Investigation of Copy Number Variation in South African Patients with Congenital Heart Defects

Nicole Aimee Saacks

SCKNIC002

Dissertation Presented for the Degree of

Master of Science in Medicine

Faculty of Health Sciences

UNIVERSITY OF CAPE TOWN

September 2020

Supervisor: Professor Liesl Zühlke

Department of Paediatrics and Child Health,

University of Cape Town

Declaration

I, Nicole Aimee Saacks hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

Date: 29 September 2020

Acknowledgements

My sincerest thanks and appreciation to the following persons:

- My incredible supervisor, Professor Liesl Zühlke, for the opportunity to work with and learn from her, for encouraging my research and teaching career, for her continuous support, expert guidance, and leadership. Thank you for always believing in me and teaching me research and writing skills beyond my studies. You are an inspiration and 'Wonder Woman' to all who know and are privileged to work with you.

- My mentor, Dr. James Eales, for training me in bioinformatics at the University of Manchester, for assisting me with the various analyses performed, for his patience, guidance, and coding expertise throughout this study. Thank you for your continuous support and for sharing your passion for computational biology and ggplot2 with me.

- My mentor and friend, Timothy Spracklen, for his invaluable expert genetic knowledge, guidance, and investment in training me throughout this project. Thank you for encouraging and supporting me every step of the way, this dissertation is a product of your input in many ways.

- Professor Bongani Mayosi (and in loving memory) for making this project possible, through his DGEMBE partnership with Professor Bernard Keavney at the University of Manchester.

- Professor Bernard Keavney, for the opportunity to work and train in his laboratory at the University of Manchester, for his cardiovascular genetics expertise, shared wisdom, and continuous support. Thank you for showing me the importance of CHD genetics research and keeping the collaboration between the University of Cape Town and the University of Manchester strong.

- Dr. Simon Williams, for teaching me invaluable bioinformatic skills at the University of Manchester, for assisting and supporting me throughout the study.

- Professor Mark Engel, for encouraging my research career, and for his expert guidance, advice, and support throughout the project.

- Dr. Thomas Aldersley and the entire PROTEA and CHDRU team members, for their tireless commitment to children's heart disease over the years, their continuous support, encouragement, and companionship.

- Associate Professor Gasnat Shaboodien for the opportunity to work and train at the Hatter Institute for Cardiovascular Research in Africa, for the long meetings and further development of my research skills.

- All the members of the Cardiovascular Genetics and Cardio Protection teams at the Hatter Institute for Cardiovascular Research in Africa, for their friendship, positivity, and assistance throughout this project. Thank you for making my time spent in the laboratory enjoyable.

- This work is based on the research supported, in part, by the National Research Foundation of South Africa (Grant number 98563). Any finding, conclusion, or recommendation expressed in this material is that of the author(s) and the NRF does not accept any liability in this regard. The financial assistance of the Medical Research Council United Kingdom is hereby acknowledged.

- This degree would not be possible without the unrelenting support, love, and encouragement of my amazing family and wonderful friends, especially my incredible mom who is always there for me, providing support, advice, and many cups of tea. Thank you for being my anchor and constantly reminding me that I can do anything and everything I set my mind to.

"If I have seen further, it is by standing on the shoulders of giants"

Isaac Newton

This dissertation is dedicated to Brave Little Heart, Thaakirah Matthews, and her wonderful family, as well as all the heart warriors and heart moms I have met, learned from and spent time with throughout this project. Without our CHD patients, this research would not be possible. Thank you for telling your stories and inspiring me with your bravery, courage, and resilience. *"Sometimes superheroes reside in the hearts of small children fighting big battles." – Unknown*

Finally, I dedicate this research to positive change in a world of opportunity.

Investigation of copy number variation in South African Patients with Congenital Heart Defects
Abstract

Background: Congenital heart disease (CHD) is the leading non-infectious cause of paediatric morbidity and mortality worldwide and a significant social and healthcare burden. The aetiology of CHD is poorly understood, though heritable genetic factors including copy number variants (CNVs) have been shown to contribute to the risk of CHD in individuals of European ancestry. However, the role of rare CNVs in the development of CHD in African populations including South Africa is unknown. This study aims to identify pathogenic and likely pathogenic CNVs in South African cases of CHD. To our knowledge, this is the first study to investigate the genetic basis of CHD in a South African cohort.

Methods: The study cohort included 105 patients presenting to the cardiac clinics at Red Cross War Memorial Children's Hospital and Groote Schuur Hospital with non-syndromic isolated CHD (n = 76), non-syndromic CHD with additional extra-cardiac anomalies (n = 17), and positive controls with syndromic CHD (n = 12). Genotyping was performed using the Affymetrix CytoScan HD platform. Rare CNVs were filtered using stringent criteria for their size and algorithm-specific quality score and were compared against a gene panel of known CHD-associated genes. Candidate genes were considered based on pLI scores and reported CHD phenotypes in mouse models. The identified CNVs were validated by quantifying the read-coverage of available whole-exome sequencing data of a similar overlapping cohort.

Results: Chromosomal microarray analysis was successful for 101 participants (including 89 non-syndromic CHD cases and 12 control cases) and led to the identification of eight CNVs overlapping genes known to be causal for CHD (*GATA4, TBX1, FLT4, CRKL, NSD1,* and *B3GAT3*), and four CNVs encompassing candidate genes likely to play a role in the development of CHD (*DGCR8, JARID2, KDM2A,* and *FSTL1)*. The CNVs were identified in nine unrelated individuals: five of the CNVs were classified as pathogenic or likely pathogenic (5.6% of the cohort) and four were classified as variants of unknown significance (4.6%). CNVs of interest were validated using the available whole-exome sequencing data.

Conclusions: In this study, we show that chromosomal microarray analysis is an effective technique for identifying CNVs in patients diagnosed with CHD and that this approach can be performed locally in South Africa, producing results similar to those seen in international CHD studies. The findings of this thesis highlight the wide genetic heterogeneity of CHD and the growing importance of CHD genetic studies for both research and clinical purposes. Advancing our understanding of CHD aetiology will help define disease risk in South Africa and improve the way we care for and assess our cardiac patients.

Table of Contents

List of Figures

Chapter 1. Introduction and Review of the Literature

Chapter 2. Methods and Materials

Chapter 3. Results

Appendices

List of Tables

Chapter 1. Introduction and Review of the Literature

Chapter 2. Methods and Materials

Chapter 3. Results

Appendices

List of Abbreviations and Acronyms

A	Aorta
ACMG	The American College of Medical Genetics
AD	Autosomal dominant
ADHD	Attention-deficit hyperactivity disorder
AV	Atrioventricular
APT	Affymetrix power tools
AR	Autosomal recessive
BA	Black African
ChAS	Chromosome analysis suite
CGH	Comparative genomic hybridisation
CHD	Congenital heart disease
CMA	Chromosomal microarray
CN	Copy number
CNVs	Copy number variants
CPGR	Centre for Proteomic and Genomic Research in Cape Town
CVD	Cardiovascular disease
DECIPHER	Database of genomic variation and phenotype in humans using Ensembl resource
DGV	Database of Genomic Variants
DNA	Deoxyribonucleic acid
ECAs	Extra-cardiac anomalies
EDTA	Ethylenediaminetetraacetic acid
ExAC	Exome Aggregation Consortium
FISH	Fluorescent in situ hybridisation
gnomAD	Genome Aggregation Database
GSH	Groote Schuur Hospital
GXD	Gene Expression Database
HICs	Higher-income countries
HIV	Human immunodeficiency virus
HOCM	Hypertrophic obstructive cardiomyopathy
HPCSA	Health Professions Council of South Africa
HREC	Human Ethics Research Committee
LA	Left atrium

LCRs	Low copy repeat regions
LLMICs	Low- and lower-middle-income countries
LoF	Loss of function
LV	Left ventricle
MA	Mixed ancestry
MAPD	Median Absolute Pairwise Difference
MGI	Mouse Genome Informatics
MLPA	Multiplex Probe Ligation Amplification
NAHR	Non-allelic homologous recombination
NCBI	National Center for Biotechnology Information
NCDs	Non-communicable diseases
NGS	Next-generation sequencing
OMIM	Online Mendelian Inheritance in Man
P	Pulmonary artery
PCR	Polymerase chain reaction
pLI	Probability of loss-of-function intolerance
PROTEA	Partnerships for Congenital Heart Disease in Africa
QC	Quality control
RA	Right atrium
RACHS-1	Risk Adjustment for Congenital Heart Surgery
RCWMCH	Red Cross War Memorial Children's Hospital
RV	Right Ventricle
RVH	Right ventricular hypertrophy
SD	Standard deviation
SNP	Single nucleotide polymorphism
SNPQC	SNP Quality Control
SNVs	Single nucleotide variants
SSA	Sub-Saharan Africa
T21	Trisomy 21 or Down syndrome
TBE	Tris-Borate-EDTA
TS	Turner syndrome
UCT	University of Cape Town
VOUS	Variants of uncertain clinical significance
WES	Whole-exome sequencing
WGS	Whole-genome sequencing

Measurements

°C	Degrees Celsius
A_{230}	Absorbance at 230nm
A_{260}	Absorbance at 260nm
A_{280}	Absorbance at 280nm
bp	Base pairs
kb	Kilobase
Mb	Mega-bases
ml	Millilitres
mM	Millimolar
ng	Nanogram
nm	Nanometre
V	Volts
w/v	Percentage weight per volume
µg	Microgram
µl	Microlitres

Nucleic Acid Codes

A	Adenine
C	Cytosine
G	Guanine
T	Thymine

List of CHD-subtypes Mentioned in this Thesis

AS	Aortic stenosis
ASD	Atrial septal defect
AVSD	Atrioventricular septal defect
BAV	Bicuspid aortic valve
BPV	Bicuspid pulmonary valve
CoA	Coarctation of the aorta
DORV	Double outlet right ventricle
HLHS	Hypoplastic left heart syndrome
IAA	Interrupted aortic arch
LVOTO	Left ventricular outflow tract obstructions
MA	Mitral atresia
MS	Mitral stenosis
PDA	Patent ductus arteriosus
PA	Pulmonary atresia
PPS	Peripheral pulmonary stenosis
PS	Pulmonary valve stenosis
RVOTO	Right ventricular outflow tract obstructions
SVAS	Supravalvular aortic stenosis
TA	Tricuspid atresia
TAPVR	Total anomalous pulmonary venous return
TAPVD	Total anomalous pulmonary venous drainage
TGA	Transposition of the great arteries
TOF	Tetralogy of Fallot
TrA	Truncus arteriosus
VSD	Ventricular septal defect

Chapter 1. Introduction and Review of the Literature

Introduction

Congenital heart disease (CHD) is the most prevalent birth defect and the leading non-infectious cause of morbidity and paediatric mortality worldwide (Cai et al., 2018; Lyu et al., 2018). Approximately 9 per 1000 children are born with CHD (Liu et al., 2019), with an estimated 11000 South African children born with this condition each year (Hoosen et al., 2011). CHD is defined as a structural malformation of the heart and/or great blood vessels that occurs before birth (Liu et al., 2019). The disease develops as a result of perturbation in normal cardiac development (Zaidi & Brueckner, 2017), and can range from asymptomatic to life-threatening depending on the severity and complexity of the cardiac lesion(s). Tremendous strides in treatment, management, and cardiothoracic surgery have led to an increased survival rate of children born with CHD, and consequently a growing adult CHD population (Van Der Linde et al., 2011). The increasing adult CHD population poses new scientific questions and challenges, including the genetic risks for the offspring of individuals with CHD. As a result, there is an increasing need for the incorporation of medical genetics in CHD management, with possible implications on diagnosis, recurrence risk, and family screening (Parrott & Ware, 2012).

Despite the many advances in diagnosis and treatment of CHD, our understanding of the causes underlying disease pathogenesis is relatively poor, though the role of genetic mutations and chromosomal rearrangement has been demonstrated (Pierpont et al., 2018). CHD is a complex genetic disorder with a suggested polygenic susceptibility and elevated recurrence risk amongst first-degree relatives. In families with an index case with CHD, the risk of the same CHD phenotype amongst siblings increases between 3-fold and 80-fold depending on the type of CHD, while the risk of another type of CHD increases 2- to 3-fold (Fotiou et al., 2019; Zaidi & Brueckner, 2017). Advanced genetic approaches including whole-exome sequencing (WES) and chromosomal microarrays (CMAs) have led to the identification of numerous rare single nucleotide variants (SNVs) and copy number variants (CNVs) associated with CHD (Fotiou et al., 2019; Page et al., 2019; Soemedi et al., 2012a). Most major studies have focused on individuals of European ancestry, and evidence-based research on the causes of disease in the African population is scarce. Importantly, the genetic basis of CHD in the South African context has been minimally explored. Advancing our understanding of the possible causes of CHD will help define disease risk in South Africa, improve the way we care for and treat individuals with CHD, and facilitate prevention (Wang et al., 2014). This thesis is part of a pilot study which focuses on exploring the genetic contribution to CHD in South Africa, by specifically investigating the role of copy number variation in CHD pathogenesis utilising the Affymetrix CytoScan HD array platform.

1.1. Structure and function of the heart

The human heart is a muscular organ that is approximately the size of a closed fist. As the heart contracts, it pumps blood throughout the body through a network of blood vessels called the circulatory system, providing the tissues with nutrients and oxygen, and removing waste. As shown in Figure 1.1, the heart is comprised of two parallel circulation systems: deoxygenated blood enters the right atrium, flows through the right ventricle to the lungs where it is oxygenated, while oxygenated blood enters the left atrium via the pulmonary vein, flows through to the left ventricle and is pumped into the systemic circulation (Fox, 2013).

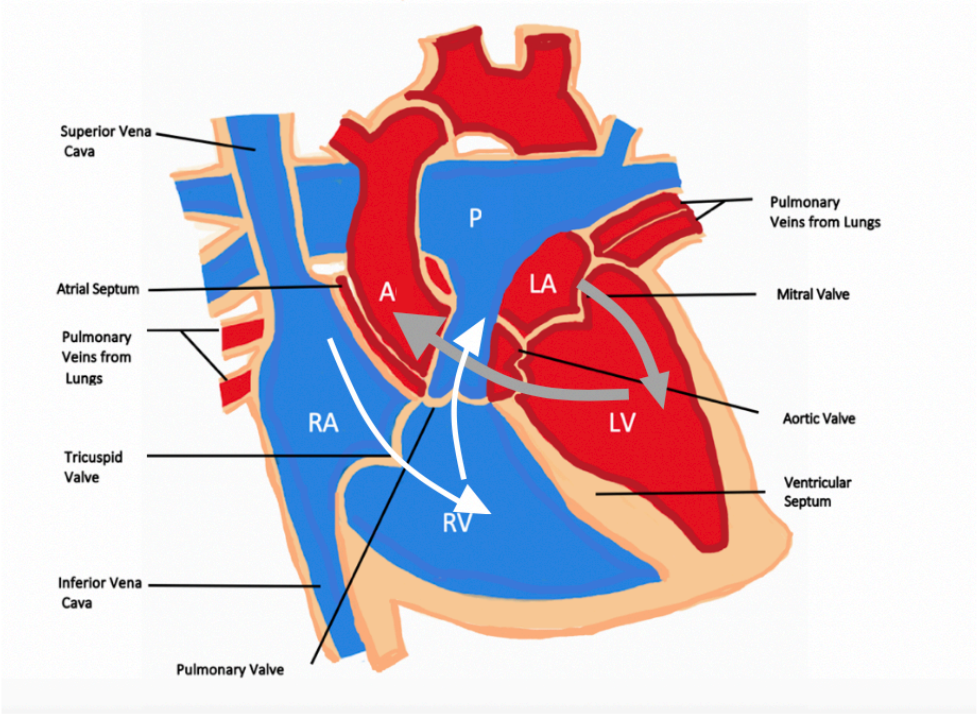

Figure 1.1. Blood flow through the human heart
The blue shading represents deoxygenated blood, which is pumped to the lungs to be oxygenated. The red shading represents oxygenated blood received from the lungs, which is pumped into systemic circulation. Direction of blood flow is indicated by white arrows (deoxygenated blood) and grey arrows (oxygenated blood). Acronyms: A – aorta; LA - left atrium; LV - left ventricle; P - pulmonary artery; RA - right atrium; RV - right ventricle. (Illustration by Nicole Saacks).

Unidirectional blood flow, occurring in parallel, is critical for optimal functionality of the heart; however, in some cases, the heart or blood vessels fail to develop properly during embryogenesis. This may result in abnormal blood flow through the heart, blockages that prevent adequate blood flow, or underdeveloped parts of the heart itself. Heart defects can arise as the various components of the heart are developing during gestation, resulting in CHD, the most common birth defect in the world (Pierpont et al., 2018).

1.2. Congenital heart defects

The heart is the first organ to develop in the human embryo through a complex series of events reviewed in more detail elsewhere (Muñoz-Chápuli & Pérez-Pomares, 2009). Any disruption during cardiogenesis can result in a cardiac defect. CHD is an umbrella term for a spectrum of cardiac anomalies of differing incidence and severity, the most common of which are shown in Figure 1.2.

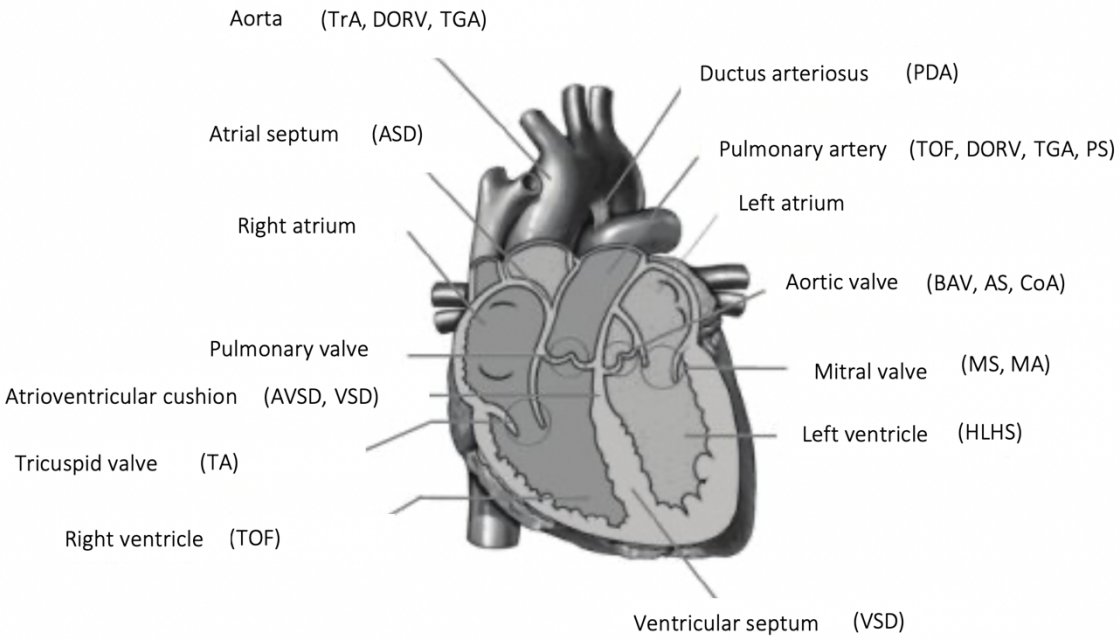

Figure 1.2. The heart structures commonly affected by congenital heart disease
Acronyms: AS – aortic stenosis; ASD – atrial septal defect; AVSD – atrioventricular septal defect; BAV – bicuspid aortic valve; CoA – coarctation of the aorta; DORV – double outlet right ventricle; HLHS – hypoplastic left heart syndrome; MA – mitral atresia; MS – mitral stenosis; PDA – patent ductus arteriosus; PS – pulmonary valve stenosis; TA – tricuspid atresia; TrA – truncus arteriosus; TGA – transpositions of the great arteries; TOF – tetralogy of Fallot; VSD – ventricular septal defect (Adapted from Ghorbel, Angelini & Caputo, 2012).

The diversity of cardiac phenotypes implicated in CHD has resulted in multiple classification systems based on anatomical, functional, and clinical features (Costain, Silversides, & Bassett, 2016). One such classification system is the Risk Adjustment for Congenital Heart Surgery (RACHS-1) method. The RACHS-1 method was developed by a panel of eleven paediatric cardiologists and surgeons to allow comparisons of outcomes for CHD patients undergoing cardiac surgery (Jenkins et al., 2002).

The RACHS-1 method groups cardiac procedures with similar mortality risks into six categories based on parameters such as age, diagnosis, prematurity, and surgical procedure codes (Jenkins, 2004). RACHS-1 scores have been widely used amongst paediatric cardiologists, and have proved to be a useful tool for assessing mortality risk (Cavalcante et al., 2016).

In 2002, Hoffman and Kaplan classified CHD into three categories, severe, moderate, and mild, based on the severity of the lesions (Table 1.1). Severe CHD comprises all cyanotic heart defects, with most infants presenting as severely ill. Moderate CHD cases require expert care, while mild CHD is the most prevalent category comprising of asymptomatic CHD and self-correcting lesions. A second method proposed by Botto, Lin, Riehle-Colarusso, Malik, and Correa (2007), classified CHD in terms of cardiac complexity, cardiac phenotype, and extra-cardiac anomalies into the following main categories: conotruncal defects, left ventricular outflow tract obstructions (LVOTO), right ventricular outflow tract obstructions (RVOTO), septal defects, atrioventricular defects (AVSD), heterotaxy, total anomalous pulmonary venous return (TAPVR), and complex cases. Multiple defects can occur together resulting in a more complex CHD such as tetralogy of Fallot (TOF). TOF is a tetrad of four cardiac anomalies including a ventricular septal defect (VSD), right ventricular hypertrophy, pulmonary valve stenosis (PS), and an overriding aorta (Bailliard & Anderson, 2009). While numerous CHD classifications systems exist, the most widely used are shown in Table 1.1 (Botto et al., 2007; Hoffman & Kaplan, 2002). The broad range of cardiac lesions and underlying developmental mechanisms has made aetiological and epidemiological studies of CHD a great challenge.

Table 1.1. Classification of CHD subtypes based on Hoffman and Kaplan (2002) and Botto et al. (2007) *

CHD subtypes	Classification by Botto et al. (2007)	Classification by Hoffman and Kaplan (2002)
D-transpositions of the great arteries	Conotruncal defect	Severe
Tetralogy of Fallot	Conotruncal defect	
Double outlet right ventricle	Conotruncal defect	
Interrupted aortic arch	Conotruncal defect	
Truncus arteriosus	Conotruncal defect	
Hypoplastic left heart syndrome	LVOTO	
Total anomalous pulmonary venous connection	TAPVR	
Right heart lesions (Tricuspid atresia, Pulmonary atresia, Ebstein anomaly)	RVOTO	
Single ventricle	Complex	
Atrioventricular septal defect	AVSD	
Coarctation of aorta	LVOTO	
Critical or severe aortic stenosis	LVOTO	
Severe pulmonary stenosis	RVOTO	
Large ventricular septal defect	Septal defect	
Large patent ductus arteriosus	None	
Mild or moderate aortic stenosis	LVOTO	Moderate
Moderate pulmonary stenosis	RVOTO	
Noncritical coarctation of the aorta	LVOTO	
Large atrial septal defect	Septal defect	
Complex forms of ventricular septal defect	Septal defect	
Small ventricular septal defect	Septal defect	Mild
Small patent ductus arteriosus	None	
Mild pulmonary stenosis	RVOTO	
Bicuspid aortic valve without aortic stenosis	None	
Small or spontaneously closed atrial septal defect	Septal defect	

*Adapted from (Botto et al., 2007; Hoffman & Kaplan, 2002)

Note. Acronyms: AVSD – atrioventricular septal defects; CHD – congenital heart disease; LVOTO – left ventricular outflow tract obstruction; RVOTO – right ventricular outflow tract obstruction; TAPVR – total anomalous pulmonary venous return.

