hum Genet, 2013. **93**(4): p. 631-40.

936. Boone, P.M., et al., *Incidental copy-number variants identified by routine genome testing in a clinical population.* Genet Med, 2013. **15**(1): p. 45-54.

937. Dheensa, S., et al., *Health-care professionals' responsibility to patients' relatives in genetic medicine: a systematic review and synthesis of empirical research.* Genet Med, 2016. **18**(4): p. 290-301.

938. Godlee, F., *Our commitment is to patient partnership.* BMJ, 2017. **356**: p. j939.

939. Mitchell, C., et al., *Exploring the potential duty of care in clinical genomics under UK law.* Med Law Int, 2017. **17**(3): p. 158-182.

940. Epstein, J.A., H. Aghajanian, and M.K. Singh, *Semaphorin signaling in cardiovascular development.* Cell Metab, 2015. **21**(2): p. 163-173.

941. Kaufman, M.H. and V. Navaratnam, *Early differentiation of the heart in mouse embryos.* J Anat, 1981. **133**(Pt 2): p. 235-46.

942. de Boer, B.A., et al., *Growth of the developing mouse heart: an interactive qualitative and quantitative 3D atlas.* Dev Biol, 2012. **368**(2): p. 203-13.

943. Hiruma, T., Y. Nakajima, and H. Nakamura, *Development of pharyngeal arch arteries in early mouse embryo.* J Anat, 2002. **201**(1): p. 15-29.

944. Kisanuki, Y.Y., et al., *Tie2-Cre transgenic mice: a new model for endothelial cell-lineage analysis in vivo.* Dev Biol, 2001. **230**(2): p. 230-42.

945. Mjaatvedt, C.H. and R.R. Markwald, *Induction of an epithelial-mesenchymal transition by an in vivo adheron-like complex.* Dev Biol, 1989. **136**(1): p. 118-28.

946. Lincoln, J., C.M. Alfieri, and K.E. Yutzey, *Development of heart valve leaflets and supporting apparatus in chicken and mouse embryos.* Dev Dyn, 2004. **230**(2): p. 239-50.

947. Kaufman, M.H. and L. Richardson, *3D reconstruction of the vessels that enter the right atrium of the mouse heart at Theiler Stage 20.* Clin Anat, 2005. **18**(1): p. 27-38.

948. Laboratory, T.J., *Inbred Strains of Mice, No. 3.* 1963: Bar Harbor, Maine. . p. 104.

949. Smith, C.L., C.A. Goldsmith, and J.T. Eppig, *The Mammalian Phenotype Ontology as a tool for annotating, analyzing and comparing phenotypic information.* Genome Biol, 2005. **6**(1): p. R7.