1.3. Global burden of disease

Cardiovascular disease (CVD) is the leading cause of death worldwide, claiming approximately 17.8 million lives each year (Mensah, Roth, & Fuster, 2019). Non-communicable diseases (NCDs) including CVD are becoming an increasingly important healthcare concern worldwide due to their chronic and progressive nature, and the widespread effect of urbanisation seen in developing countries over the past century (Fuster, 2014). According to the World Heart Federation (2017), up to 80% of NCD-related deaths occur in low- and lower-middle-income countries (LLMICs), where detection of CVDs and cardiac care is not readily available (World Heart Federation, 2017). Though the majority of CVD-NCD deaths are atherosclerotic and hypertensive, CHD remains an important source of disease burden (Mensah et al., 2019). CHD is a major cause of cardiovascular morbidity and mortality in children, and a significant health burden worldwide (Lander & Ware, 2014). A systematic literature review by al. (2019) reported a substantial increase in CHD birth prevalence from ~4 per 1000 live births in 1970 to the estimated 9 per 1000 live births seen today, largely due to improved screening and diagnosis of minor defects. This corresponds to about 1.5 million babies being born with CHD each year, representing a significant burden in worldwide public healthcare (Liu et al., 2019). Notably, the distribution of the various CHD subtypes is fairly uniform across different countries, and the global prevalence rates of these subtypes have remained the same from 1990 to 2017, except for septal defects which have decreased as a result of an increased proportion of defects being surgically corrected (Zimmerman et al., 2020).

Despite the significant burden of CHD, the prognosis for children born with CHD in developed countries has drastically improved over the last fifty years, with over 90% of affected individuals surviving to adulthood (Zühlke et al., 2019). This improved outlook can be attributed to the increased availability of prenatal echocardiography and improved medical and surgical care for patients living in higher-income countries (HICs) (Liu et al., 2019). With the increased prevalence of CHD, the issues of long-term prognosis, and the underlying causes of CHD are gaining more global attention (Van Der Linde et al., 2011).

1.3.1. African burden of congenital heart disease

The socio-economic burden of CVD, including CHD, falls heavily on developing LLMICs, many of which are situated in sub-Saharan Africa (SSA) (Markbreiter & Philippa, 2016; Mensah et al., 2019). NCD including CHD is a rising epidemic in SSA, predicted to surpass infectious diseases such as human immunodeficiency virus (HIV) and tuberculosis as the leading cause of death over the next decade

(Keates, Mocumbi, Ntsekhe, Sliwa, & Stewart, 2017). The majority of children born with CHD live in LLMICs and face a starkly different prognosis to children living with CHD in HICs (Zühlke et al., 2019). There are several reasons for these regional differences. Limited antenatal screening is available to patients in LLMICs, and as a result, few cases of CHD are diagnosed before birth leading to an increased rate of mortality during the infancy period. Furthermore, children born with CHD in LLMICs have inadequate access to life-saving surgeries, intervention, and cardiac care (Zühlke et al., 2019). Paediatric cardiac services are an extremely expensive area of medicine and are not readily available to the majority of patients living in SSA. In addition, health infrastructure and education are insufficient, and with the demands of HIV/AIDs, paediatric cancer, and malnutrition on the health budget, children living in SSA with CHD are often neglected (Hewitson, Brink, & Zilla, 2002). Another significant aspect is the underestimated birth prevalence of CHD in Africa. The estimated prevalence of 9 per 1000 live births is generally accepted worldwide, with genetic, environmental, and epigenetic factors accounting for the variation seen between regions (Zühlke, Mirabel, & Marijon, 2013). However, a recent systematic review by Liu et al. (2019) reported a significantly lower prevalence rate in Africa (2.315 per 1000 live births). Liu et al. (2019) suggest that this disparity in prevalence between Africa and other continents reflects severe resource constraints and limited access to healthcare, leading to a low detection rate and a paucity of available estimates from Africa.

The challenges that African countries face in caring for patients with CHD are aggravated by the lack of understanding of the basic epidemiology of individuals living with CHD in Africa. Few clinical epidemiology studies on the prevalence of CHD phenotypes in SSA have been published, which show a variable prevalence of cardiac defects with TOF, ASD, and VSD occurring frequently (De Decker et al., 2016; Teteli et al., 2014; Thomford et al., 2018; Wonkam et al., 2017). However, further research including African-based genetic studies is required to help us better understand the African CHD epidemiology data obtained (Thomford et al., 2018). Understanding the extent of the CHD burden in SSA is the first step toward the development of preventative and treatment strategies for CHD on the African continent.

1.4. Available treatment for congenital heart disease

Advances in treatment, management, and surgery have improved the prognosis and increased the lifespan of individuals with CHD (Van Der Linde et al., 2011). Various treatment options are available to individuals with CHD depending on the type and severity of their cardiac defects. Not all CHD patients require medical treatment and certain defects, such as small ASDs and VSDs, often correct themselves with age. Patients with more complex defects often require treatment early in their lives.

This treatment can range from medications such as angiotensin-converting enzyme inhibitors to cardiac catheterization, cardiothoracic surgery, and even heart transplants (American Heart Association, 2018). The advanced surgical techniques available to individuals diagnosed with CHD have enabled many patients to undergo corrective, reparative, or palliative heart surgery (Yuan & Jing, 2009). Furthermore, advances in cardiac surgery over the last few decades have led to improved surgical outcomes for CHD patients with genetic syndromes, who were often not operated on in the past due to the high risks of postoperative mortality (Robinson & Newburger, 2003). Although advances in cardiac care have increased the survival rate for children born with CHD, the underlying aetiology of CHD is poorly understood (Lander & Ware, 2014).

1.5. The aetiology of congenital heart disease

The aetiology of CHD has been the focus of numerous studies over the past decades (Baban et al., 2014; Cai et al., 2018; Mlynarski et al., 2016; Zhu, Kartiko, & Finnell, 2009). Though our understanding of the molecular pathways involved in heart development has improved greatly over the years, the underlying causes of approximately 75% to 80% of CHD cases remain unclear (Bassili et al., 2000; Blue, Kirk, Sholler, Harvey, & Winlaw, 2012; Tomita-Mitchell et al., 2012). It has long been appreciated that environmental, genetic, and epigenetic factors can cause CHD, often in the context of a multifactorial disease (Cai et al., 2018). The interaction between these risk factors is thought to increase susceptibility to the development of a heart defect (Wang et al., 2014). With more children born with CHD surviving to adulthood and starting families of their own, improving our understanding of disease aetiology and recurrence risks has become critical. Importantly, understanding the causes of CHD will help clinicians determine the prognostic outcome for surgery or treatments, and identify patients at higher risk of operative morbidity and mortality (Blue et al., 2012).

1.5.1. Environmental risk factors for congenital heart disease

Environmental risk factors for CHD include any non-genetic factors that have been associated with the risk of developing a cardiac defect, many of which occur *in utero* (Blue et al., 2012). Environmental factors that have been associated with CHD include maternal exposure to cigarette smoke, alcohol, thalidomide, isotretinoin and antiseizure medication (Zhu et al., 2009), infectious agents such as rubella (Dewan & Gupta, 2012), and teratogens such as dioxins and pesticides (Kopf & Walker, 2009). Novel non-genetic causes and risk factors for CHD are continuously arising despite efforts to minimize these modifiable influences (Fahed, Gelb, Seidman, & Seidman, 2013). Certain modifiable risk factors

for CHD such as folate deficiencies and air pollution are likely to significantly affect patients in LLMICs, including South Africa, where exposure to these risk factors is at higher levels when compared to HICs. Although data regarding these factors could improve public health priorities worldwide, the impact of these risk factors on the incidence of CHD in SSA has been minimally explored (Darnton-Hill & Mkparu, 2015; Zhang et al., 2016).

1.5.2. Heritable risk factors for congenital heart disease

Epidemiology of CHD suggests that the majority of cases have a genetic basis, though the genetic underpinnings of CHD remain unclear. There is a multitude of evidence that supports the role of genetics in CHD, including population-based studies, twin studies, and the recurrence risk between 2- and 80-fold for first-degree relatives of CHD patients (Zaidi & Brueckner, 2017). Population-based studies have revealed an elevated incidence of certain CHD subtypes such as septal defects in consanguineous populations, suggesting a recessive genetic contribution to the development of CHD (Bassili et al., 2000; Becker, Halees, Molina, & Paterson, 2001; Shieh, Bittles, & Hudgins, 2012). Wang et al. (2014) and Øyen et al. (2009, 2010) found an increased risk of recurrence of both similar and discordant forms of CHD amongst relatives when compared to the general population. The elevated recurrent risk in consanguineous populations and relatives, who share a genetic background emphasises the genetic contribution to CHD pathogenesis.

1.5.3. Genetics of congenital heart disease

CHD is a complex heterogeneous genetic disorder associated with both familial and sporadic inheritance patterns (Figure 1.3). Familial CHD mutations can be inherited in an autosomal dominant, autosomal recessive, or X-linked manner, and can manifest in a variety of clinical phenotypes (Chung & Rajakumar, 2016). Advances in genetic technology such as WES and CMAs have led to the identification of numerous defective genes implicated in CHD. It has been postulated that several hundred genes may be involved in disease pathogenesis with many still to be discovered (Aburawi, Aburawi, Bagnall, & Bhuiyan, 2015; Fotiou et al., 2019). Variants across the frequency spectrum can affect CHD, from common variants with low but cumulative impacts, to rare mutations or CNVs with high impact.

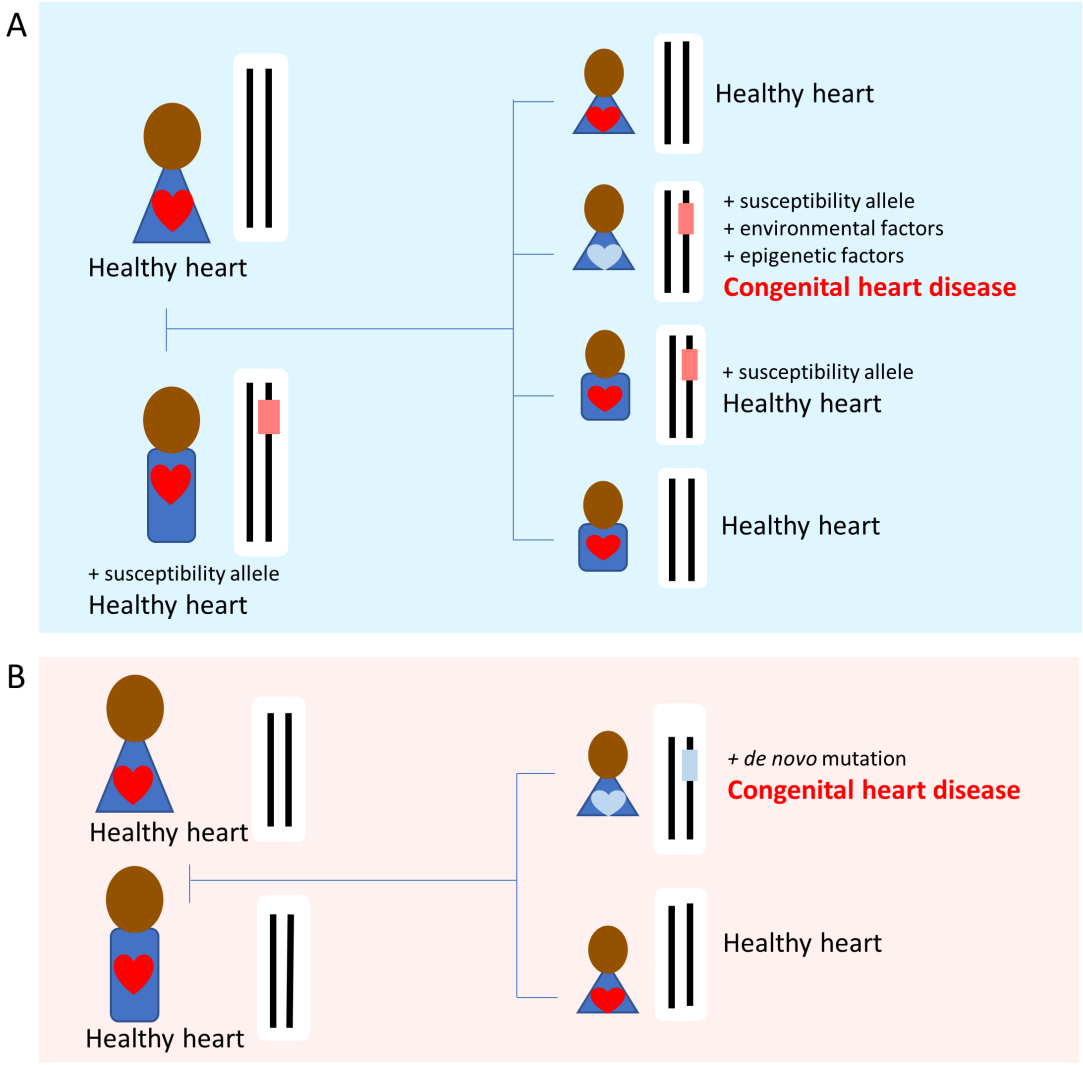

Figure 1.3. Typical inheritance patterns in CHD

A) Complex inheritance, in which CHD is influenced by genetic factors, as well as environmental and epigenetic factors. B) *De novo* inheritance, in which sporadic mutations (or chromosomal rearrangements) may cause CHD. (Illustration by Nicole A. Saacks).

For the majority of cases, especially severely affected patients, CHD occurs sporadically in families with no prior history of the disease, with only 2% of CHD cases exhibiting familial disease (Øyen et al., 2009). Sporadic CHD may arise as a result of *de novo* genetic events, which include single-gene mutations, point mutations such as SNVs, chromosomal aberrations, and smaller CNVs of particular chromosomal regions (Figure 1.4). It is important to note that CHD is usually oligogenic or complex in its genetics and can result from a combination of these genetic factors (Akhirome, Walton, Nogee, & Jay, 2017; Pierpont et al., 2007; Zaidi & Brueckner, 2017).

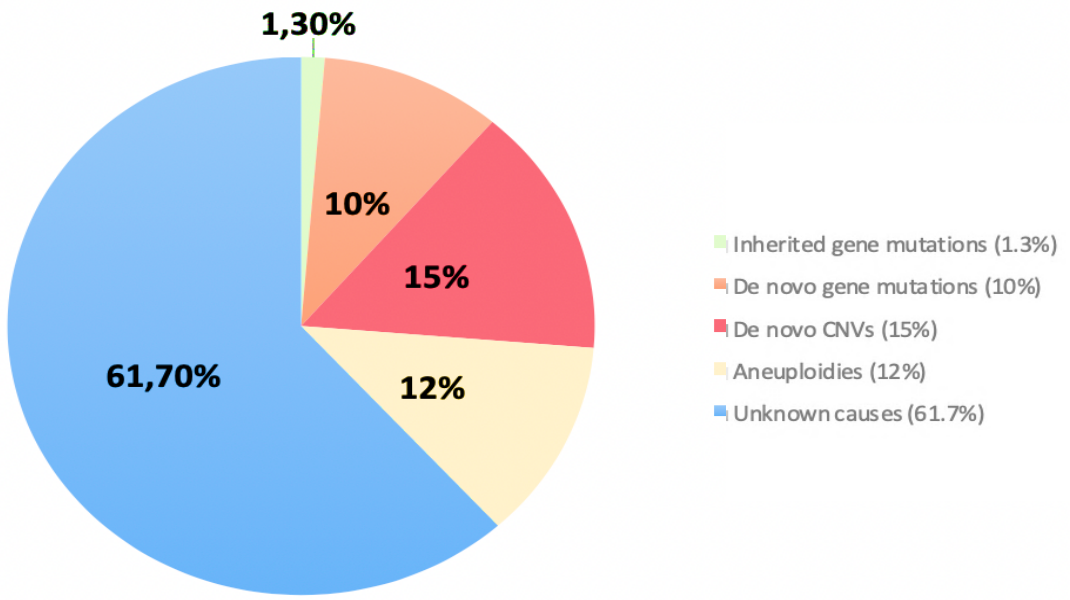

Figure 1.4. Genetic causes of congenital heart disease

Inherited and *de novo* gene mutations comprise the majority of known causes for congenital heart disease. Previous studies have estimated that 12% of CHD pathology is attributable to chromosomal abnormalities. *De novo* CNVs include well-established deletion syndromes such as the 22q11.2 deletion as well as recently discovered CNVs (Adapted from Akhirome et al., 2017).

1.5.4. Single gene mutations associated with congenital heart disease

Traditional genetic techniques including linkage analysis and candidate gene approaches have enabled the discovery of numerous causative genes implicated in CHD pathogenesis as shown in Table 1.2. However, these techniques rely on multiple affected family members and an understanding of the underlying molecular pathways of cardiac development, which may be challenging in CHD. Previous studies have identified rare causal mutations in genes encoding cardiac transcription factors such as *NKX2-5, GATA4*, and *TBX5* in patients with non-syndromic CHD (Durocher, Charron, Warren, Schwartz, & Nemer, 1997; Garg et al., 2003; Schott et al., 1998). These transcription factors control critical events during cardiac development and regulate genes important for cardiomyocyte differentiation, proliferation, and apoptosis (Durocher et al., 1997; Hiroi et al., 2001; Singh et al., 2010).

Genetic mutations in these transcription factors have been most commonly associated with ASD, VSD, and TOF (Benson et al., 1999; Garg et al., 2003; Okubo et al., 2004). Additionally, genes that encode structural proteins including cardiac actins and myosins have been linked to CHD. Examples of the implicated genes include *MYH6, MYH7*, *MYH11,* and *ACTC1* (Wessels & Willems, 2010).

Table 1.2. Genes associated with non-syndromic congenital heart disease*

GENE	*FUNCTION*	*ASSOCIATED CARDIAC LESIONS*
NKX2-5	Transcription factor	ASD–AV block, TOF, HLHS, TGA, DORV, Ebstein anomaly, VSD
NKX2-6	Transcription factor	TrA
GATA4	Transcription factor	ASD, PS, TOF, VSD, DORV
GATA6	Transcription factor	TA, TOF, AVSD
TBX1	Transcription factor	IAA, aortic arch anomalies, VSD
FOG2	Transcription factor	TOF, DORV
FOXH1	Transcription factor	TGA, TOF, VSD
TBX5	Transcription factor	ASD, VSD, AVSD, conduction abnormalities
TBX20	Transcription factor	ASD, VSD, valve defects, LVOTO
CITED2	Transcription factor	ASD, VSD, TOF, TGA
ZIC3	Transcription factor	Heterotaxy, ASD, AVSD, TGA, VSD, TAPVR, PS
ZFPM2	Transcription factor	TOF
HAND1	Transcription factor	HLHS (somatic mutation)
TFAP2B	Transcription factor	PDA
NOTCH1	Membrane ligand–receptor	AS, BAV, TOF
NODAL	Membrane ligand–receptor	Heterotaxy, TGA
JAG1	Membrane ligand–receptor	PS, TOF
CFC1	Membrane ligand–receptor	Heterotaxy, TGA, DORV, TOF
LEFTY2	Membrane ligand- receptor	Heterotaxy
MYH6	Sarcomeric protein	ASD
MYH7	Sarcomeric protein	ASD, Ebstein anomaly
MYH11	Sarcomeric protein	PDA
ACTC1	Sarcomeric protein	ASD, VSD
GJA1	Gap junction protein	HLHS (somatic mutation)
GJA5	Gap junction protein	TOF
CRELD1	Matricellular protein	AVSD, dextrocardia
ELN	Structural protein	SVAS
VEGFA	Mitogen	TOF

*Adapted from (blue et al., 2012; pierpont et al., 2018; wessels & willems, 2010)

Note. Acronyms: AS – aortic stenosis; ASD – atrial septal defect; AV – atrioventricular; AVSD – atrioventricular septal defect; BAV – bicuspid aortic valve; DORV – double outlet right ventricle; HLHS – hypoplastic left heart syndrome; IAA – interrupted aortic arch; LVOTO – left ventricular outflow tract obstruction; PDA – patent ductus arteriosus; PS – pulmonary valve stenosis; SVAS – supravalvular aortic stenosis; TA – tricuspid atresia; TrA – truncus arteriosus; TAPVR – total anomalous pulmonary venous return; TGA – transposition of the great arteries; TOF – tetralogy of Fallot; VSD – ventricular septal defect.

CHD has been attributed to Mendelian syndromes in 3% to 5% of cases (Blue et al., 2012). A selection of the more well-known single gene disorders linked to CHD is shown in Table 1.3. For example, truncating mutations in the T-box transcription factor, *TBX5*, have been linked to Holt-Oram syndrome, a Mendelian disease often associated with cardiac abnormalities (Baban et al., 2014). Haploinsufficiency of the transcription factor *TBX1* is a common finding in individuals with 22q11.2 deletion syndrome and is responsible for many of the associated cardiovascular phenotypes (Zhang et al., 2006).

The NOTCH signalling pathway gene, *NOTCH1*, has been identified as a major susceptibility gene for defects such as BAV, AS and TOF, as shown in Table 1.3 (Garg et al., 2005; Page et al., 2019). Mutations in the *NOTCH1* ligand, *JAG1*, and *NOTCH2* have been associated with Alagille syndrome, an autosomal dominant disorder linked to CHD (Li et al., 1997; McDaniell et al., 2006). Tables 1.2 and 1.3 also illustrate the genetic complexity of the disease, where one gene can give rise to more than one type of CHD, and one CHD subtype can be caused by mutations in more than one gene. The complex nature of CHD has made establishing phenotype-genotype correlations a major challenge for researchers (Øyen et al., 2009).

Table 1.3. Single gene disorders associated with CHD*

CAUSAL GENE	*SYNDROME*	*ASSOCIATED CARDIAC LESIONS*	*LOCUS*	*INHERITANCE*	*REFERENCE*
CBL	Noonan	PS and RVH	11q23.2	AD	(Lepri et al., 2014)
CHD7	CHARGE	ASD, VSD, valve defects	8q12.2	AD and *de novo*	(Pranckėnienė, Preikšaitienė, Gueneau, Reymond, & Kučinskas, 2019)
EVC1	Ellis-van Creveld	ASD/single atrium	4p16	AR	(Tompson et al., 2007)
JAG1	Alagille	TOF, PS	20p12	AD	(Li et al., 1997; McDaniell et al., 2006)
KMT2D/MLL2	Kabuki	CoA, ASD, VSD	12q13.12	AD and *de novo*	(Ang et al., 2016; Digilio et al., 2017)
MAP2K1	Cardiofaciocutaneous	PS, ASD, HOCM	15q22.31	AD and *de novo*	(Lepri et al., 2014; Rodriguez-Viciana et al., 2006)
MAP2K2	Cardiofaciocutaneous	PS, ASD, HOCM	19p13.3	AD and *de novo*	(Lepri et al., 2014; Rodriguez-Viciana et al., 2006)
NRAS	Noonan	PS and RVH	1p13.2	AD	(Cirstea et al., 2010; Lepri et al., 2014)
NOTCH2	Alagille	TOF, PS	1p12	AD	(Kamath et al., 2012; McDaniell et al., 2006)
PTPN11	Noonan	PS and RVH	12q24	AD	(Aoki, Niihori, Narumi, Kure, & Matsubara, 2008; Lepri et al., 2014)
RAF1	Leopard and Noonan	HOCM and PS	3p25.2	AD	(Lepri et al., 2014)
SHOC2	Noonan	PS and RVH	10q25.2	AD	(Cordeddu et al., 2009; Lepri et al., 2014)
SOS1	Noonan	PS and RVH	2p21	AD	(Lepri et al., 2014)
TBX1	22q11.2 del	TA. IAA. TOF	22q11.21	AD and *de novo*	(Zhang et al., 2006)
TBX5	Holt-Oram	Isolated ASD, isolated VSD, PDA	12q24	AD	(Baban et al., 2014)
TFAP2B	Char	PDA	6p12p21	AD	(Zhao et al., 2001)

*Adapted from (Aburawi et al., 2015)

Note. Acronyms: AD – autosomal dominant; AR – autosomal recessive; ASD – atrial septal defect; CoA – coarctation of the aorta; del – deletion; HOCM – hypertrophic obstructive cardiomyopathy; IAA – interrupted aortic arch; PDA – patent ductus arteriosus; PS – pulmonary valve stenosis; RVH – right ventricular hypertrophy; TA – tricuspid atresia; TOF – tetralogy of Fallot; VSD – ventricular septal defect.

1.6. Chromosomal abnormalities associated with congenital heart disease

A chromosomal abnormality occurs when chromosomal material is lost or gained and can cause a range of genetic disorders if dosage-sensitive genes are affected (Blue et al., 2012). Chromosomal abnormalities contribute to approximately 8% to 20% of CHD cases (Akhirome et al., 2017; Fotiou et al., 2019; Roos-Hesselink, Kerstjens-Frederikse, Meijboom, & Pieper, 2005). The chromosomal causes of CHD can be divided into two categories: gross chromosomal anomalies (or aneuploidies), and smaller CNVs as shown in Figure 1.5 (Blue et al., 2012).

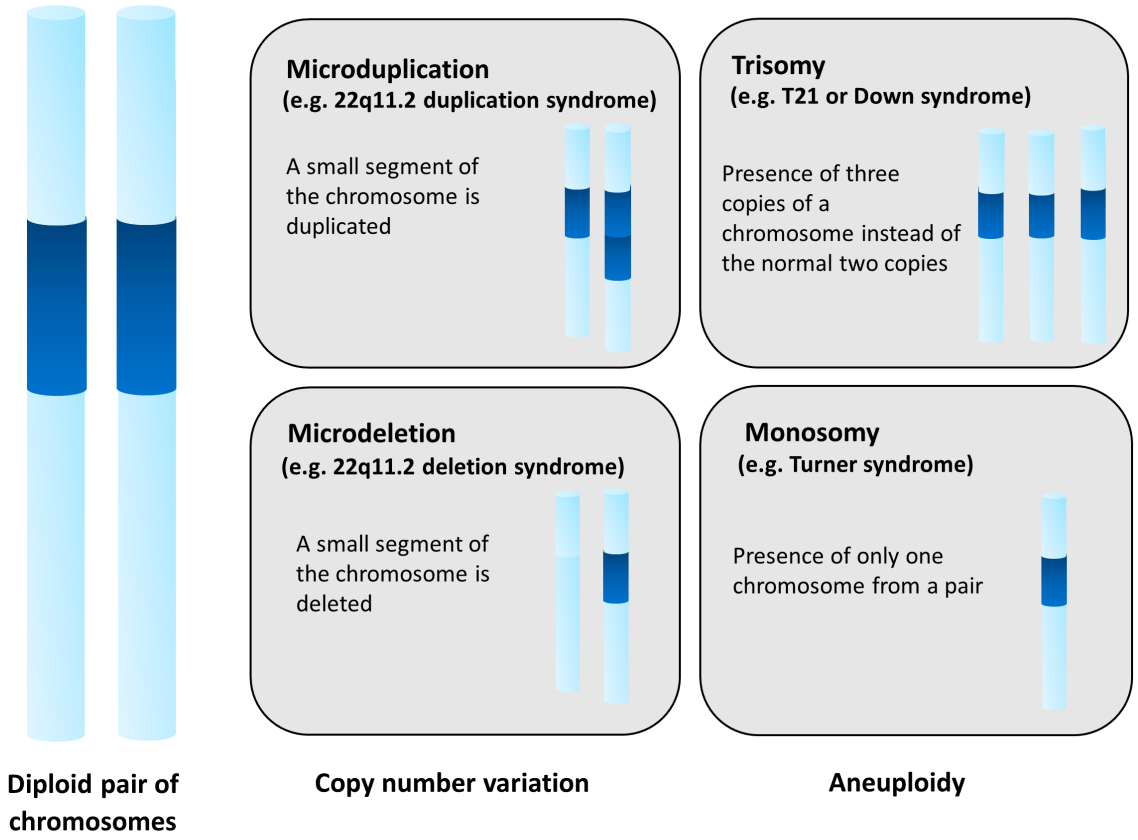

Figure 1.5. Chromosomal causes of CHD
Representation of typical chromosomal alterations. Chromosomes are depicted in light blue and regions of deletion or duplication are in dark blue (Illustration by Nicole A. Saacks).

1.6.1. Aneuploidies

Aneuploidy is typically defined as an abnormal number of chromosomes (Pierpont et al., 2018). Chromosomal aneuploidies were the first genetic causes of CHD to be discovered, and continue to play an important role in CHD pathology today (Fahed et al., 2013; Zaidi & Brueckner, 2017). The most common aneuploidy is Trisomy 21 (T21 or Down syndrome), whereby children are born with an extra copy of chromosome 21 (trisomy 21). This genetic syndrome affects approximately 1 in 800 individuals and is the most common chromosomal disorder seen in individuals with CHD (Pierpont et al., 2018). Cardiac complications are the most common cause of death amongst T21 syndrome patients (Benhaourech, Drighil, & El Hammiri, 2016), and approximately 40% to 50% of individuals with T21 syndrome have an associated cardiac defect. T21 syndrome patients commonly present with an ASD, VSD, PDA, TOF, or AVSD (Pierpont et al., 2018). Turner syndrome (TS) results as a partial or complete loss of the X chromosome in females (Pierpont et al., 2007), and 33% of cases occur in conjunction with CHD, usually on the left side of the heart (Zaidi & Brueckner, 2017). Bicuspid aortic valve is the most common heart defect associated with TS, with a prevalence of 15% to 30%, followed by coarctation of the aorta which has a prevalence of 7% to 18%. These cardiac anomalies can lead to serious complications for individuals with TS, including aortic dilation and dissection (Silberbach et al., 2018). CHD is observed in 60% to 80% of individuals with trisomy 13 (Patau syndrome) and trisomy 18 (Edwards syndrome) (Zaidi & Brueckner, 2017). Another common aneuploidy is Klinefelter syndrome. Approximately 50% of males born with Klinefelter syndrome have an associated CHD, usually presenting with PDA or an ASD (Pierpont et al., 2007). A selection of well-established chromosomal abnormalities associated with CHD is shown in Table 1.4.

Table 1.4. Chromosomal abnormalities associated with congenital heart disease*

Chromosomal anomaly	Locus	Most common CHD
Chromosomal aneuploidies		
Trisomy 8 mosaicism	Chromosome 8	VSD, PDA, CoA, TAPVR, TrA
Trisomy 9 mosaicism	Chromosome 9	PDA, LSVC, VSD, TOF, PA, DORV
Patau syndrome	Chromosome 13 (Trisomy 13)	ASD, VSD, PDA, HLHS
Edwards syndrome	Chromosome 18 (Trisomy 18)	ASD, VSD, PDA, TOF, DORV, TGA, CoA, BAV
Down syndrome	Chromosome 21 (Trisomy 21)	AVSD, ASD, VSD, TOF, TGA
Turner syndrome	Chromosome X (Monosomy X)	CoA, BAV, AS, HLHS
Klinefelter syndrome	Chromosome X (47- XXY)	MVP, PDA, ASD
Copy number variants		
1p36 microdeletion	1p36	PDA, VSD, ASD, BAV, Ebstein anomaly
1q21.1 microdeletion	1q21.1	PDA, VSD, ASD, TrA, TOF TOF, TGA, PS
1q21.1 duplication	1q21.1	TOF, TGA, ASD, pulmonary atresia
1q41q42 microdeletion	1q41q42	BAV, ASD, VSD, TGA
1q43q44 microdeletion	1q43q44	VSD, CoA, HLHS
2q31.1 microdeletion	2q31.1	VSD, ASD, PDA
2q37 microdeletion	2q37	VSD, ASD, CoA
Wolf-Hirschhorn syndrome	4p	ASD, VSD, PDA, aortic atresia, dextrocardia, TA, TOF
Cri-du-chat syndrome	5p	VSD, ASD, PDA
Williams- Beuren syndrome	7q11.23 deletion	AS and PS, PPS
8p23.1 microdeletion	8p23.1	AVSD, PS, VSD, TOF
Kleefstra syndrome	9q34.3 deletion	ASD, VSD, TOF, pulmonary arterial stenosis
Jacobsen syndrome	11q deletion	HLHS, AS, VSD, CoA
15q11.2 microdeletion	15q11.2	TOF, BAV
15q24 microdeletion	15q24	PDA, pulmonary arterial stenosis, PS
16p11.2p12.2 microdeletion	16p11.2p12.2	TOF, BAV, pulmonary atresia
17q21 microdeletion	17q21	PS, ASD, VSD, BAV
Alagille syndrome	20p12 deletion	Peripheral pulmonary artery hypoplasia, TOF, PS
22q11.2 microdeletion	22q11.2	IAA type B, TrA, TOF
22q11.2 microduplication	22q11.2	TOF, HLHS, VSD, PS, TrA
Phelan-McDermid syndrome	22q13 microdeletion	PDA, VSD, ASD, TAPVR

*Adapted from (Blue et al., 2012; Pierpont et al., 2018; Soemedi, Wilson, et al., 2012)

Note. Acronyms: AS – aortic stenosis; ASD – atrial septal defect; AVSD – atrioventricular septal defect; BAV – bicuspid aortic valve; BPV – bicuspid pulmonary valve; CoA – coarctation of the aorta; DORV – double-outlet right ventricle; HLHS – hypoplastic left heart syndrome; IAA – interrupted aortic arch; LSVC – persistent left superior vena cava; MVP – mitral valve prolapse; PA - pulmonary atresia; PDA - patent ductus arteriosus; PS – pulmonary valve stenosis; PPS – peripheral pulmonary stenosis; TA – tricuspid atresia; TAPVR – total anomalous pulmonary venous return; TGA – transposition of the great arteries; TOF - tetralogy of Fallot; TrA,- truncus arteriosus; VSD - ventricular septal defect.

1.6.2. Copy number variation

Copy number variation is a type of structural genetic variation whereby segments of the genome are duplicated or deleted. CNVs are a subgroup of structural variants comprising insertions, deletions, and complex rearrangements of any size. In this study, CNVs are defined as microdeletions or microduplications of the genome that are larger than one kilobase in size and affect the dosage of one or more genes. CNVs typically arise from nonallelic homologous recombination (NAHR) within the genome (Figure 1.6). This process usually occurs during meiosis and is mediated by the flanking of non-allelic low copy repeat regions (LCRs) at homologous sequences in or between chromosomes, which may result in CNVs in the LCRs as well as any sequence therein (Malhotra & Sebat, 2012). Other mechanisms responsible for CNV generation include non-homologous end-joining, fork stalling, and template switching, and L1-mediated retro-transposition discussed elsewhere (Malhotra & Sebat, 2012; Zhang, Carvalho, & Lupski, 2015).

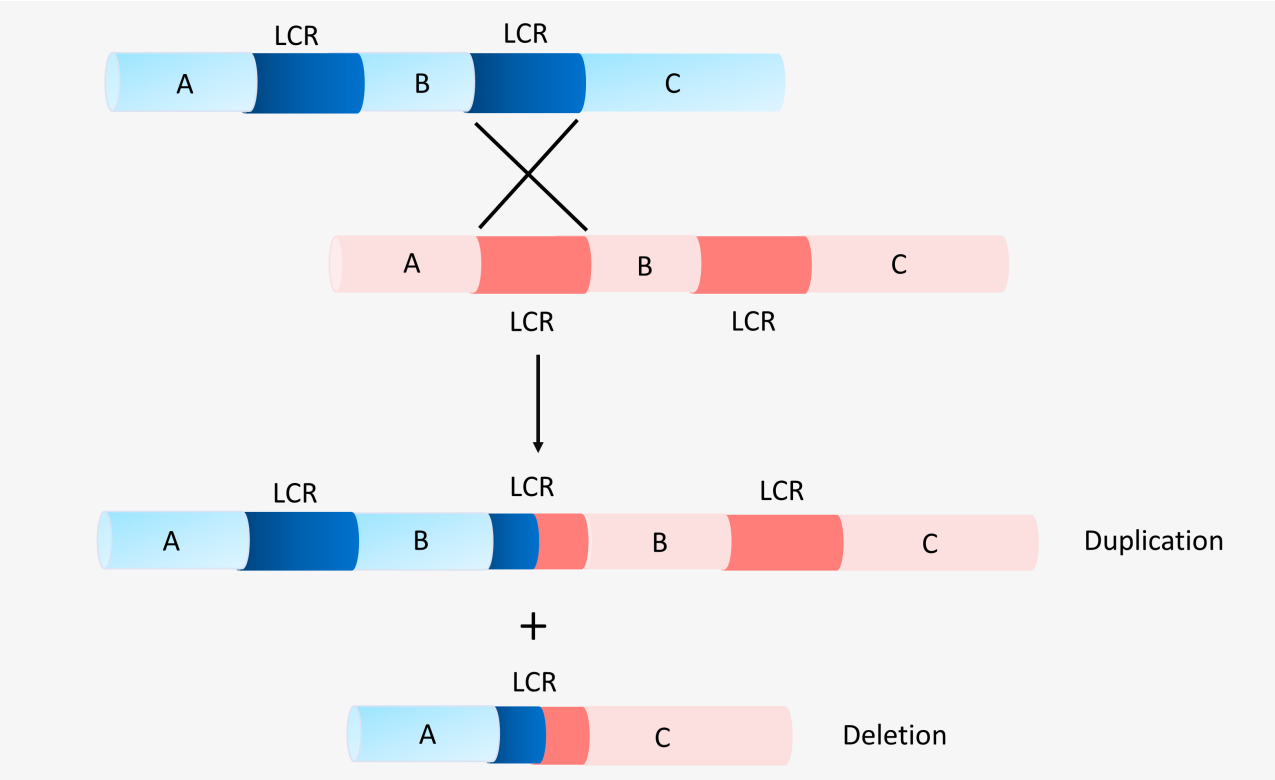

Figure 1.6. Nonallelic homologous recombination
NAHR is a form of recombination between two DNA regions with high sequence similarity, and one of the major mechanisms underlying CNV formation. This unequal crossing over results in reciprocal deletion and duplication of the intervening sequence. If this occurs during meiosis, resultant offspring can inherit a CNV. A, B and C represent different genes. LCR – low copy repeats (Illustration by Nicole A. Saacks).

Genomic microduplications and microdeletions can range from one kilobase (kb) to several mega-bases (Mb) in size and are a common source of genetic variation associated with many Mendelian diseases and genetic disorders (de Ligt et al., 2014). Emerging evidence has indicated that CNVs are important contributors to numerous disorders including cancer (Dixon et al., 2018), neuropsychiatric disorders (Walsh et al., 2008), neurodevelopmental disorders (Zarrei et al., 2018), and congenital defects including CHD (Pang et al., 2010; Tomita-Mitchell et al., 2012). CNV mutations can be inherited or *de novo* and can lead to altered copies of dosage-sensitive genes, the effects of which range from benign to fatal depending on the function of the genes implicated (Truty et al., 2019). Large CNVs comprising several million base pairs can be detected by cytogenic analyses and/or fluorescent in situ hybridisation (FISH), whereas smaller CNVs are identified using high-resolution microarrays that assess SNVs and CNVs (Fahed et al., 2013).

1.6.2.1. Copy number variation in health and disease

CNVs occur frequently amongst healthy individuals, comprising approximately 12% of the average person's genome (Redon et al., 2009). Although the majority of CNVs have no phenotypic consequence in healthy individuals, microduplications and/or microdeletions that implicate dosage-sensitive genes can be detrimental. If critical genetic regulatory elements are disrupted, dependent genes may be over- or under-expressed which can significantly contribute to disease pathogenesis (Lander & Ware, 2014). It is therefore important to differentiate pathogenic CNVs from likely benign CNVs that are commonly found in the general population.

CNVs are commonly associated with incomplete penetrance in which CNV carriers in a family do not always develop disease. Therefore, the inheritance pattern of a CNV is not an adequate measure to determine pathogenicity (Vermeesch, Balikova, Schrander-Stumpel, Fryns, & Devriendt, 2011). In 2011, The American College of Medical Genetics (ACMG) developed a set of guidelines to aid the evaluation and interpretation of CNVs (Kearney, Thorland, Brown, Quintero-Rivera, & South, 2011). In the context of CHD, researchers and clinicians have defined CNV pathogenicity according to the following criteria: a CNV overlapping a known disease-associated region or known dosage-sensitive CHD gene; a CNV located in a gene-rich region; a CNV that comprises a large deletion or duplication; a *de novo* mutation or a CNV associated with a specific phenotype within a family; and/or a rare CNV found in less than 1% of healthy individuals. It is generally agreed upon that pathogenic CNVs share a combination of the above qualities (Lander & Ware, 2014).

The ACMG groups CNVs into three main categories of clinical significance: pathogenic, uncertain clinical significance, and benign (Kearney et al., 2011). A CNV is classified as pathogenic if it has been documented as clinically significant by numerous peer-reviewed publications. This category will include most structural aberrations that are visible by cytogenetic analysis (> 3Mb - 5Mb). CNVs are categorized as variants of uncertain clinical significance (VOUS) when insufficient evidence is available for determination of clinical significance at the time of reporting. Variants within this category can be subclassified as likely pathogenic (for example, the CNV is described in a single report but has characteristics of a pathogenic CNV and a phenotype that is relevant to the patient's phenotype); likely benign (for example, the CNV does not overlap with any genes or is described in the general population but does not represent a polymorphism); and no subclassification (for example, the CNV overlaps with genes which may or may not be dosage-sensitive). A variant is categorized as benign if it has been reported in numerous peer-reviewed studies as a benign variant and/or represents a common polymorphism found in more than 1% of the healthy population (Kearney et al., 2011).

1.6.2.2. Copy number variation in congenital heart disease

The discovery of pathogenic and potentially pathogenic CNVs associated with CHD has significantly improved our understanding of the aetiology of the disease (Pierpont et al., 2018). Investigation of the role of CNVs in CHD pathology has led to the identification of numerous dosage-sensitive genes that are critical for cardiac development. Previous studies of large CHD cohorts have detected a 1.8-fold to 3.9-fold greater burden of CNVs in CHD cases compared to controls, with large, rare, gene-containing CNVs having a greater impact on CHD (Lander & Ware; 2014; Soemedi et al., 2012a; Tomita-Mitchell et al., 2012; Walsh et al., 2008).

1.6.2.3. Copy number variation in syndromic congenital heart disease

Cardiac defects commonly occur in conjunction with a multitude of genetic disorders (syndromic CHD) characterised by large CNVs (Table 1.4). The most common microdeletion in humans is 22q11.2 deletion syndrome, caused by a deletion that is not visible by standard karyotyping on the long (q) arm of chromosome 22 as a result of NAHR (Zaidi & Brueckner, 2017). The cardiac phenotype for 22q11.2 deletion syndrome varies but usually presents with TOF, truncus arteriosus (TrA), and/or interrupted aortic arch- type B (IAA-type B) (Costain et al., 2016). The 22q11.2 deletion overlaps with the T-Box transcription factor gene *TBX1.* It appears that the deletion of *TBX1* is the primary causal mechanism to the clinical phenotype associated with the syndrome (Fulcoli et al., 2016). Additional well

characterised CHD-associated CNVs include a deletion at 7q11.23, which causes Williams-Beuren syndrome (Chung & Rajakumar, 2016), a deletion at 11q24-25, which results in Jacobsen syndrome (Grossfeld et al., 2004), a deletion at 8p23, which leads to haploinsufficiency of *GATA4,* resulting in a variety of CHD phenotypes with neurodevelopmental delay (Wat et al., 2009), and a 1p36 deletion, which commonly occurs in conjunction with a septal defect, and is associated with orofacial malformations, microcephaly and mental retardation (Chung & Rajakumar, 2016). The majority of these genetic disorders cannot be identified by conventional karyotyping and require CMA or FISH for accurate diagnosis (Lander & Ware, 2014).

1.6.2.4. Copy number variation in non-syndromic congenital heart disease

Non-syndromic CHD can occur in isolation or in conjunction with extra-cardiac anomalies (ECAs) such as neurodevelopmental delay and/or dysmorphism. Several studies have demonstrated that CNV is a significant contributory factor to the development of CHD with ECAs. For example, a study by Thienpont et al. (2007) found a rare causative CNV in 17% of 60 patients including CNVs in regions of known cardiac transcription factors (*NKX2-5* and *NOTCH1*), suggesting an association between CNV and CHD. A similar study by Richards et al. (2008) identified rare CNVs in 25% of 40 patients with CHD who showed normal karyotypes. Half of the cohort presented with isolated CHD and the other half had CHD with ECAs. However, causative CNVs were only detected in study subjects who presented with CHD and ECAs. Syrmou et al. (2013) detected CNVs in 37 of 55 individuals with CHD (67%), 81% of the CNV-positive CHD patients presented with ECAs. Collectively, these studies demonstrate the important role of CNVs in the development of CHD with ECAs, and that the genes implicated in CHD tend to have multiple phenotypic effects. These studies also indicate that CMAs can be a useful tool to identify causative CNVs in individuals presenting with CHD and ECAs when the standard karyotype appears normal (Richards et al., 2008).

Most CHD cases (up to 85%) occur in isolation without ECAs. The role of CNVs in isolated CHD has been minimally explored. Identifying isolated CHD is often challenging, as ECAs can be easily missed, or not yet present at the time of diagnosis, especially in very young study populations. Longitudinal studies with carefully phenotyped study subjects are required to define the role of CNV in isolated CHD (Lander & Ware, 2014). Previous large-scale studies have investigated CNVs in cohorts which include patients with isolated CHD and patients with CHD and ECAs. A study by Soemedi et al. (2012b) found that a duplication of the gap-junction gene *GJA5* increased the risk of TOF by ten-fold. This study also showed that microdeletions of the 1q21.1 region corresponded to a population-attributable risk of approximately 1% for TOF. Furthermore, a large-scale genome-wide investigation of CNV data

from 2,256 CHD-affected probands, 283 trio CHD-affected families, and 1,538 controls was performed by Soemedi, et al., 2012a). This study showed that rare deletions account for 3% to 4% of the population attributable risk for TOF and other CHDs (Soemedi et al., 2012a). Tomita-Mitchell et al. (2012) explored the effect of CNVs in 945 individuals diagnosed with CHD and detected pathogenic CNVs in 4.3% of their study subjects (excluding 135 patients with syndromic CHD-associated chromosomal abnormalities). Additionally, a CNV analysis by Erdogan et al. (2008) identified *de novo* causative CNVs in 3% of 105 patients with isolated CHD presenting with varied phenotypes. Many of the identified CNVs contained genes important for cardiac development, and/or genes critical for correct left-right patterning of the heart according to animal models. A recent study by Kim et al. (2016) identified large pathogenic CNVs (> 300kb) that were significantly associated with increased postoperative mortality in non-syndromic CHD patients.

Overall, these studies highlight the significant impact of CNV in CHD pathology and indicate that as genetic testing progresses, investigating and identifying pathogenic CNVs associated with specific forms of CHD will become an increasingly useful tool in gene discovery and accurate CHD diagnosis. This is particularly important for patients with complex CHD and CHD with ECAs, whereby identifying additional anomalies not easily detected by standard karyotyping, could significantly improve disease prognosis and patient outcome. Furthermore, exploring the role of CNVs in CHD will contribute to our understanding of healthy cardiac development and related perturbations, and provide knowledge relevant to clinical practice and potential therapeutic strategies (Edwards & Gelb, 2016).

1.7. Discovering genes for congenital heart disease

Over the years, many genes involved in cardiac development have been discovered. However, the complete process of cardiogenesis is not fully understood (Aburawi et al., 2015). Emerging sophisticated, high-throughput genetic technology such as next-generation sequencing (NGS) has rapidly advanced the pace at which genes are being discovered. NGS technologies include gene panel tests, whole-genome sequencing (WGS), and WES. WES is a genomic technique that sequences all the protein-coding regions of the genome and has been used to identify rare, causative SNVs in CHD-associated genes as well as previously unreported genes, in both small and large patient cohorts. For example, Page et al. (2019) used WES to detect unique pathogenic variants in a cohort of 829 non-syndromic TOF patients. Similarly, a study by Zaidi et al. (2013) performed WES on 362 sporadic severe CHD patients, their parents, and 264 control trios and found a significant excess of *de novo*, damaging mutations in genes involved in cardiac development amongst the patients.

In the past, detection of aneuploidies and gross chromosomal aberrations has relied heavily on standard karyotyping. However, smaller CNVs are not easily detected by this method and require a higher resolution CMA analysis (Figure 1.7). CMA is a clinical genetic technique that includes single nucleotide polymorphism (SNP) arrays or comparative genomic hybridisation (CGH) and allows for the interrogation of rare CNVs known to be associated with a disease, as well as other emerging chromosomal deletions and duplications in the genome not visible by standard karyotyping (Liao et al., 2014). Although the resolution of CMAs depends on many factors, array CGH and SNP arrays are usually able to detect CNVs upwards of 50-100 kb in size (Figure 1.7). It is likely that 5% to 15% of non-syndromic CHD is attributable to CNVs above 100 kb, although the contribution of CNVs smaller than 50 kb is unclear due to the limited resolution of these techniques (Figure 1.7). Sequencing technologies such as WES are only able to detect insertions or deletions less than 100 bp in size, but the role of CNVs between 100 bp and 50 kb may be delineated using long-read sequencing, which has not been utilised in CHD studies as yet. Nevertheless, CMA is becoming an increasingly important tool used in both prenatal and postnatal clinical genetic settings (Liao et al., 2014) and has led to the identification of numerous pathogenic CNVs implicated in CHD discussed elsewhere (Erdogan et al., 2008; Goldmuntz et al., 2011; Richards et al., 2008; Soemedi et al., 2012b; Thienpont et al., 2007; Tomita-Mitchell et al., 2012; Wu et al., 2017). The International Standard Cytogenomic Array Consortium recommends CMA as the first-tier cytogenetic diagnostic test for children born with congenital abnormalities including CHD, as it provides the most comprehensive coverage of the genome (Miller et al., 2010). The CytoScan HD array by Affymetrix is the highest performing CMA for detecting chromosomal abnormalities, with 2.67 million-copy number markers and a genotyping accuracy of 99% (Wang et al., 2015; Yu, Michalopoulos, Ding, Tseng, & Luo, 2014). This high-density array has been coined the ‘gold standard’ for detecting CNVs within a population and this technique is largely replacing FISH and standard karyotyping at many institutions (Cai et al., 2018; de Ligt et al., 2014; Shen et al., 2018).

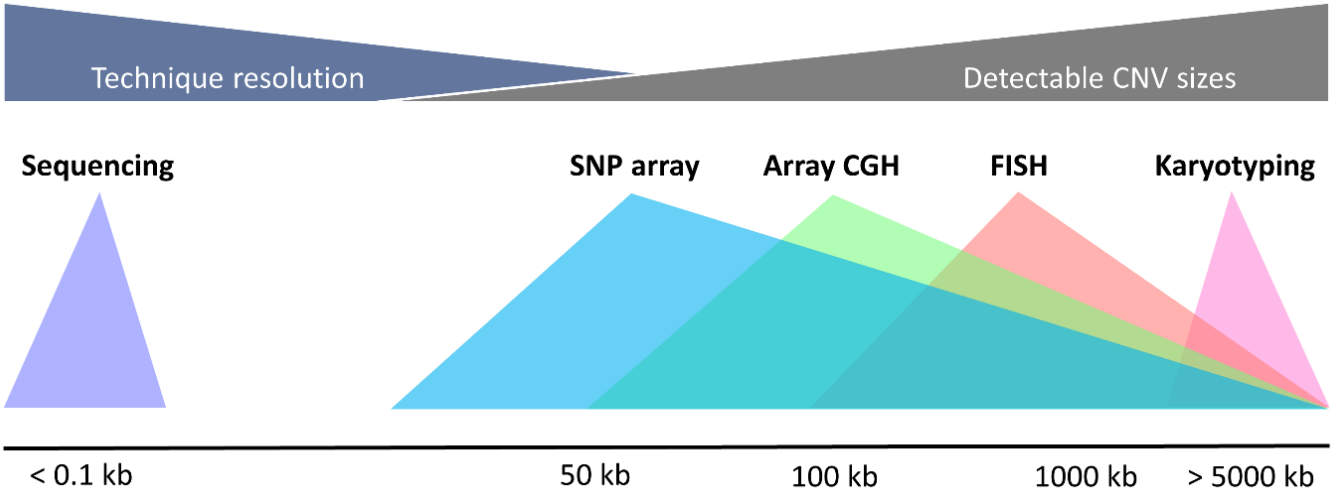

Figure 1.7. Comparison of different CNV detection methods
Indicated are the sizes of CNVs that are detectable by each genetic technique. Karyotyping is limited to the detection of whole and partial chromosomal aneuploidy, while FISH, array CGH and SNP array can be used to detect smaller CNVs. Sequencing technologies such as whole-exome sequencing and whole-genome sequencing can only detect small insertions or deletions (approximately 100 bp). Acronyms: CGH – comparative genomic hybridisation; CNV – copy number variation; FISH - fluorescent in situ hybridisation; kb – kilobase; SNP – single nucleotide polymorphism (Illustration by Timothy F. Spracklen).

1.8. Genetics of congenital heart disease in sub-Saharan Africa

CHD has been described as a 'neglected' condition in South Africa, with an underestimated reported prevalence of approximately 2.35 per 1000 live births (Liu et al., 2019). In 2013, Zühlke, Mirabel, and Marijon provided evidence that suggested the burden of CHD is vastly underestimated as a result of poor prognoses for African children born with CHD. There is a lack of African-based evidence on the genetic basis of cardiovascular disease in SSA due to poor funding and limited local expertise (Laing, Kraus, Shaboodien, & Ntusi, 2019; Mocumbi, 2012). Consequently, the major genetic break-throughs for CHD that have been seen in developed regions of the world over the past few decades have not been replicated in most LLMICs including South Africa.

The epidemiology of CHD in Africa shows a spectrum of CHD phenotypes with varied prevalence (Thomford et al., 2018). However, there are relatively few genomic studies in African CHD populations; these studies could contribute to our understanding of the epidemiological data, and how we manage and treat African children with CHD (Thomford et al., 2018). A comprehensive review of the available

literature on the genetics of CHD in SSA revealed three independent studies that used molecular genetic approaches to investigate the causes of CHD in African populations (De Decker et al., 2016; Teteli et al., 2014; Wonkam et al., 2017). A genomic study in Rwanda was done in 2014 where echocardiography, standard karyotyping, and Multiplex Ligation-dependent Probe Amplification (MLPA) was performed on 125 patients with clinical features of genetic disorders. The study showed that 64 of the 125 study subjects had CHD, and a genetic cause was found for 61 of the 64 subjects, although this high detection rate may have been influenced by the selection of participants with signs of genetic syndromes. Of the 22 patients who presented with normal karyotypes, MLPA and FISH analyses enabled Teteli and colleagues to detect a 7q11.23 duplication, a 13qter deletion, and a 22q11.2 deletion within the study cohort (Teteli et al., 2014). Two years later, De Decker et al. (2016) used FISH to determine the prevalence of 22q11.2 deletion syndrome in South African children presenting at Red Cross War Memorial Children's Hospital (RCWMCH) with CHD. This genetic study identified deletions at the 22q11.2 locus in six out of 125 patients (4.8%). Similarly, Wonkam et al. (2017) investigated the prevalence of 22q11.2 deletion syndrome in CHD patients in Cameroon using MLPA and FISH. In this study, 22q11.2 deletion was detected in 2 of 70 patients (2.8%). Both patients had conotruncal heart defects in conjunction with ECAs (Wonkam et al., 2017). Collectively these studies illustrate the ability to diagnose syndromic CHD using standard karyotyping and FISH, and their relevance in genomic studies on the African content. However, these CHD studies focus mainly on individuals with known genetic disorders which contribute to a small fraction of the total burden of CHD. The low resolution and coverage of the techniques limit the ability to identify smaller CNVs which may contribute, at least in part, to the genetic basis of CHD in Africa.

To our knowledge, there has not been any research into the genetic basis of non-syndromic CHD in Africa, and implementing techniques that can identify potential causative SNVs and CNVs in genes involved in cardiac development is becoming increasingly important. Advanced genetic platforms including next-generation technologies (WGS and WES) and the high-resolution CMA platform, CytoScan HD, allow for the interrogation of massive gene numbers and CNVs of all sizes in a single experiment (Thomford et al., 2018). Using these platforms to investigate the genetic underpinnings of CHD in Africa will provide valuable insight into the complexities of CHD, by validating the disease-causing genes found in other populations, refining disease-associated loci, and possibly identifying new genes that may contribute to CHD pathogenesis. The study presented in this thesis is the first to explore the genetic basis of CHD in Africa using the high-throughput CytoScan HD array platform. Improving our understanding of the genetic architecture and risk factors associated with CHD patients of African descent is the first step toward improving the accuracy of CHD diagnosis, and a crucial step toward identifying potential measures to combat CVD. Moreover, exploring the genetics of CHD has

the potential to improve the quality of life for children born with CHD in Africa, and will enable clinicians to predict recurrent risk, identify familial inheritance patterns and predict prognostic outcome pre- and post-surgical intervention (Thomford et al., 2018).

In summary, there is a dearth of data on the genetic predisposition to CHD in African populations. Genomic studies could help to elucidate preventative risk factors specific to the SSA CHD population, which could eventually translate into prevention strategies ultimately decreasing the burden of CVD in Africa (Thomford et al., 2018).

1.9. Partnerships for Congenital Heart Disease in Africa

Partnerships for Congenital Heart Disease in Africa (PROTEA) is a collaborative project between the University of Cape Town (UCT) and the University of Manchester designed to address the gaps in CHD epidemiology in Africa. The project aims include developing a comprehensive phenotypic/genotypic registry, creating a biorepository for DNA extraction and genetic analyses, and using computational fluid dynamics to develop potential new treatments for CHD. Through this project, a UCT-based CHD research infrastructure has been established to determine the feasibility of conducting genetic research in South African CHD patients, and if feasible, establish a model that can be applied to collaborating African countries in the future.

1.10. The rationale for this study

Data on the genetic factors that may cause CHD in South Africa is scarce, for which the rate of morbidity and mortality differs vastly from Western populations. There are large gaps in knowledge regarding risk stratification and patient outcome for children born with heart defects in South Africa. CNV has been shown to contribute to the risk of being born with CHD. Many large-scale CNV studies have investigated the contribution of CNVs to the development of cardiac defects in individuals of European ancestry. While this approach has identified numerous pathogenic and likely pathogenic variants linked to CHD, to our knowledge, no study to date has investigated the contribution of rare CNVs to the development of CHD in African populations including South Africa. Performing a CMA on the CytoScan HD array platform will enable us to identify potential genetic causes for CHD and likely causative CNVs in a South African CHD population.

This sub-study of the PROTEA project investigates the genetic basis of CHD, specifically focusing on the role of CNVs in the development of CHD for a subset of the PROTEA cohort, where the cause of their disease remains unknown.

This genetic investigation will contribute to our understanding of CHD epidemiology in SSA and contribute toward the establishment of a densely phenotyped, genotyped, and prospectively followed CHD cohort in Africa, in accordance with the aims of PROTEA. This will significantly impact the efficacy of care and improve the outcome for African children born with CHD.

1.11. Aims and objectives

This study aims to identify pathogenic and likely pathogenic CNVs in South African cases of CHD. This will be addressed through the following objectives:

a) Identify known and novel genomic regions (CNVs and/or genes) associated with CHD.

b) Determine the frequency of pathogenic and likely pathogenic CNVs in a cohort of 101 CHD probands.

c) Develop methods to validate CNVs of interest using WES data.

In the next chapter, I will describe in detail the methods used to investigate the role of CNVs in a South African CHD cohort. The third chapter of this thesis will present the scientific findings, followed by a chapter in which the results will be put in context and future directions for the research will be discussed.

Chapter 2. Methods and Materials

2.1. Ethical considerations

This study was approved by the Human Ethics Research Committee (HREC) of the Faculty of Health Sciences, UCT (UCT HREC number: 339/2019) (Appendix A). The work undertaken in this study was in keeping with the standards stipulated in the declaration of Helsinki.

2.2. Study subjects

Patients from the Cardiac Clinics at RCWMCH and Groote Schuur Hospital (GSH) were selected to participate in the genetic investigation. The study cohort was recruited over two years from April 2017 and comprised 105 unrelated individuals with CHD (48 males and 57 females) from birth to 100 years old. Participants were eligible for the study if they presented with isolated CHD (n = 76) or CHD with additional ECAs (n = 17) but did not have a specific genetic diagnosis or recognized chromosomal anomaly detected by standard techniques (non-syndromic CHD patients). An additional twelve patients presenting with known genetic syndromes and chromosomal abnormalities such as T21 syndrome and 22q11.2 deletion syndrome were included as positive controls for the study (syndromic patients).

2.2.1. Patient recruitment and enrolment

Patients presenting with CHD at RCWMCH and GSH were recruited by PROTEA co-investigators to participate in this study. Participants and their family members underwent a process of genetic counselling and history taking to assist in constructing a genealogy and to ensure that participants fully understood the implications of their participation. If family members agreed to participate in the study, informed consent was obtained for each participant (or from their parents, if affected probands were too young to consent themselves) (Appendix B). In the adult institutions, or if participants were older than seven and younger than eighteen, assent and informed consent from the participant was taken. As one of the main objectives of the PROTEA study is to set up a prospective CHD cohort and lay the groundwork for long-term follow-up, it was necessary to use personal information, and study subjects needed to remain identifiable after enrolment. Therefore, basic demographic information such as sex,

date of birth, and self-reported ethnicity was provided by each participant. Additionally, blood samples (4 - 8ml) were obtained from each participant for use in downstream molecular investigations.

2.3. DNA extraction

Two vials of peripheral blood (4 - 8ml) were drawn from each study subject by an HPCSA registered clinician. DNA was extracted from each blood sample using the Gentra PureGene™ DNA isolation kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions (Appendix C) (Gentra® PureGene® Handbook, Qiagen, 2014). In brief, red blood cells, which lack nuclei, were removed from the whole blood samples using a Red Blood Cell lysis Buffer (Gentra® PureGene® Handbook, Qiagen, 2014). The remaining pellets were then resuspended in Cell Lysis solution (Qiagen) to lyse the white blood cell and nuclear membranes, allowing access to the nucleus where DNA is stored. A Protein Precipitation Solution (Qiagen) was then added to remove protein contaminants. Finally, the genomic DNA was recovered by precipitation with alcohol and dissolved in DNA hydration solution (1 mM EDTA, 10 mM Tris·CI pH 7.5). The process is presented in Figure 2.1. DNA extracts were stored at -80°C. Each sample was given a code to ensure patient confidentiality and anonymity.

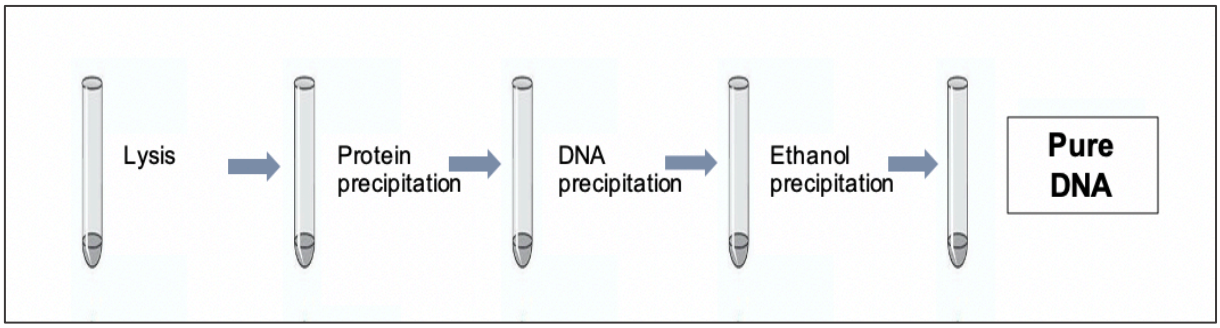

Figure 2.1. The process of DNA extraction (Adapted from Gentra® PureGene® Handbook, Qiagen, 2014)

2.3.1. DNA quality control

The quality of DNA extracts was assessed before use in downstream molecular investigations. This included the spectrophotometric quantification of DNA and its visualisation by agarose gel electrophoresis.

2.3.1.1. Spectrophotometric quantification of DNA

The quality and quantity of DNA were measured using a NanoDrop™2000 spectrophotometer (Thermo Fisher Scientific, USA). DNA concentration was indicated at an absorbance of 260nm (A_{260}). The NanoDrop spectrophotometer measures protein contamination by providing a ratio between the maximum absorbance wavelength for DNA (A_{260}), and the maximum absorbance wavelength for protein (A_{280}). Additionally, the machine measures organic solvent contamination by providing a ratio between the maximum absorbance wavelength for DNA (A_{260}) and the maximum absorbance wavelength for organic solvents (A_{230}) such as ethanol. Pure isolated DNA samples have an $A_{260/280}$ ratio of approximately 1.8 and an $A_{260/230}$ ratio which ranges between 2.0 - 2.2. Deviation from these ranges suggests protein and organic solvent contamination respectively.

2.3.1.2. Gel visualisation of DNA

The integrity of the extracted DNA was measured by gel electrophoresis through a 1% (w/v) agarose gel at a voltage of 100 V for 1 hour (Appendix D). DNA was loaded onto an agarose gel with 3 µl of 1X loading dye and stained with fluorescent nucleic acid dye, GelRed™ (Biotium, California, USA), at a ratio of 1:10. GelRed™ intercalates between the base pairs of the DNA strands and can be visualised under ultraviolet light. In addition, a 100 bp DNA ladder (New England Biolabs®, Massachusetts, USA) was loaded onto each gel and electrophoresed alongside the extracted DNA samples (Appendix D). Gel electrophoresis is a technique used to separate nucleic acid fragments according to their size. DNA fragments are negatively charged, and therefore move towards the positive electrode when a current is passed over the gel. Since all DNA fragments have the same amount of charge per mass, smaller fragments move through the gel faster than large ones. When a gel is stained with a DNA-binding dye, the DNA fragments can be visualised as bands, which represent a group of DNA fragments of the same size (Lee, Costumbrado, Hsu, & Kim, 2012). The CytoScan HD assay requires that all input DNA should be of a high molecular weight where high-quality genomic DNA will run as a clear band on the gel.

2.4. Genotyping of samples

Genotyping was performed using the Affymetrix CytoScan HD platform (Affymetrix, Santa Clara, CA) according to the manufacturer's specifications at the Centre for Proteomic and Genomic Research in Cape Town (CPGR). Briefly, 250 ng of patient DNA was digested with the NspI restriction enzyme, amplified with DNA polymerase chain reaction (PCR), and further purified and fragmented. The

samples were then labelled with biotin-end labelled nucleotides, hybridised to an Affymetrix CytoScan HD array for 17 hours and subsequently washed, stained, and scanned according to the manufacturer's instructions (ThermoFisherScientific). The array is designed specifically for cytogenetic research with over 2700000 markers spanning the genome, including 750000 SNP probes and 1950000 probes to detect CNVs. A reference genomic sample, supplied by Affymetrix, was genotyped alongside every study subject (Santa Clara, CA).

2.4.1. Chromosomal microarray quality control

To be included in downstream analyses, CMA runs were required to meet or exceed the quality control (QC) thresholds specified by Affymetrix. These included a Median Absolute Pairwise Difference (MAPD) of ≤ 0,25, an SNP Quality Control (SNPQC) of ≥ 15,0, and a Waviness Standard Deviation (SD) of ≤ 0,12. These parameters will be explained below.

MAPD estimates the variability in each microarray by calculating the median pairwise copy number (CN) difference between adjacent markers across the genome. A high MAPD value can be attributed to intrinsic variability in the DNA sample, hybridisation preparation, microarray, or scanner, and suggests that CN calls may be inaccurate, increasing the rate of false-positive/negative results. SNPQC estimates the distribution of homozygous and heterozygous genotype alleles, and calculates the distance between them, therefore measuring how well the alleles are resolved in the microarray data. A high resolution of homozygotes and heterozygotes will produce a high SNPQC score. A low SNPQC value indicates compromised quality of SNP allele data, due to higher noise within the microarray (CytoScan ® HD Suite DataSheet CytoScan ® HD Array specifications, 2012). Waviness-SD is a global measure of the long-range variation of microarray probes. Microarray data with Waviness-SD > 0,12 indicates either sample or processing batch effects that will result in reduced quality CN calls. However, it is important to note that elevated Waviness-SD with adequate MAPD and SNPQC metrics can occur in samples with large regions of chromosomal change or samples with multiple copy number changes, and therefore should be assessed with caution. Samples that failed QC were removed from the analysis (CytoScan ® HD Suite DataSheet CytoScan ® HD Array specifications, 2012).

2.4.2. Bioinformatic analysis of copy number variants

Affymetrix Intensity (CEL) files were processed using Affymetrix Power Tools (APT) software (v2.11.0), with default settings and CNV calling thresholds. The detected CNVs were filtered under stringent QC

criteria for their size, probe content, and algorithm-specific quality score as presented in Figure 2.2. All reported CNVs were based on NCBI human genome build 37 (hg 19).

2.4.2.1. Quality filtering of copy number data

In a first exploration, the CN data sets from all 101 study subjects were extracted into a text-based form. Karyotypes of the study subjects were generated using the ggplot2 package of the statistical software R (http://www.r-project.org/), to investigate gross chromosomal rearrangements and confirm samples with clinically diagnosed genetic syndromes.

CN data from the 12 positive controls were removed from the analysis, and the data sets from the remaining 89 non-syndromic CHD patients were combined for further investigation. Poor quality CNV calls that did not meet the QC criteria stipulated by the Affymetrix Chromosome Analysis Suite (ChAS) software [Confidence > 0,85; Marker count > 50; Mean marker distance < 15 000 bp] were removed from the analysis. The CN state of each autosome was normalised by subtracting two from the standard autosome CN state of two so that a normal diploid state would have a score of zero. These scores were then filtered to identify chromosomes with an abnormal CN state. Similarly, the CN state of each sex chromosome was normalised in females by subtracting two from the standard CN state of the X chromosome (CN = 2), and in males by subtracting one from the standard CN state of the X (CN = 1) and Y (CN = 1) chromosomes. Once again, chromosomes with a normal diploid state with a score of zero were removed to identify chromosomes with an abnormal CN state. Detected CNVs that met the following criteria were selected for further analysis: deletions ≥ 100 kb, duplications ≥ 100 kb, and CNVs with less than 70% overlap with known centromeric/telomeric chromosomal regions.

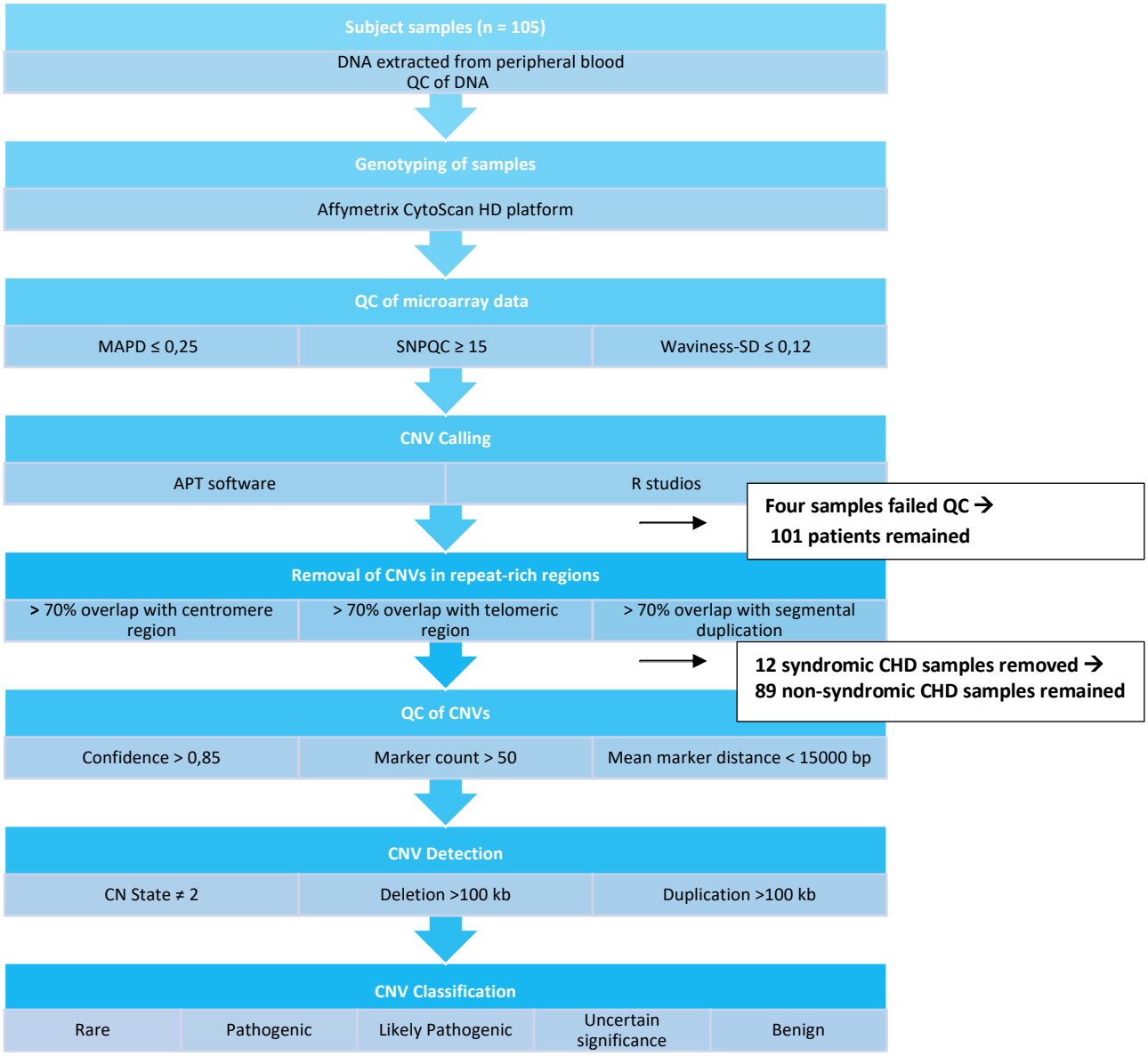

Figure 2.2. Overview of CNV analysis and QC workflow

2.4.2.2. Copy number variant classification

Rare CNVs were distinguished from common CNVs, by monitoring the degree of overlap between CNVs of interest and previously identified common CNVs annotated by the UK Biobank database (https://www.ukbiobank.ac.uk/). The UK Biobank database comprises 2792 CHD cases and 472378 control samples, and in our analysis, we selected rare CNVs found in less than 1% of the healthy population. CNVs were classified as either pathogenic, likely pathogenic, benign, or of uncertain clinical significance in accordance with the guidelines suggested by the ACMG (Kearney et al., 2011) and Battaglia et al. (2013) as mentioned in the previous chapter.

2.5. Identification of known copy number variants and genes associated with congenital heart disease

To identify CNVs occurring in gene regions known to be associated with CHD, all CNVs that overlapped with one or more CHD genes listed in the Genomics England PanelApp gene panel for non-syndromic CHD were investigated (available at https://panelapp.genomicsengland.co.uk/panels/212/). The Genomics England PanelApp is a publicly available crowdsourcing tool that represents a consensus of causative genes for many diseases. The genes in each panel are rated according to a traffic-light system, in which green genes represent genes with a high level of evidence for the gene-disease association, amber genes represent genes with a moderate level of evidence, and red genes represent genes with minimal evidence. A list of the CHD genes in the panel alongside their genetic rating is presented in Table 2.1.

Table 2.1. Non-syndromic CHD gene panel (Version 1.51) – Genomics England Panel App

GENE SYMBOL	RATING	GENE SYMBOL	RATING	GENE SYMBOL	RATING
ABL1	Green	***TAB2***	Green	***FOXH1***	Red
ACVR2B	Green	***TBX5***	Green	***FOXL1***	Red
CFAP53	Green	***TRAF7***	Green	***GJA1***	Red
CFC1	Green	***ZIC3***	Green	***HAND2***	Red
ELN	Green	***ACTC1***	Amber	***IRX4***	Red
FLNA	Green	***CRELD1***	Amber	***MED13L***	Red
GATA4	Green	***GATA5***	Amber	***MYOM2***	Red
GATA6	Green	***LEFTY2***	Amber	***NFATC1***	Red
GDF1	Green	***NOTCH1***	Amber	***NKX2-6***	Red
JAG1	Green	***SHROOM3***	Amber	***PLXND1***	Red
MMP21	Green	***TBX20***	Amber	***RPSA***	Red
MYH6	Green	***ALDH1A2***	Red	***SEMA3D***	Red
NKX2-5	Green	***CITED2***	Red	***SMAD6***	Red
NODAL	Green	***CRKL***	Red	***TBX1***	Red
NOTCH2	Green	***DISP1***	Red	***TLL1***	Red
NR2F2	Green	***FLT4***	Red	***ZFPM2***	Red

In a second investigation, all CNVs overlapping a less stringent, in-house list of candidate causative CHD-associated genes were further analysed. The candidate gene list comprises 157 genes that have been associated with CHD in at least one published paper (Appendix E). The gene list was generated at UCT through a comprehensive literature review. CNVs of interest were visualised using the Affymetrix ChAS software (version 4.0) (Thermofischer Scientific).

2.6. Identification of candidate genes

To identify candidate disease-causing genes, we first excluded CNVs overlapping known CHD-associated genes from further investigation. Next, we selected for genes that were likely intolerant of loss-of-function (LoF) mutations. The Exome Aggregation Consortium (ExAC) (http://exac.broadinstitute.org/) has assigned genes a computed score, which indicates the probability that the gene is intolerant of a LoF mutation (pLI). A gene with a high pLI score (≥ 0,8) is considered extremely intolerant of LoF mutations, whereas genes with low pLI scores are more tolerant of LoF mutations. The computed scores were developed based on high-quality exome sequence data for over 60000 individuals of different ethnicities (Lek et al, 2016). We therefore filtered for CNVs containing genes with a pLI score ≥ 0,8. To further refine our candidate gene list, we examined the gene expression pattern in the embryonic mouse heart using the Gene Expression Database (GXD) in Mouse Genome Informatics (MGI) (http://www.informatics.jax.org/expression.shtml). Finally, we selected genes with a reported cardiac phenotype when mutated in mice, again using the MGI database.

2.7. Validation of the identified copy number variants by exome sequencing

A related, ongoing sub-study of PROTEA involves exome-sequencing of a similar overlapping study cohort. We used the available data to independently validate our CMA findings. Exome target enrichment and sequencing library generation were performed by the Ion AmpliSeq Exome RDY kit (ThermoFisher Scientific). We used the Ion AmpliSeq target enrichment metadata file to identify the targeted regions (exons) within our CNV regions of interest in the patient and 79 control samples. We then used Rsamtools to count the reads that lie within the targeted regions of each CNV region of interest. Aligned reads were only counted if they were non-duplicate, both paired reads were mapped to the reference genome (GRCh37), primary alignments and with a mapping quality of greater than ten. Read counts for each CNV of interest were calculated as the sum of aligned reads covering the targeted regions of the exome-sequencing platform (Ion AmpliSeq Exome RDY) within the CNV regions

for each sample within the cohort. Read count ratio was calculated for all probands displaying a CNV of interest, the ratio was calculated as the proband read count divided by the population mean read count. Finally, p-values were calculated by a one-tailed Z-test (in the direction implied by the array-based discovery analysis) using the population mean and standard deviation for each CNV of interest. A p-value of less than 0.05 was considered statistically significant. In the event of partial gene overlaps, we manually refined the partially-overlapped region to fit the area where we saw the largest evidence of copy number variation and then calculated the read coverage, read count ratio, and significance. All calculations and visualisations were performed using the data.table and ggplot2 R packages (http://www.r-project.org/).

In summary, CN data from 89 non-syndromic CHD cases were analysed using the CytoScan HD microarray platform, and CNVs of interest were investigated and validated through bioinformatic analyses. The following chapter will show how the above methods were applied to patient samples and what results were obtained.

Chapter 3. Results

3.1. The study cohort

Patients presenting to the Cardiac clinics at RCWMCH and GSH for cardiac care at various ages were consented to participate in a study to investigate the genetic basis of CHD. At the time of writing, the PROTEA biorepository consisted of over 500 participants, and CMA data was available for 105 of them. All 105 study subjects underwent clinical genetic evaluation and were eligible for the study if they presented with an isolated heart defect (n = 76) or CHD with additional ECAs (n = 17) but did not have a specific genetic diagnosis or recognized chromosomal aberration detected by standard techniques. A total of 93 patients met these criteria and were included in downstream molecular analyses.

Twelve additional patients were ascribed to known CHD-related genetic and/or chromosomal abnormalities including T21 syndrome (n = 5), 22q11.2 deletion syndrome (n = 2), Goldenhar syndrome (n = 1), foetal alcohol syndrome (n = 1), Klinefelter syndrome (n = 1), and 13q deletion (n = 1). These patients were classified as syndromic CHD cases and used as positive controls for the microarray and subsequent bioinformatic analyses (Table 3.1).

Table 3.1. Demographic characteristics of the study cohort

Demographic characteristics	*Isolated CHD (n = 76)*	*CHD with ECAs (n = 17)*	*Positive controls (n = 12)*	*Total (n = 105)*
Age at enrollment (years)				
Median	4	11	1	4
IQR	1 – 9,5	3 – 16	0 – 1,25	1 – 11
Sex				
Male (%)	30 (39,47%)	11 (64,71%)	6 (50%)	47 (44,76%)
Female (%)	46 (60,53%)	6 (35,29%)	6 (50%)	58 (55,24%)
Ethnicity				
Black African (%)	25 (32,89%)	5 (29,41%)	3 (25%)	33 (31,43%)
Caucasian (%)	2 (2,63%)	0 (0%)	0 (0%)	2 (1,90%)
Mixed Ancestry (%)	49 (64,47%)	12 (70,59%)	9 (75%)	70 (66,67%)

Note. Values are rounded off to two decimal points. CHD – congenital heart disease; ECA – extra-cardiac anomaly; IQR – interquartile range

The study cohort consisted of 58 females (55,24%) and 47 males (44,76%) as summarised in Table 3.1. The self-reported ethnicity of the cohort was predominantly of Cape mixed ancestry (66,67%), and also included Black Africans (31,43%) and Caucasians (1,9%). Similarly, the preponderance of patients of Cape mixed ancestry was observed in the cohorts with isolated CHD (64,47%) and CHD with ECAs (70,59%).

3.1.1. Clinical characteristics of the South African cohort

The study patients presented with a wide spectrum of cardiac phenotypes and were classified into sub-phenotypic groups as listed in Table 3.2, based on the method by Botto et al. (2007). The most frequent CHD diagnoses included TOF (n = 13), VSD (n = 11), DORV (n = 11) and PDA (n = 10).

Table 3.2. Phenotypic characteristics of 105 CHD study subjects based on the classification system by Botto et al. (2007) *

CHD phenotype	*Number of patients*
Non-syndromic CHD cases	*93*
Conotruncal defects (n =29)	
TOF	13
TGA	4
DORV	11
TrA	1
Septal defects (n = 22)	
ASD	8
VSD	11
AVSD	3
LVOTO (n=7)	
CoA	5
AS-Multilevel	
MS	
Cleft MV	2
RVOTO (n=19)	
TA	5
PA	9
PS	3
Pulmonary Artery anomaly	2
PDA	10
Complex (n=6)	6

Table 3.2. Phenotypic characteristics of 93 CHD study subjects based on the classification system by Botto et al. (2007) * Continued…

Syndromic CHD cases	12
Trisomy 21	5
22q11.2 syndrome	2
Noonan syndrome	1
Goldenhar syndrome	1
Klinefelter syndrome	1
Fetal alcohol syndrome	1
13q deletion	1

Note. Acronyms. ASD – atrial septal defect; AS- aortic stenosis; AVSD – atrioventricular septal defect; DORV – double outlet right ventricle; LVOTO – left ventricular outflow tract obstruction; MS – mitral stenosis; PA – pulmonary atresia; PDA – patent ductus arteriosus; PS – pulmonary valve stenosis; RVOTO – right ventricular outflow tract obstruction; TrA – Truncus arteriosus; TGA – transposition of the great arteries; TA – tricuspid atresia; TOF – tetralogy of Fallot; VSD - ventricular septal defect.

3.2. Chromosomal microarray analyses of South African congenital heart disease patients

A total of 105 South African probands diagnosed with CHD were analysed using the CytoScan HD array platform. Of the 105 samples subjected to CMA, 101 samples met the QC criteria (Section 2.4.1) with all the metrics within the acceptable ranges, and four samples (2_00173; 2_00197; 2_00326; 2_00384) failed. Of the four failed samples, three were from isolated CHD patients and one was from a patient with CHD and ECAs. Once CMA was complete, a CN plot was generated for each sample. Figure 3.1 shows a comparison between a sample of high quality and a sample that failed QC.

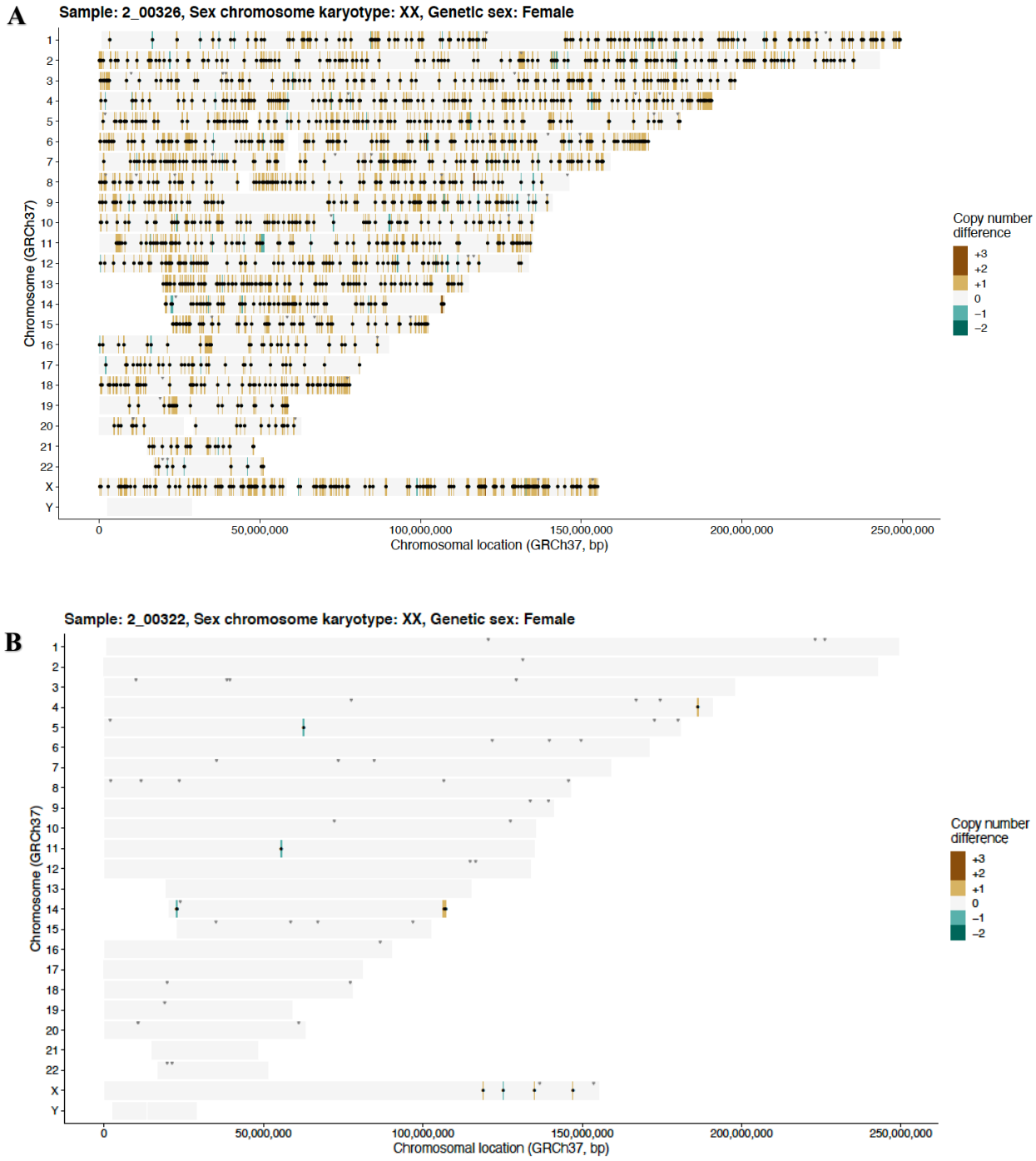

Figure 3.1. Distribution of CNVs for a sample which failed QC compared to a sample with high-quality microarray data

A. One individual (2_00326) shows approximately 1800 singly duplicated regions (all < 1 Mb) distributed across all the autosomes and the X chromosome. This is a clear example of a sample that failed microarray analysis. B. A second individual with a more typical CN plot. The vertical axis represents the different chromosomes, and the horizontal axis indicates the chromosomal coordinates. Each CNV is marked by a black circle to aid visibility and is denoted by a coloured region the size of the CNV. CNVs are coloured according to the difference from standard CN.

3.3. Global analyses of 101 congenital heart disease cases

A microarray profile was obtained for the 101 successfully genotyped DNA samples, including 89 non-syndromic CHD cases and 12 syndromic CHD cases. A karyotype for each study subject was generated to detect whole chromosomal aberrations and confirm cases with a diagnosed genetic disorder (Figure 3.2). The karyotypes revealed a CN state of three on chromosome 21 for five individuals (2_00004; 2_00059; 2_00090; 2_00098; 2_00142). This matched the clinical diagnosis of T21 syndrome for these five patients. One individual (2_00388) presented with a 46-XXY karyotype, which matched the clinical diagnosis of Klinefelter syndrome for this patient.

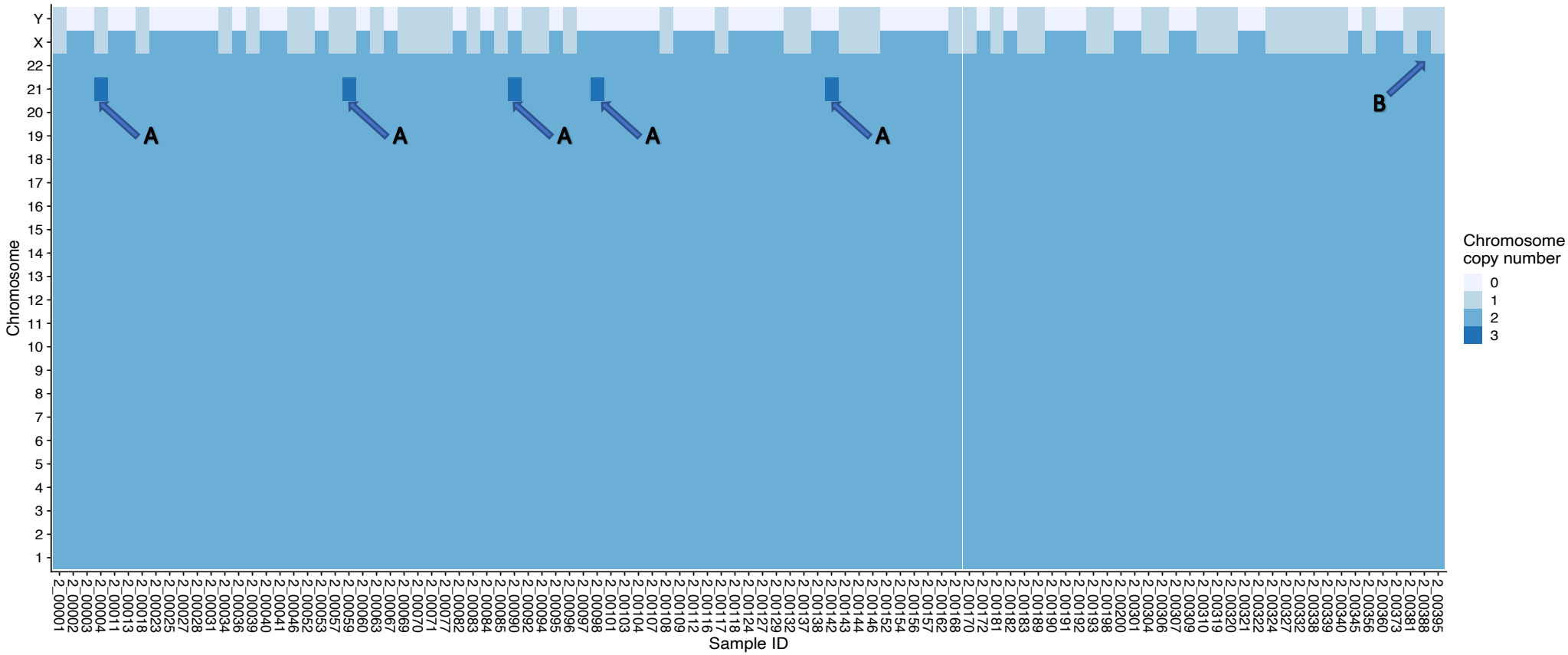

Figure 3.2. The karyotypes of 101 CHD cases

The graphic represents a summarised karyotype for each participant in the study. The anonymized sample identifier of each study subject is represented on the horizontal axis, while the 23 chromosomes that make up the human genome are represented on the vertical axis. A sky-blue filled box represents a normal CN state = 2. An abnormal CN state can be seen by a light-blue filled box (CN state = 1) or a navy filled box (CN = 3). A. Five individuals present with T21 syndrome (CN state = 3). B. One individual presented with a 46-XXY karyotype.

In a second analysis, the sex-chromosomes for each study subject were investigated to confirm the genetic sex of each participant in the study (Figure 3.3). The sex-chromosome plot identified 56 females (XX), 44 males (XY), and one individual with Klinefelter syndrome (XXY). The individual with Klinefelter syndrome is represented by a purple dot, indicating a karyotype of XXY, as expected. The individuals with gross chromosomal abnormalities and/or genetically confirmed syndromes were excluded from subsequent CNV investigations. Therefore, 89 non-syndromic CHD cases were included in downstream bioinformatic analyses.

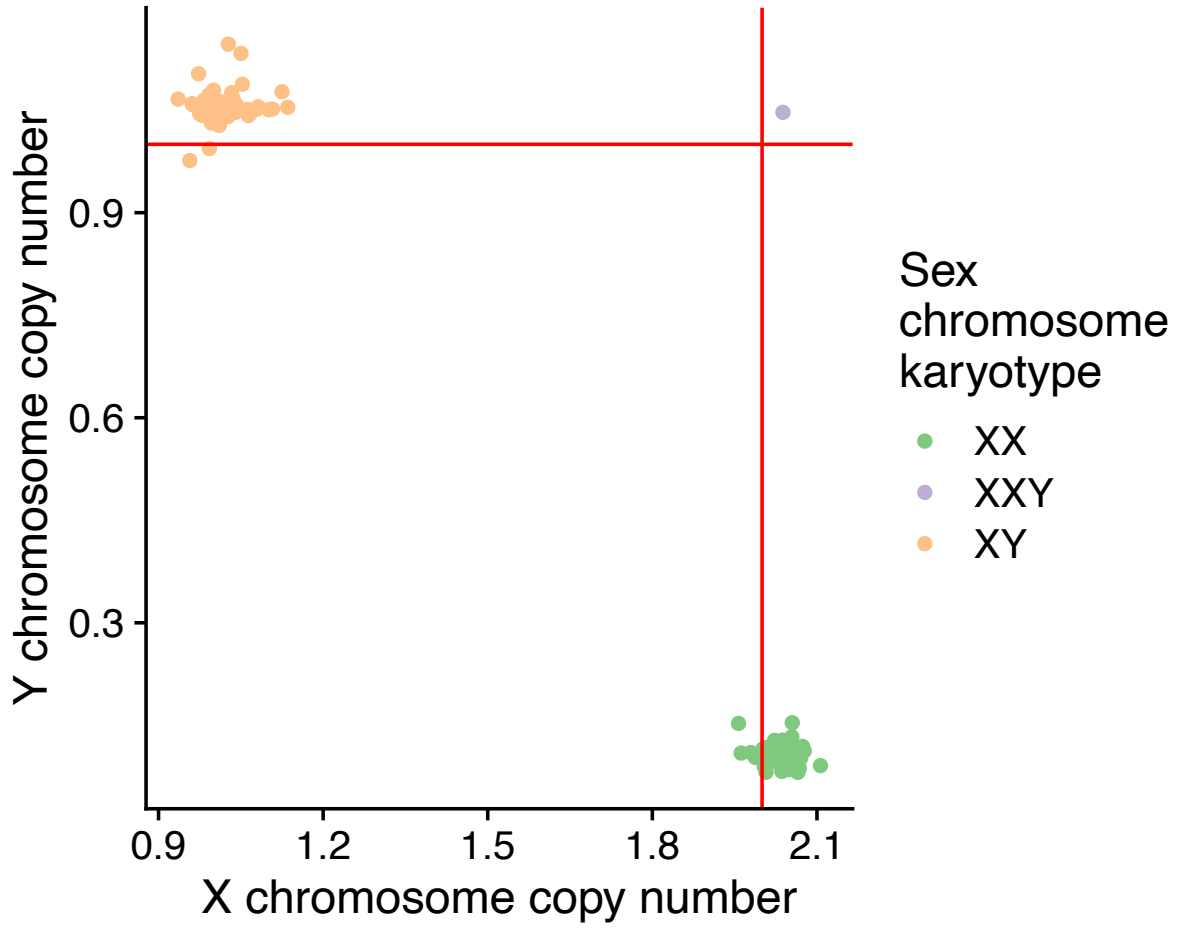

Figure 3.3. Genetic sex-chromosome copy number states for 101 study subjects
This graph illustrates the CN state of the sex chromosomes for 101 study participants. The vertical axis represents the CN state of the Y chromosome. The horizontal axis represents the CN state of the X chromosome. Each dot represents one individual in the study. Male individuals have one Y chromosome (CN = 1) and one X chromosome (CN = 1) depicted in orange. Female individuals have two X chromosomes (CN = 2) and no Y chromosomes (CN = 0) depicted in green. One individual with two X chromosomes and one Y chromosome is depicted in purple.

3.4. Chromosome microarray analysis of 89 non-syndromic congenital heart disease cases

To identify rare CNVs that might increase the risk of CHD development, a CNV analysis was performed using Affymetrix CytoScan HD array data from 89 non-syndromic CHD subjects, of which 73 were CHD cases with isolated heart defects and 16 were CHD cases with ECAs. CN datasets from the 89 non-syndromic CHD cases were extracted from the output Affymetrix CEL files and combined into a text-based format for further investigation. In total, 35 768 CN calls were identified (Figure 3.4). The next step was to identify and remove CN datasets of low quality. At this stage, we were left with 19 333 CN calls. To identify chromosomes with an abnormal CN state, the CN of each autosome and sex chromosome was normalised so that a diploid state would result in a score of zero, as described in the methods section. These scores were filtered to identify chromosomes with an abnormal CN state (CN state ≠ 2). CNVs overlapping known centromeric/telomeric chromosomal regions were removed from the analysis. As a result, 856 CNVs of interest remained.

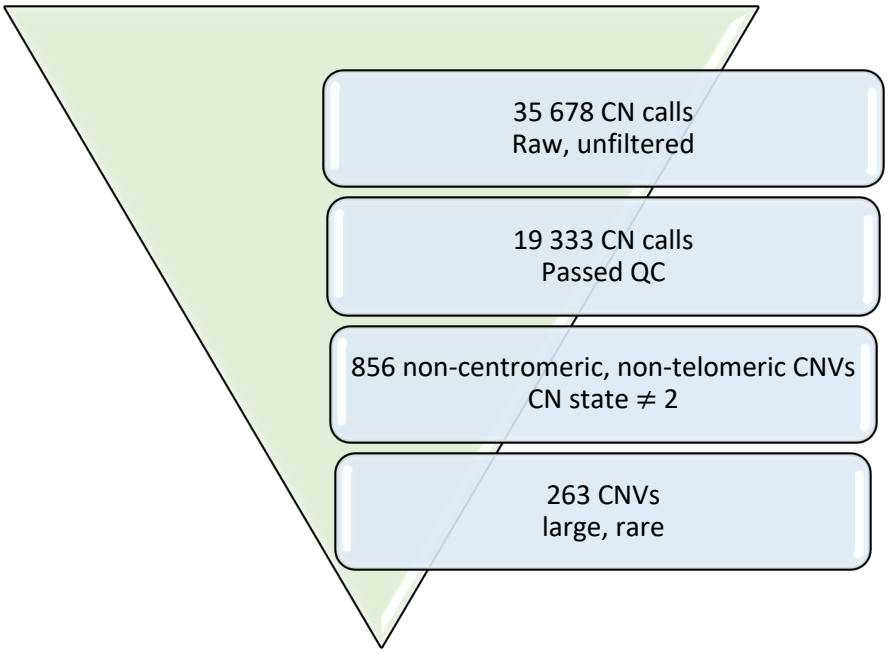

Figure 3.4. Filtering CMA data of 89 non-syndromic CHD cases to identify CNVs of interest
Note. Acronyms: CN – copy number; CNV – copy number variation; QC – quality control

Finally, the CNVs of interest were identified after filtering by size and rarity. CNVs were classified as rare variants when occurring in less than <1% of the UK Biobank control population (472 378 healthy individuals). As a result, 263 CNVs of interest remained. Of the 263 CNVs, 99 microdeletions and 164 microduplications were detected within the South African CHD patient cohort (Figure 3.5).

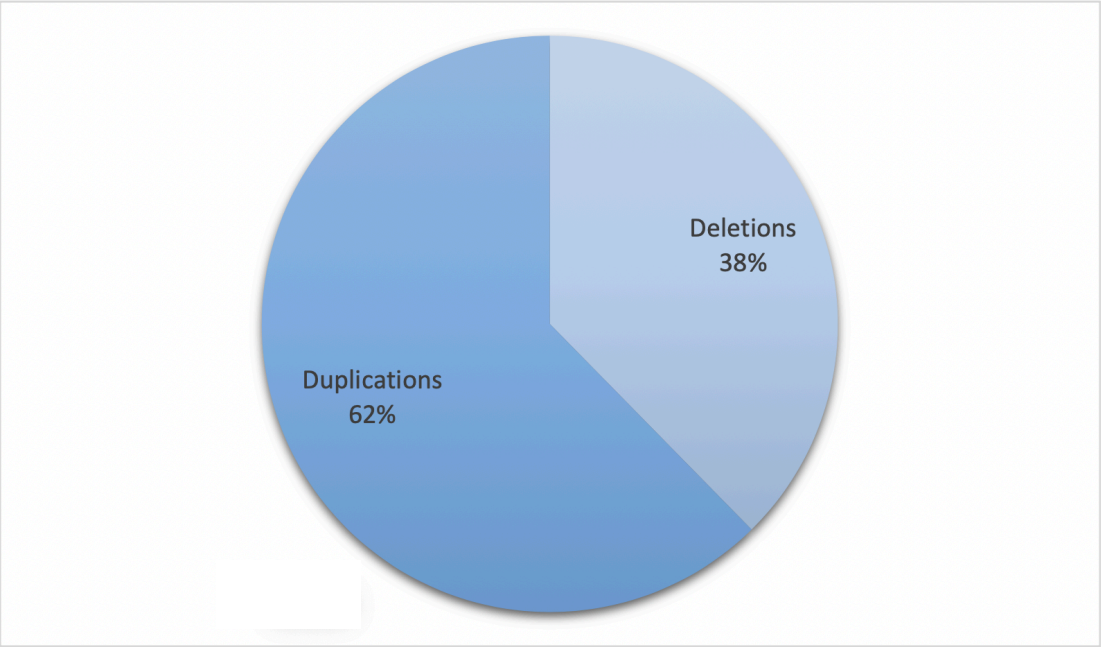

Figure 3.5. Identification of large rare CNVs in South African patients with CHD

In summary, bioinformatic analysis of the microarray data of 89 study subjects led to the identification of 263 large, rare CNVs throughout the genomes of 89 patients ranging in size from 100 kb to 6,1 Mb (10.25375/uct.11953749). Figure 3.6 summarises the complete list of CHD study subjects with large, rare CNVs. An average of two rare CNVs was detected per sample, representing a mean of 1,8 deletions and 2,1 duplications per individual.

The next step of the analysis involved investigating the genes found within the detected CNV regions. An initial analysis focussed on identifying and validating known causative CHD-associated CNVs and genes. Thereafter, analyses focussed on identifying candidate CHD genes that could be potentially pathogenic in the South African CHD cohort.

Figure 3.6. The frequency of CNVs in the cohort per patient

The horizontal axis represents the 89 study subjects within the cohort and their corresponding sample identifier. The vertical axis represents the number of CNVs per patient. Red represents a deletion. Blue represents a duplication.

3.5. Identification of known copy number variants and genes associated with congenital heart disease

To identify genes that have been previously associated with CHD, we investigated CNVs that overlapped genes listed in the Genomics England PanelApp gene panel for non-syndromic CHD. We identified six large, rare CNVs that contained genes known to be causal for CHD in 5/89 (5,62%) patients. Implicated genes included *FLT4, GATA4, CRKL,* and *TBX1* (Table 3.3). When considering genes on our in-house list of CHD-associated genes, two additional CNVs were identified in two patients. Both CNVs represented large microdeletions (Table 3.3), with a <1% frequency in the UK Biobank control population. Implicated genes were *NSD1* and *B3GAT3.* Combined with the PanelApp findings, a total of eight large, rare CNVs overlapping known CHD-associated genes were detected in 6/89 individuals (6,74%). The CNV set comprised 5/8 (62,5%) deletions and 3/8 (37,5%) duplications. These will be explored in more detail below.

Table 3.3. CNVs overlapping with known CHD genes in the study cohort

Genomic region	*Chromosomal coordinates*	*Event*	*Size (kb)*	*Gene symbol*	*Patient ID*	*Ethnicity*	*Patient diagnosis*	*Literature*
CHD Panel								
5q35.3	5:177356690-180679497	Duplication	3322,81	*FLT4*	2_00013	BA	AVSD	(Page et al., 2019; Soemedi et al., 2012a)
5q35.3	5:180035579-180206228	Partial Duplication	170,65	*FLT4*	2_00097	MA	ASD + PS	(Page et al., 2019; Soemedi et al., 2012a)
8p23.1	8:8093066-11877967	Deletion	3784,90	*GATA4*	2_00168	MA	ASD + PS	(Blue et al., 2012)
22q11.21	22:21050078-21800471	Deletion	750,39	*CRKL*	2_00162	MA	VSD	(Tomita-Mitchell et al., 2012)
22q11.21	22:21092382-21465662	Deletion	373,28	*CRKL*	2_00025	MA	TOF	(Tomita-Mitchell et al., 2012)
22q11.21	22:18916828-19784699	Deletion	877,87	*TBX1*	2_00025	MA	TOF	(Fulcoli et al., 2016)
In-house genes								
5q35.3	5:176598618-177198436	Partial Duplication	599,82	*NSD1*	2_00013	BA	AVSD	(Thienpont et al., 2007)
11q12.3	11:61206220-63379913	Deletion	2173,69	*B3GAT3*	2_00018	MA	ASD	(Bloor, Giri, Didi, & Senniappan, 2017)

Note. Values are rounded off to two decimal points. Acronyms. ASD – atrial septal defect; AVSD – atrioventricular septal defect; BA – Black African; CHD – congenital heart disease; ID – identifier; MA – mixed ancestry; PS – pulmonary valve stenosis; TOF – tetralogy of Fallot; VSD – ventricular septal defect.

3.5.1. Deletion of the 8p23.1 region

In one male proband (2_00168), a 3784,90 kb microdeletion was identified at the 8p23.1 cytoband depicted in Figure 3.7. The deletion involves 19 genes in total (Database of genomic variation and phenotype in humans using Ensembl resource - DECIPHER), five of which are classified as morbid genes in Online Mendelian Inheritance in Man (OMIM) including *RP1L1* (OMIM no. 608581), *BLK* (OMIM no. 191305), *FDFT1* (OMIM no. 184420), *CTSB* (OMIM no. 116810), and *GATA4* (OMIM no. 600576). Notably, the 8p23.1 deletion encompasses the well-characterised CHD-associated gene, *GATA4,* which encodes a zinc-finger transcription factor that plays an essential role in mammalian cardiac development (Garg et al., 2003). This deletion was identified in a patient clinically diagnosed with ASD and PS as well as ECAs including intellectual disabilities and dysmorphisms. Similar deletions were reported in several patients in the DECIPHER database (Table 3.4); these deletions were associated with a wide range of CHDs including ASD and ECAs such as intellectual disability. This indicates that the deletion observed in 2_00168 is likely to contribute to that patient's phenotype.

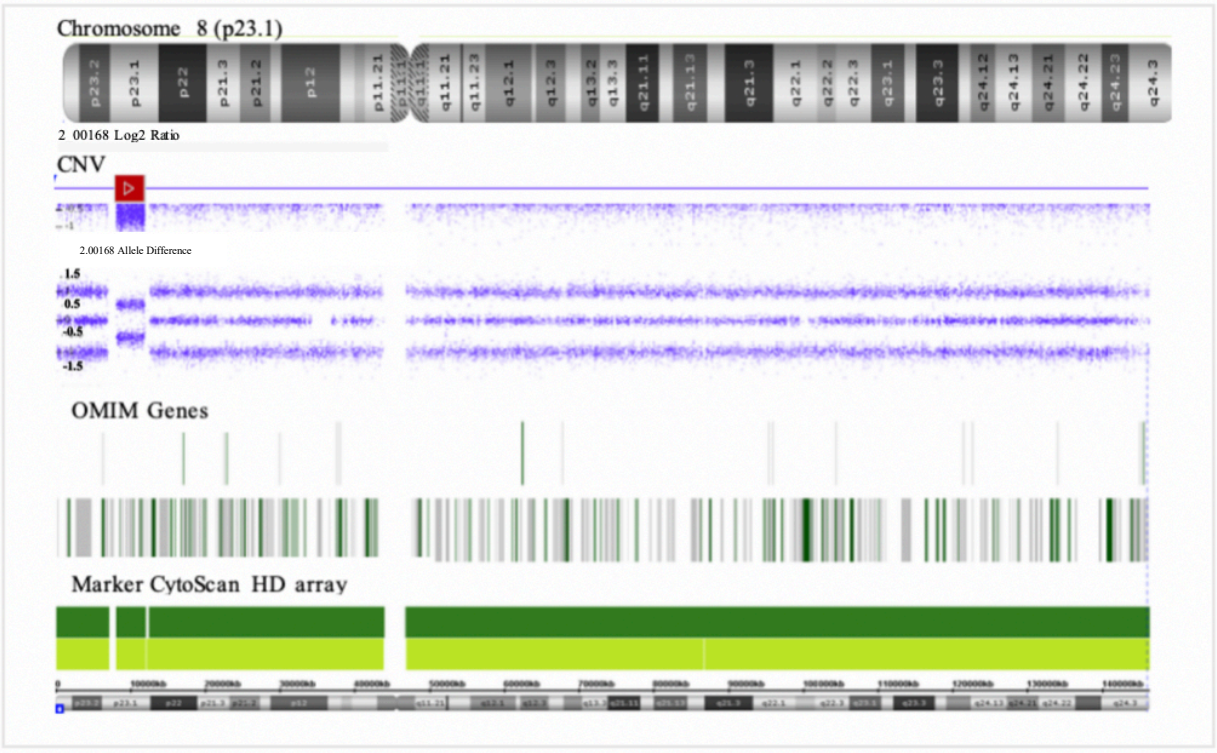

Figure 3.7. Deletion of the 8p23.1 region in male proband 2_00168
CMA from a male proband (2_00168) showing a 3,78 Mb microdeletion at 8p23.1 involving 19 genes. The red block indicates the chromosomal region that has been deleted. This image was generated using the Affymetrix ChAS software (version 4.0)

Table 3.4. Pathogenic deletions overlapping the CNV identified in patient 2_00168

DECIPHER patient ID	*Chromosomal coordinates*	*Size (Mb)*	*Associated CHD phenotypes*	*Nervous system ECAs*
355754	8:8079853-11723203	3,64	ASD, PS	Mild intellectual disability
332789	8:8093065-11888779	3,80	PS, abnormal atrial septal morphology	ADHD, global developmental delay
277490	8:8100384-11860569	3,76	Arteriovenous malformation	Global developmental delay
287741	8:8101641-11898209	3,80	VSD, abnormal heart morphology	Global developmental delay, delayed speech and language development
288613	8:8103647-11852988	3,75	ASD, peripheral PS	None
128	8:8119295-11892064	3,77	PS, atrioventricular canal defect, TAPVR	Hyperactivity, intellectual disability, short attention span
356365	8:8130630-11841901	3,71	AS, atrioventricular canal defect, tachycardia	Agitation, global developmental delay, delayed speech and language development, short attention span
381384	8:8130630-11841901	3,71	Atrioventricular canal defect	None
384421	8:8130630-11841901	3,71	BAV, intermediate atrioventricular canal defect	None
400002	8:8175585-11853852	3,68	Abnormal morphology of myocardial trabeculae	ADHD, motor delay

Note. Values are rounded off to two decimal points. Acronyms. ADHD – attention-deficit hyperactivity disorder; ASD – atrial septal defect; BAV – bicuspid aortic valve; CHD – congenital heart disease; ECA – extra-cardiac anomaly; ID – identifier; PS – pulmonary valve stenosis; TAPVR – total anomalous pulmonary venous return; VSD – ventricular septal defect.

3.5.2. Deletion of the 22q11.21 region

A 750,39 kb microdeletion of the 22q11.21 cytoband was identified in a female proband (2_00162) diagnosed with isolated VSD (Figure 3.8). The microdeletion intersected 12 genes (DECIPHER), four of which were classified as morbid in OMIM: *PI4KA* (OMIM no. 600286), *SERPIND1* (OMIM no. 142360), *SNAP29* (OMIM no. 604202), and *LZTR1* (OMIM no. 600574). The microdeletion also overlaps the *CRKL* gene, which has been associated with cardiac defects including VSD, and 22q11.2 deletion syndrome (Chen, Zhou, Zhang, & Zhang, 2014; Gelb, 2016; Tomita-Mitchell et al., 2012). This CNV is smaller than the 22q11.1 deletion typically associated with 22q11.2 deletion syndrome; however, two pathogenic deletions in the same region have been reported in DECIPHER patients with CHD phenotypes including VSD (Table 3.5). We compared the CN plot for patient 2_00162 to two control patients within the cohort diagnosed with 22q11.2 deletion syndrome (Appendix F).

Table 3.5. Pathogenic deletions overlapping the CNV identified in patient 2_00162

DECIPHER patient ID	*Chromosomal coordinates*	*Size (Mb)*	*Associated CHD phenotypes*	*Nervous system ECAs*
284733	22:21134126-21440514	0,31	Abnormal heart morphology	Moderate intellectual disability
300629	22:20733427-21505417	0,78	VSD	Global developmental delay

Note. Values are rounded off to two decimal points. Acronyms. CHD, congenital heart disease; ECA – extra-cardiac anomaly; ID – identifier; VSD – ventricular septal defect

Additionally, we identified two microdeletions within the 22q11.21 region in a female proband (2_00025) diagnosed with TOF. The first 373,28 kb microdeletion overlaps eight genes, three of which are classified as morbid: *SERPIND1* (OMIM no. 142360), *SNAP29* (OMIM no. 604202), and *LZTR1* (OMIM no. 600574). This CNV also overlaps the adaptor-protein encoding gene, *CRKL*. The second CNV is 867,87 kb in size, overlapping 15 genes (DECIPHER) including four morbid genes: *SLC25A1* (OMIM no. 190315), *CDC45* (OMIM no. 603465), *GP1BB* (OMIM no. 138720), and *TBX1* (OMIM no. 602054). As mentioned previously (Section 1.5.4), the haploinsufficiency of *TBX1* significantly contributes to the clinical phenotype associated with 22q11.2 deletion syndrome (Fulcoli et al., 2016). Additionally, mutations of the *TBX1* gene have been associated with various

CHD phenotypes including TOF, TA, IAA, and VSD (Aburawi et al., 2015; Blue et al., 2012; Pierpont et al., 2018). A third microdeletion, 1293,71 kb in size, was identified in this region during the candidate gene analysis (Section 3.6). These three adjacent CNVs cumulatively delete 2534,86 kb of genomic material and encompass the coordinates 22:18916828-21465662, which makes up most of the 22q11.21 cytoband, deletion of which causes 22q11.2 syndrome (Costain et al., 2016). However, deletions in this region (and including the genes *TBX1* and *CRKL*) have been associated with numerous forms of CHD, as evidenced in the DECIPHER database, including patients with isolated TOF as seen in patient 2_00025 (Table 3.6).

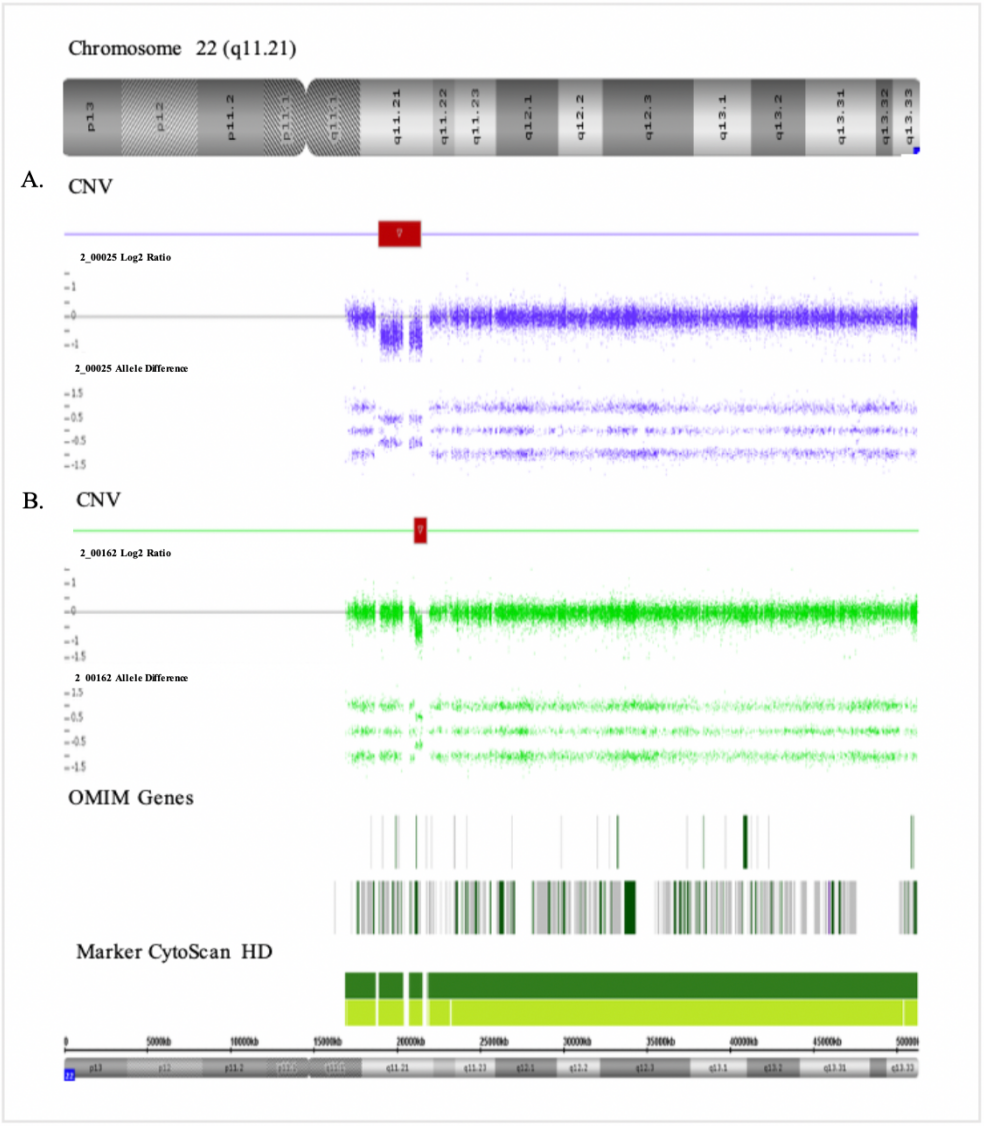

Figure 3.8. Deletion of the 22q11.21 region in female probands 2_00025 and 2_00162

A. CMA from a female proband (2_00025) showing a microdeletion of the 22q11.21 cytoband including CHD-associated genes, *TBX1,* and *CRKL*. B. CMA from a female proband (2_00162) showing a 750,39 kb microdeletion overlapping CHD-associated gene, *CRKL*. The red blocks indicate the deleted regions. This image was generated using the Affymetrix ChAS software (version 4.0)

Table 3.6. Pathogenic deletions overlapping the CNVs identified in patient 2_00025

DECIPHER patient ID	*Chromosomal coordinates*	*Size (Mb)*	*Associated CHD phenotypes*	*Nervous system ECAs*
282275	22:18644790-21798907	3,15	PDA, VSD, vascular ring	Global developmental delay, mild intellectual disability, hyperactivity
331247	22:18646835-21661435	3,01	TOF	None
331294	22:18661699-21661435	3,00	Abnormal aortic arch morphology, abnormal ventricular septum morphology	None
331351	22:18661699-21661435	3,00	Abnormality of cardiovascular system morphology	None
331184	22:18661699-21722313	3,06	Abnormality of cardiovascular system morphology	None
366443	22:18661724-21025713	2,36	VSD	Delayed speech and language development, global developmental delay
360833	22:18844632-21462353	2,62	DORV	None
331301	22:18847961-21499494	2,65	Abnormal aortic morphology	None
331204	22:18875830-21441944	2,57	Abnormal heart morphology	None
259106	22:18889039-21464119	2,58	Abnormal heart morphology	Global developmental delay
300629	22:20733427-21505417	0,77	VSD	Global developmental delay
288305	22:18890162-21441944	2,55	Inlet VSD	None
288045	22:18890162-21540347	2,65	VSD	None
278276	22:18894635-21464260	2,57	Abnormal ventricular septum morphology	Delayed speech and language development, global developmental delay, aggressive behaviour

Table 3.6. Pathogenic deletions overlapping the CNVs identified in patient 2_00025 Continued…

DECIPHER patient ID	*Chromosomal coordinates*	*Size (Mb)*	*Associated CHD phenotypes*	*Nervous system ECAs*
303619	22:18910248-21409634	2,50	VSD	Global developmental delay
356289	22:18916842-21798907	2,88	ASD, VSD, bicuspid aortic valve, interrupted aortic arch	Intellectual disability
315022	22:18916842-21800797	2,88	VSD	Delayed gross motor development
326366	22:18919942-20311763	1,39	TOF	None
338676	22:18919942-21440514	2,52	VSD	None
287304	22:18919942-21801661	2,88	Conotruncal defect	Global developmental delay
362163	22:18967371-21462353	2,49	ASD, VSD	Psychomotor retardation
286124	22:19405375-19792642	0,39	TOF	Delayed speech and language development, global developmental delay

Note. Values are rounded off to two decimal points. Acronyms. ASD – atrial septal defect; CHD – congenital heart disease; ECA – extra-cardiac anomaly; DORV – double outlet right ventricle; ID – identifier; PDA – patent ductus arteriosus; TOF – tetralogy of Fallot; VSD – ventricular septal defect.

3.5.3. Duplication of the 5q35.3 region

The CMA analysis revealed two adjacent microduplications of the 5q35.3 cytoband region in one female proband (2_00013) diagnosed with isolated AVSD (Figure 3.9). The first 599,818 kb microduplication overlaps 19 genes (DECIPHER), four of which are classified as morbid in OMIM including *SLC34A1* (OMIM no. 182309), *F12* (OMIM no. 610619), *DDX41* (OMIM no. 608170), *B4GALT7* (OMIM no. 604327). Additionally, this CNV partially intersects the CHD-associated gene *NSD1* (OMIM no. 6066810). *NSD1* consists of five transcripts, four of which are small (< 100 amino acids) but

contained in the duplicated region. The effects of this are unclear based on current knowledge and this CNV is best regarded as a VOUS. The *NSD1* gene encodes a histone methyltransferase involved in normal growth and development (OMIM). Deletions and duplications of the 5q region encompassing *NSD1* have been associated with TOF, ASD, VSD and AS (Pierpont et al., 2018; Rosenfeld et al., 2012; Tomita-Mitchel el al., 2012).

The second 3,32 Mb microduplication of the 5q35.3 cytoband encompasses 46 genes (DECIPHER), including eight morbid genes, *PROP1* (OMIM no. 601538), *NHP2* (OMIM no. 606470), *PHYKPL* (OMIM no. 614693), *GRM6* (OMIM no. 604096), *ADAMTS2* (OMIM no. 604539), *LTC4S* (OMIM no. 246530), *SQSTM1* (OMIM no. 601530), and the CHD-associated gene *FLT4* (OMIM no. 136352). The *FLT4* gene encodes a receptor kinase called VEGFR3, which is essential for lymphatic development in adults and the development of the cardiovascular system during embryogenesis (Karkkainen et al., 2000). Two patients with similar CNVs were identified in the DECIPHER database: although one was a microdeletion, both included *FLT4* and the 3' region of *NSD1* and both were associated with ASD, VSD, and ECAs (Table 3.7). These variants were classified as likely pathogenic, indicating that CNVs including *FLT4* and partially *NSD1* may contribute to the development of CHD phenotypes.

Table 3.7. Likely pathogenic CNVs overlapping the deletions identified in patient 2_00013

DECIPHER patient ID	*Chromosomal coordinates*	*Size (Mb)*	*Associated CHD phenotypes*	*Nervous system ECAs*
396069	5:176667394-180905260	4,24	ASD, VSD	Delayed speech and language development, intellectual disability
395588*	5:176667394-180905260	4,24	ASD	Intellectual disability

*Note. Values are rounded off to two decimal points. Acronyms. ASD – atrial septal defect; CHD – congenital heart disease; ECA – extra-cardiac anomaly; ID – identifier; VSD – ventricular septal defect. *This variant is a deletion*

A smaller 170,65 kb duplication of the 5q35.3 region was identified in a second female proband within the cohort (2_00097) diagnosed with ASD and PS (Figure 3.9). The duplication intersects with one gene, *OR2Y1,* and partially overlaps the CHD-associated morbid gene, *FLT4*. In this case, only one protein-coding *FLT4* transcript (ENST00000512795) is duplicated. Therefore, the effect of haploinsufficiency at this locus is unclear based on current knowledge and this CNV is best regarded as a VOUS.

Figure 3.9. Duplication of the 5q35.3 region in female probands 2_00013 and 2_00097

A. CMA from a female proband (2_00013) showing a microduplication of the 5q35.3 cytoband including CHD-associated genes, *NSD1,* and *FLT4*. B. CMA from a female proband (2_00097) showing a 170 kb microduplication partially overlapping CHD-associated gene, *FLT4.* The blue blocks indicate the duplicated regions. This image was generated using the Affymetrix ChAS software (version 4.0).

3.5.4. Deletion of the 11q12.3 region

In a male proband (2_00018) clinically diagnosed with ASD with ECAs including intellectual disabilities and dysmorphisms, a 2173,69 kb microdeletion was detected at the 11q12.3 cytoband (Figure 3.10). The microdeletion overlaps eight OMIM morbid genes: *MYRF* (OMIM no. 608329), *BEST1* (OMIM no. 607854), *FTH1* (OMIM no. 134770), *ROM1* (OMIM no. 180721), *GANAB* (OMIM no. 104160), *C11orf83* (OMIM no. 616097), *BSCL2* (OMIM no. 606158) and the CHD-associated gene *B3GAT3* (OMIM no. 606374). Previous studies have associated *B3GAT3* mutations with cardiovascular anomalies and developmental delays (Bloor et al., 2017; Colman et al., 2019; Jones, Schwarze, Adam, Byers, & Mefford, 2015). These include LoF missense, splice site, and nonsense mutations (ClinVar no. VCV000577885), which means that the gene may indeed exhibit haploinsufficiency. Reported CHDs amongst *B3GAT3* mutation carriers include ASD, VSD, PS, and PDA, as well as syndromic phenotypes such as intellectual disability and craniofacial dysmorphism (Colman et al., 2019).

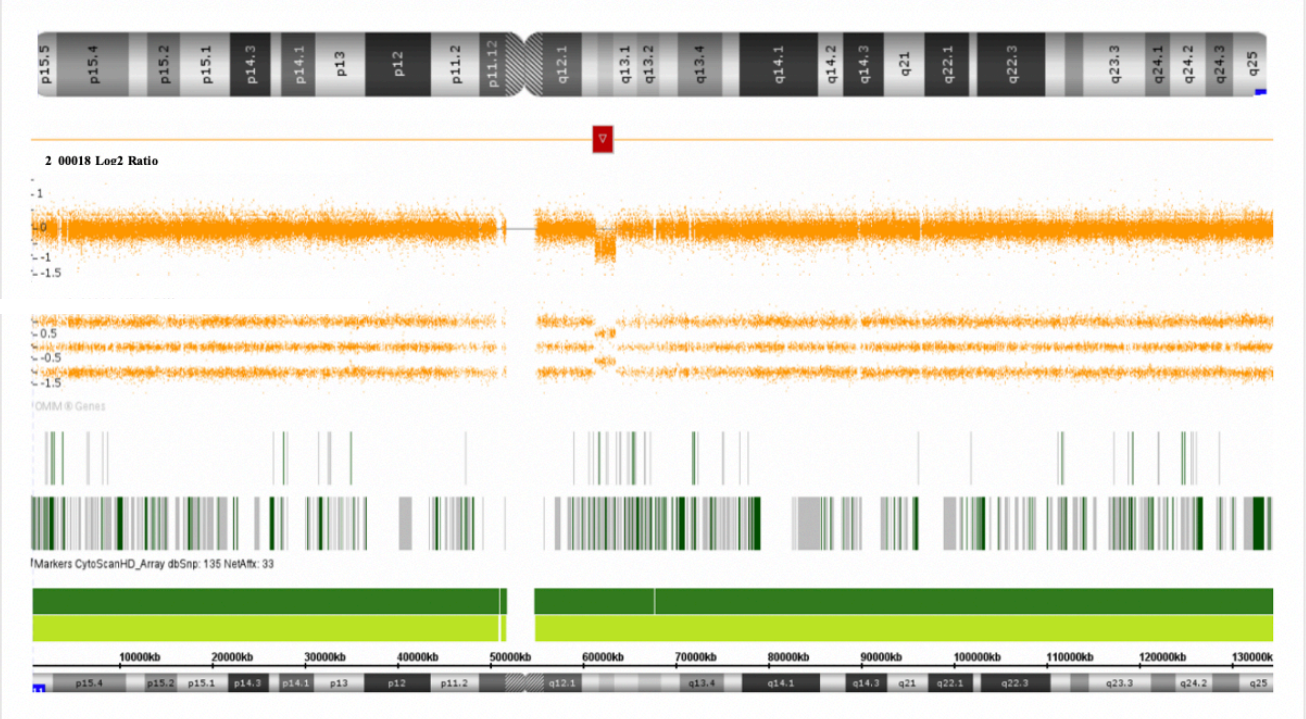

Figure 3.10. Deletion of the 11q12.3 region in male proband 2_00018
CMA from a male proband (2_00018) showing a 2173,69 kb microdeletion at 11q12.3, involving eight morbid genes including CHD-associated gene, *B3GAT3*. The red block indicates the chromosomal region that has been deleted. This image was generated using the Affymetrix ChAS software (version 4.0)

3.6. Identification of candidate genes

Our initial analysis sought to identify CNVs spanning known CHD-associated genes within a South African CHD cohort. Amongst 263 CNV regions of interest, we identified eight CNVs overlapping genes known to be causal for CHD in six probands. Given that a large proportion of CHD genetics is still unaccounted for, and the genetic basis of CHD in a South African cohort has been minimally explored, the next step of the analysis was to investigate the remaining CNVs for candidate genes that may play a role in disease pathogenesis.

To identify candidate CHD genes, we examined the genes within the remaining 255 CNV regions (10.25375/uct.12032202). Starting from 807 genes in duplication CNVs and 474 genes in deletion CNVs, we filtered for protein-coding genes that were likely intolerant of LoF mutations (pLI score ≥ 0,8). This led to the identification of 33 candidate CHD genes including 26 genes in duplication CNVs and seven genes in deletion CNVs (Appendix G). To further refine our list of candidate genes, we next evaluated which genes have recorded expression in the mouse embryonic heart in GXD. Of the 33 candidate CHD genes, 32 genes were found to be expressed in the mouse embryonic heart. The resulting candidate gene list was checked in the MGI database for prior mouse phenotypes. Genes that were not associated with cardiac phenotypes were removed from the analysis. This led to the identification of four genes that may be involved in the development of CHD, namely *DGCR8* (OMIM no. 609030)*, FSTL1* (OMIM no. 605547)*, JARID2* (OMIM no. 601594), and *KDM2A* (OMIM no. 605657) as listed in Table 3.8. *DGCR8* occurred in the 22q11.1 microdeletion region in patient 2_00025 which, as discussed earlier, contains genes that have been linked to CHD before (Section 3.5.2). In contrast, the dosage sensitivity of the other three genes has not been established or associated with CHD. No patients with CHD or other disorders were found to harbour similar CNVs in DECIPHER. However, these genes were identified as CHD candidates due to their high pLI scores combined with their expression in the developing murine heart and the reported cardiac phenotypes in mouse models.

Table 3.8. Four protein-coding candidate genes in South African non-syndromic CHD cases

Gene symbol	*Gene function*	*pLI score*	*Patient ID*	*Ethnicity*	*Chromosomal coordinates*	*Event*	*Size (kb)*	*Patient diagnosis*
DGCR8	*DGCR8* encodes a subunit of the microprocessor complex involved in microRNA biogenesis (McDonald-McGinn et al., 2015)	1,00	2_00025	MA	22:19790229-21083936	Deletion	1293,71	TOF
FSTL1	*FSTL1* encodes a glycoprotein similar to follistatin, a protein that binds activin. This gene has been shown to play a role during embryogenesis and possibly regulates the action of certain growth factors on cell proliferation and differentiation (Chaly, Hostager, Smith, & Hirsch, 2014)	0,96	2_00132	MA	3:120032413-124764110	Duplication	4731,70	VSD + patent foramen ovale, autism spectrum disorder
JARID2	*JARID2* regulates histone methyltransferase complexes which are critical for heart and liver development, haematopoiesis and fusion of the neural tube (Mysliwiec, Bresnick, & Lee, 2011)	1,00	2_00039	MA	6:15283550-16036315	Partial Duplication	752,77	DORV
KDM2A	*KDM2A* is a lysine demethylase of H3 and its function is therefore essential to the histone code (Kawakami, Tokunaga, Ozawa, Sakamoto, & Yoshida, 2015)	1,00	2_00036	MA	11:66832101-66967055	Duplication	134,95	ASD

Note. Values are rounded off to two decimal points. Acronyms. ASD – atrial septal defect; BA – Black African; DORV – double outlet right ventricle; MA – mixed ancestry; TOF – tetralogy of Fallot; VSD – ventricular septal defect; The pLI score shows the probability that the given gene falls into the haploinsufficiency category, and as a result is extremely intolerant of loss-of-function variation. A gene with a high pLI score (pLI ≥ 0.9) is extremely intolerant of loss-of-function mutation. A gene with a low pLI score (pLI ≤ 0.1) can tolerate loss-of-function variation.

3.7. Validation of the identified known and candidate genes

In this study, we identified twelve CNVs of interest, these overlap with eight known CHD-associated genes, as well as four candidate genes. Of the twelve samples displaying CNVs of interest, exome-sequencing data was available for eleven, we used these data to independently validate our CMA findings. Exome-sequencing data was not available for individual 2_00132, so we were unable to validate the CNV overlapping the candidate gene *FSTL1*.

To validate the identified CNVs, we first quantified total aligned read coverage across all samples in the region of interest. We then calculated a read count ratio, to contrast read coverage in the CNV-containing sample with the reference population (patient/reference population) for each CNV of interest (Table 3.9). The read count ratio allows us to determine copy number; for example, a read count ratio of 1.5 indicates a sample presenting with a copy number state 1.5 times higher than the standard state (two), thereby indicating a duplication event. Similarly, a read count ratio of approximately 0.5 indicated a sample presenting with a CN state half that of the normal CN state of two, therefore confirming a deletion event. A Z-test of proportions was used to determine whether the differences in read counts between the patient and reference samples were of statistical significance. Of the eleven samples, nine were considered statistically significant (p-value < 0.05). The CNVs overlapping the CHD gene *FLT4* and candidate gene *JARID2* were unable to be replicated using the exome-sequencing data, although the *FLT4* partial duplication neared statistical significance (p = 0.077). In Figure 3.11, the distribution of reads in the reference population for each CNV is compared to the patient read counts in that region.

Table 3.9. Total read coverage analysis

Chromosomal coordinates	*Event*	*Gene symbol*	*Patient ID*	*Patient read coverage*	*Reference mean read coverage*	*SD CNV reads*	*Read count ratio (patient/reference population)*	*P- value (z-test)*
5:176598618-176727216	Partial Duplication	*NSD1*	2_00013	26376	14394,59	3074,38	1,83	**4.87E-05**
5:177356690-180679497	Duplication	*FLT4*	2_00013	246349	160819,99	31835,03	1,53	**3.61E-03**
11:61206220-63379913	Deletion	*B3GAT3*	2_00018	135920	256053,06	50788,27	0,53	**9.01E-03**
22:18916828-19784699	Deletion	*TBX1*	2_00025	19193	48324,19	10670,55	0,40	**3.17E-03**
22:19790229-21083936	Deletion	*DGCR8*	2_00025	29306	72303,08	16787,96	0,41	**5.22E-03**
22:21050078-21800471	Deletion	*CRKL*	2_00025	13807	35913,09	7928,80	0,38	**2.65E-03**
11:66832101-66967055	Duplication	*KDM2A*	2_00036	2346	1466,70	331,73	1,60	**4.02E-03**
6:15283550-15522252	Partial Duplication	*JARID2*	2_00039	9903	8105.21	1556,681	1.22	1.24E-01
5:180035579-180206228	Partial Duplication	*FLT4*	2_00097	12119	8755,076	2357,483	1,38	7.68E-02
22:21092382-21465662	Deletion	*CRKL*	2_00162	23202	44743,09	10001,59	0,52	**1.56E-02**
8:8093066- 11877967	Deletion	*GATA4*	2_00168	38086	83741,14	17164,60	0,45	**3.91E-03**

Note. Values are rounded off to two decimal points, P-values in bold are significant ($p < 0.05$). CNV – copy number variation; ID – identifier; SD – standard deviation.

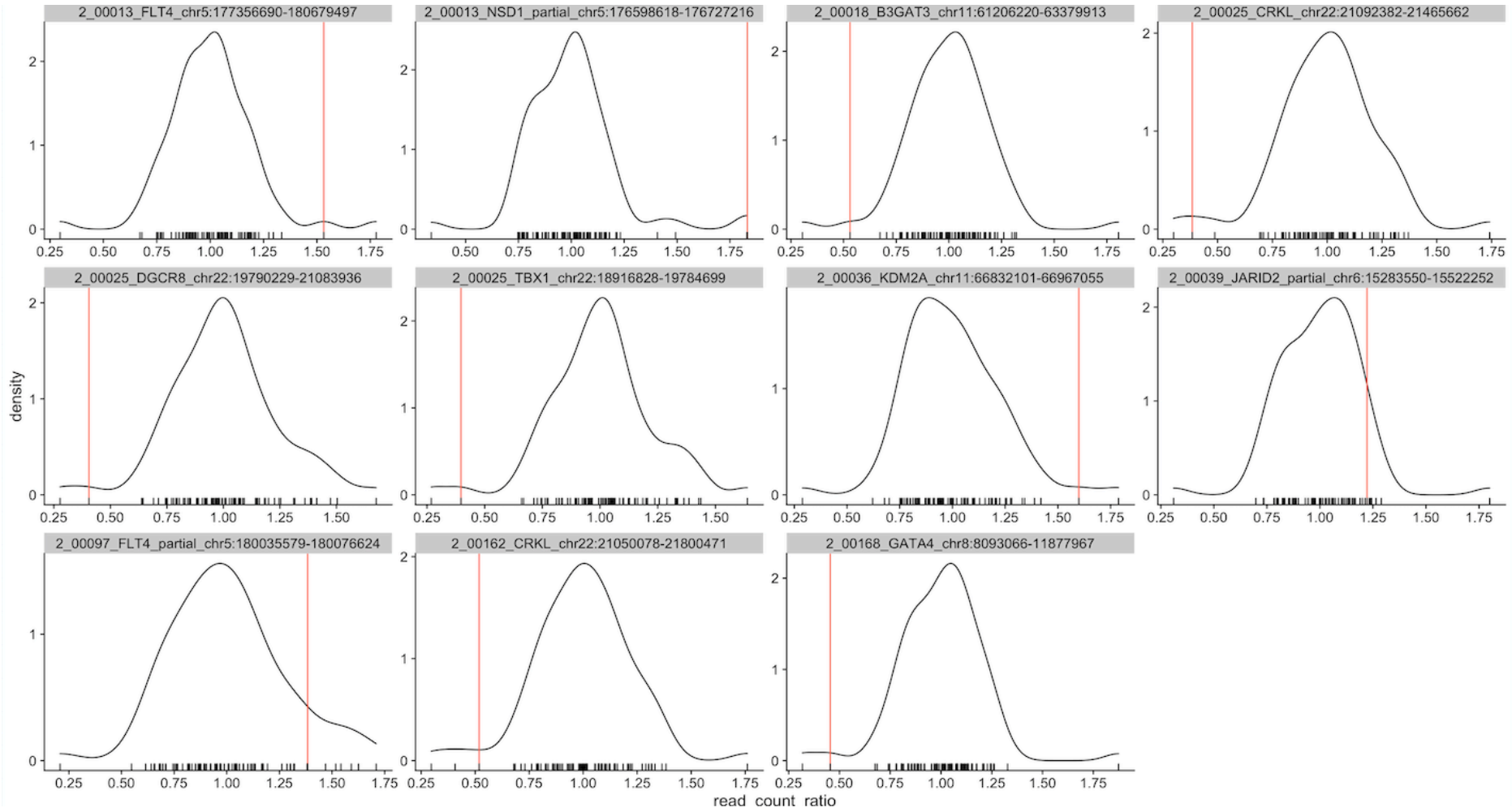

Figure 3.11. Validation density plots of sample read count ratio for the detected CNVs

The data are summarised by a gaussian smoothed kernel density estimate of read count ratio values. The x-axis represents the read count ratio. The y-axis shows the proportion of our observations per unit of the x-axis (on average). The density of read count ratios from the reference population is indicated with a black curve, and the patient of interest is indicated with a red line. Individual read count ratios from the reference population are marked on the x-axis.

In summary, this chapter explored the role of CNV in a South African CHD population. This led to the identification of eight CNVs in six probands encompassing genes known to be causal for CHD. In four probands, the CNVs overlapped similar deletions or duplications that have been previously reported as pathogenic or likely pathogenic for CHD; therefore, the identified CNVs will presumably contribute to the development of CHD in those patients. In a fifth proband, a deletion was identified which overlapped the gene *B3GAT3*, LoF mutations in which have been described in patients with similar phenotypes (Baasanjav et al., 2011; Jones et al., 2015; von Oettingen, Tan, & Dauber, 2014). Therefore, this CNV may too be considered causative of CHD in that patient. In one case, a VOUS was described due to its partial overlap of the CHD gene *FLT4*. We next investigated the genes underlying the remaining 255 CNVs within the cohort and identified four CNVs in an additional three probands overlapping genes likely to play a role in the development of CHD. However, because these genes have not been reported in CHD patients as yet, they should be considered VOUSs at this stage. Combined with the results of the panel analysis, a total of five pathogenic or likely pathogenic CNVs were identified in five patients in the cohort (5.6%) as well as four VOUSs in four additional patients (4.5%). Notably, of the nine probands in which we identified large, rare CNVs, eight are individuals of Cape mixed ancestry and one individual is Black African. No CNVs were identified in Caucasian patients. The results in this chapter indicate that CNVs may contribute to the development of CHD phenotypes in a subset of South African CHD patients. The next chapter will discuss these results in the context of CHD pathogenesis as well as further directions for the research.

Chapter 4. Discussion

CHD is the most common birth defect and a major source of paediatric morbidity and mortality worldwide (Lander & Ware, 2014). Current research indicates that 9 per 1000 children are born with CHD, however, the reported prevalence of CHD in Africa is significantly lower due to poor prognosis, missed diagnoses, and a lack of evidence-based research on the African continent (Liu et al., 2019; Zühlke et al., 2013). CHD is a complex genetic disorder with a poorly understood aetiology, though previous genetic studies have identified CNV as a significant contributor to CHD pathogenesis in individuals of European descent (Fotiou et al., 2019; Geng et al., 2014; Pierpont et al., 2018; Soemedi et al., 2012a; Tomita-Mitchell et al., 2012; Xia et al., 2019). By contrast, the genetic basis of CHD in the South African context has been minimally explored. This study aimed to identify potential pathogenic and likely pathogenic CNVs in a South African population diagnosed with CHD. We found likely causative CNVs covering known CHD-associated genes in 5.6% of our cohort. To our knowledge, this is the first study to investigate the role of CNVs in the development of CHD in an African population.

Our study has three key findings. Firstly, the CytoScan HD array by Affymetrix, the highest performing CMA for detecting chromosomal abnormalities (Wang et al., 2015; Yu et al., 2014), can be used to successfully identify likely causative CNVs involved in CHD pathogenesis. Secondly, we were able to demonstrate that CMA can be performed locally in South Africa, and will yield similar results to international studies. The CNV detection rate of 5.6% in our cohort is in accordance with the internationally reported range of 3% to 25% (Blue et al., 2017; Erdogan et al., 2008; Tomita-Mitchell et al., 2012). Thirdly, the results of the present study highlight the wide genetic heterogeneity of CHD. We identified CNVs overlapping six genes that have been previously associated with CHD, and four additional candidate CHD genes, in nine unrelated probands with a variety of CHD phenotypes including ASD, ASD and PS, AVSD, TOF, DORV, VSD, and ECAs such as autism spectrum disorder and dysmorphisms.

The interpretation of chromosomal structural changes is a significant challenge. Approximately 12% of the human genome consists of frequently occurring CNVs that have no phenotypic consequence (Redon et al., 2009). It is therefore important to distinguish truly causative microduplications and microdeletions from benign CNVs commonly found in the general population. Consequently, the genotype-phenotype relationship is often complex and difficult to establish (Geng et al., 2014). In this study, an average of two rare CNVs was detected per sample, including both likely pathogenic CNVs and VOUSs. CNV pathogenicity is typically determined based on CNV size, gene content, frequency in

healthy individuals, and comparison with peer-reviewed publications and/or public databases (Kearney et al., 2011; Lander & Ware, 2014). We present a comprehensive CNV data curation which involved stringent filtering of CNVs based on their size (≥ 100 kb) and quality, the use of a gene list comprising 157 recognised CHD-associated genes, comparison to a control population of 472 378 healthy individuals, and comparison to patients with similar chromosomal imbalances and disease phenotypes in the literature. Using the above approach, we were able to identify CNVs overlapping genes known to cause CHD, namely *GATA4, TBX1, CRKL, FLT4, NSD1,* and *B3GAT3.* The roles of these genes in CHD will be explored in more detail below.

GATA4 encodes a zinc-finger transcription factor that regulates pathways involved in embryogenesis, heart development, and myocardial differentiation and proliferation (Holtzinger & Evans, 2006). Consequently, mutations in *GATA4* have been linked to cardiac anomalies including ASD and VSD (Garg et al., 2003). The *GATA4* gene is located on the short arm of chromosome 8 at location 23.1 (GRCh38/hg38). Deletions of this region have been associated with microcephaly, growth impairment, intellectual disability, developmental delays, congenital diaphragmatic hernia, and CHD (Claeys et al., 1997; Garg et al., 2003; Wat et al., 2009). Many of these genetic associations were discovered by analysing mice harbouring disease-causing *Gata4* mutations, the mouse ortholog for *GATA4*. Previous studies by Molkentin, Lin, Duncan, and Olson (1997) and Kuo et al. (1997) found that homozygous *Gata4*-deficient mice failed to develop past embryonic stage E10.5 because of severe ventral abnormalities, including the failure to form a linear heart tube. The findings from these studies suggest that *GATA4* plays an essential role in regulating the folding of the embryo that is required for normal cardiac development (Kuo et al., 1997; Molkentin et al., 1997). A CNV analysis by Tomita-Mitchell et al. (2012) detected large, rare CNVs in 35 individuals diagnosed with CHD. Microdeletions of the 8p23.1 cytoband overlapping *GATA4* were enriched in the cohort, presenting in four of the 35 study subjects (11,43%). These patients were diagnosed with VSD, partial AV canal, ASD, sinus venosus, and HLHS (Tomita-Mitchell et al., 2012). Furthermore, Cooper et al. (2012) identified one duplication and three deletions overlapping *GATA4* in their study cohort of 575 CHD patients. In the current study, one microdeletion involving the *GATA4* gene was identified in a patient diagnosed with ASD with PS as well as dysmorphisms. This finding is consistent with previous genetic studies which have associated *GATA4* haploinsufficiency with various CHD phenotypes including ASD, PS, TOF, VSD, DORV as well as neurodevelopmental delay (Aburawi et al., 2015; Blue et al., 2012; Pierpont et al., 2007; Wat et al., 2009). Notably, CNVs involving *GATA4* have not been reported in any of the control groups catalogued in the Database of Genomic Variants (DGV) (Soemedi et al., 2012a). We, therefore, conclude that it is

highly likely that the CHD phenotype seen in proband 2_00168 resulted from the haploinsufficiency of *GATA4.*

The transcription factor *TBX1* is a member of the dosage-sensitive T-Box gene family known to play a critical role in organ and tissue development during embryogenesis (Yagi et al., 2003). Haploinsufficiency of *TBX1* has been commonly described in individuals diagnosed with 22q11.2 deletion syndrome, a well-established chromosomal syndrome with an incidence of 1 per 1000 live births (McDonald-McGinn et al., 2015). The microdeletion syndrome is known to have a variable presentation which often includes immunodeficiency, hypoparathyroidism, and CHD in conjunction with additional congenital abnormalities (McDonald-McGinn et al., 2015). Studies by Lindsay et al. (2001) and Merscher et al. (2001) found that *Tbx1*-deficient mice presented with persistent TrA, cleft palate, and absent thymus and parathyroid glands which led to embryonic lethality. Furthermore, heterozygous LoF *Tbx1* mutations in the mice resulted in cardiac, thymic, and parathyroid defects, similar to those seen in patients with 22q11.2 deletion syndrome. There are 46 known protein-coding genes present in the typical 3 Mb 22q11.2 region that is hemizygously deleted, yet these studies show that *TBX1* plays a critical role during the development of various organ systems including the heart, and indicate that *TBX1* in the region may contribute the CHD phenotype (McDonald-McGinn et al., 2015; Merscher et al., 2001).

Another well-characterised gene implicated in the 22q11.2 region is *CRKL.* The dosage-sensitive gene *CRKL* encodes a signalling protein that is thought to modulate the function of natural killer cells and has been associated with cardiac defects seen in individuals with nested distal 22q11.2 deletions (McDonald-McGinn et al., 2015). Previous mouse studies found that *Crkl* LoF mutations often led to embryonic lethality due to developmental anomalies including CHD (Racedo et al., 2015). A study by Guris, Fantes, Tara, Druker, and Imamoto (2001) found that *CRKL*-null mouse embryos presented with VSDs or conotruncal heart defects similar to those seen in 22q11.2 deletion syndrome patients, suggesting that haploinsufficiency of *CRKL* plays an important role in the development of CHD (Guris et al., 2001). Studies have shown that *TBX1* and *CRKL* interact with one another and compound heterozygosity of the two genes in mice led to an increased penetrance and expressivity of the cardiac defects characteristic of 22q11.2 deletion syndrome (Guris, Duester, Papaioannou, & Imamoto, 2006). Mutations involving *TBX1* and/or *CRKL* have been associated with a spectrum of CHD phenotypes, but are commonly seen in individuals presenting with cardiac outflow tract anomalies including TOF, TrA, TGA, DORV, and VSD (Aburawi et al., 2015; Blue et al., 2012; Pierpont et al., 2018). A CNV analysis by Tomita-Mitchell et al. (2012) identified 35 pathogenic CNVs, two of which overlapped with *TBX1* and

CRKL, and an additional three CNVs which overlapped *CRKL.* These patients presented with a variety of cardiac defects including AVSD with TOF, double-chamber right ventricle, subaortic stenosis, TOF, and VSD (Tomita-Mitchell et al., 2012). Similarly, four of the 17 pathogenic CNVs reported in a CMA study by Xia et al. (2019) involved *TBX1* and *CRKL* (23,53%). Geng et al. (2014) reported 58 pathogenic CNVs, 12 of which were caused by haploinsufficiency of *TBX1* (20,69%). Collectively, these studies highlight the role of *TBX1* and *CRKL* in CHD pathogenesis (Geng et al., 2014; Tomita-Mitchell et al., 2012; Xia et al., 2019). In the current study, we detected two likely-causative microdeletions of the 22q11.2 region in two unrelated individuals. The first CNV overlapped the CHD-associated gene *CRKL* in a proband diagnosed with VSD. This was in line with previous studies that reported individuals presenting with VSD as a result of microdeletions encompassing *CRKL* (Racedo et al., 2015; Tomita-Mitchell et al., 2012). The second CNV overlapped *TBX1* and *CRKL* in a proband diagnosed with TOF. Notably, approximately 15% of TOF patients are diagnosed with 22q11.2 deletion syndrome caused by a *TBX1* mutation, which has led to the classification of *TBX1* as a well-established TOF risk gene (Griffin et al., 2010; Lindsay et al., 2001).

FLT4 encodes VEGFR3 (vascular endothelial growth factor 3), a tyrosine kinase receptor which plays a crucial role in the development of the lymphatic system (Karkkainen et al., 2000; Jin et al., 2017). Mutations in the *FLT4* gene are commonly associated with Milroy's disease, which is characterised by hereditary lymphoedema, commonly in the lower limbs, as a result of congenital anomalies of the lymphatic system (Butler, Dagenais, Rockson, & Glover, 2007). In adults, VEGFR3 expression is mainly restricted to the lymphatic vessels, however during embryogenesis, VEGFR3 is also expressed in vascular endothelial cells and is essential for both angiogenesis and vasculogenesis (Karkkainen et al., 2000; Partanen et al., 2000). Studies by Dumont et al. (1998) and Tammela et al. (2008) found that disrupting the *Flt4* gene in the mouse embryo prevented remodelling of the primary vascular network, leading to lethality at E9.5 as a result of abnormal blood vessel formation and heart failure. A possible role for *FLT4* in the development of CHD has previously been suggested by CNV analysis when a study of 283 TOF probands revealed a female patient with a duplication encompassing the *FLT4* gene (Soemedi et al., 2012a). A WES study by Jin et al. (2017) identified *FLT4* LoF mutations in 2.3% of 426 probands with TOF. Similarly, Page et al. (2019) detected *FLT4* mutations in 2.4% of 829 individuals with TOF, thereby confirming *FLT4* to be a significant contributor to the incidence of TOF. The findings of these studies suggest that both duplications and LoF mutations of the *FLT4* gene can result in CHD. Our study did not find any TOF patients with CNVs encompassing the *FLT* gene. This may be due to the smaller size of the South African cohort compared to those of previous studies. Notably, we detected a duplication of the 5q35.3 region, involving *FLT4* in two unrelated individuals, 2_00013 and 2_00097,

diagnosed with AVSD and ASD with PS respectively. It should be noted that one CNV was not validated, the partial duplication identified in proband 2_00097. Possible explanations for the lack of replication could be that we did not correctly identify the region of the partial duplication, we did not have sufficient coverage of targeted exons in the exome enrichment kit and therefore were unable to detect the duplication with exome-sequencing, or there was no real duplication found in the region. However, the fact that the p-value neared statistical significance ($p = 0.077$), indicates that it could be real but the z-test was underpowered (for example by insufficient coverage) rather than it not being a real duplication. Validating this CNV using the high-throughput MLPA technique could resolve this issue. To the best of our knowledge, this is the first study to report a suggestive association of *FLT4* duplications with AVSD and ASD with PS. Interestingly, the mixed septal defects and PS phenotypes reported in the current study, are similar to two of the four cardiac anomalies that make up TOF, namely VSD and PS (Bailliard & Anderson, 2009). While this is a possible novel association, LoF point mutations of other transcription factors in individuals presenting with TOF, including *Nkx2-5* and *GATA4,* have been reported in ASD and AVSD before (Fahed et al., 2013; Schott et al., 1998).

A second microduplication of the 5q35.3 region overlapping CHD-associated gene *NSD1* was identified in patient 2_00013. *NSD1* is a protein-coding gene that functions as a histone methyltransferase involved in the regulation of the epigenome during growth and development (Tatton-Brown, 2014). LoF mutations of the *NSD1* gene have been associated with Sotos syndrome, a disease characterised by overgrowth, dysmorphisms, and learning disabilities (Tatton-Brown, 2014). In contrast, duplications of the region have been identified in individuals presenting with a reversed phenotype characterised by growth retardation, developmental delay, microcephaly, and CHD (Chen et al., 2006; Dikow et al., 2013). Previous studies have reported microduplications overlapping *NSD1* in individuals presenting with ASD, and AS (Rosenfeld et al., 2012; Tomita-Mitchell et al., 2012). We were able to identify a microduplication of the 5q35.3 region involving *FLT4* and the distal region of *NSD1* in a proband diagnosed with AVSD. Previous studies have reported similar likely causative CNVs in individuals diagnosed with septal defects and ECAs (Section 3.5.3). Taken together, these results suggest that CNVs including *FLT4* and partially *NSD1* may play a role in the development of CHD, and are likely responsible for the cardiac phenotype seen in patient 2_00013.

Finally, *B3GAT3* encodes glucuronyltransferase I (GlcAT-I), an enzyme that plays a critical role in the final step of proteoglycan biosynthesis (Jones et al., 2015). Proteoglycans are an essential component of the extracellular matrix and influence both cell-cell and cell-matrix interactions (von Oettingen et

al., 2014). A study by Izumikawa, Sato, and Kitagawa (2014) found that GlcAT-I deficient mice synthesized smaller chondroitin sulphate and heparan sulphate chains required for differentiation of embryonic stem cells and cytokinesis during development. The resultant failure of cytokinesis in these mice led to embryonic lethality (Izumikawa et al., 2014). Furthermore, disruptions during proteoglycan synthesis can lead to severe developmental disorders, often involving the skeletal and cardiovascular system (Bloor et al., 2017). The functioning of proteoglycans is essential to the cardiovascular system as they are involved in the growth of new blood vessels through interactions with angiogenic factors including VEGF, and influence the differentiation of precursor cells to mature cardiomyocytes (Baasanjav et al., 2011). A previous study by Baasanjav et al. (2011) showed that reduced GlcAT-I activity, caused by a homozygous missense *B3GAT3* mutation (c.830G>A, p.Arg277Gln) resulted in Larsen-like syndrome, a disorder characterised by short stature, multiple joint dislocations, craniofacial dysmorphic features as well as cardiac and/or aortic disease. The missense mutation segregated with the disease in a consanguineous family with five affected children and was not present in 1144 control chromosomes (Baasanjav et al., 2011). In a similar study, von Oettingen et al. (2014) identified the same missense *B3GAT3* mutation in a patient of a different consanguineous family. While both studies found the same missense *B3GAT3* variant in their cohorts, the affected individuals presented with a wide spectrum of cardiac anomalies including BAV, MV prolapse, ASD, patent foramen ovale, VSD, aortic root dilation, and pulmonary artery dilation (Baasanjav et al., 2011; von Oettingen et al., 2014). Additionally, Bloor et al. (2017) identified a pathogenic heterozygous splice site *B3GAT3* mutation in a female born to a nonconsanguineous family, presenting with short stature, facial dysmorphisms, growth hormone deficiency, VSD, and PS amongst other symptoms. In contrast, we identified a microdeletion of the 11q12.3 region involving *B3GAT3* in a male proband, 2_00018. The affected individual presented with ASD and ECAs including dysmorphisms and developmental delay, in accordance with the *B3GAT3* associated-clinical phenotype described in the literature (Baasanjav et al., 2011; Jones et al., 2015; von Oettingen et al., 2014). We, therefore, conclude that it is highly likely that the CHD phenotype and associated ECAs seen in proband 2_00018 resulted from a *B3GAT3* deletion.

While our study identified five CNVs likely to cause disease in five probands, we also considered candidate CHD genes because the majority of patients in the cohort were genotype-negative. As it is likely that other genes are involved in CHD pathogenesis, the investigation of these patients may be useful in the identification of putative new CHD genes. The analysis identified four candidate genes in three additional probands, namely *DGCR8, JARID2, FSTL1,* and *KDM2A.*

DGCR8 encodes a subunit of the microprocessor complex involved in the biogenesis of microRNA, which plays an important role in global gene regulation (McDonald-McGinn et al., 2015; Sellier et al., 2014). *DGCR8* is located within the 22q11.2 cytoband region, and haploinsufficiency of this gene has been associated with 22q11.2 deletion syndrome (Sellier et al., 2014). Previous mouse studies by Stark et al. (2008) found that LoF mutations in *Dgcr8* resulted in neuronal defects similar to those seen in patients diagnosed with 22q11.2 deletion syndrome. Furthermore, inactivating *Dgcr8* in murine cardiac neural crest cells led to various heart defects including persistent TrA and VSD, suggesting that *DGCRC8* is critical for normal patterning of the outflow tract and plays a key role in the development of the heart (Chapnik, Sasson, Blelloch, & Hornstein, 2012). While the entire 22q11.2 region is known to be associated with CHD, traditionally researchers tend to report on the other genes within the region, namely *TBX1* and *CRKL* (Geng et al., 2014; Tomita-Mitchell et al., 2012; Xia et al., 2019). However, previous studies have suggested that there may be a whole set of genes in the region contributing to these phenotypes (McDonald-McGinn et al., 2015). We were able to identify a microdeletion overlapping *DGCR8* in a proband diagnosed with TOF, therefore providing further evidence for the role of *DGCR8* in the development of CHD. These results suggest that the 22q11.2 locus may be more complex than previously thought.

JARID2 is a protein-coding gene that functions as a transcriptional regulator, playing a critical role in embryonic development of the heart and liver, haematopoiesis, and fusion of the neural tube (Hu, Yuan, Rao, Zheng, & Hu, 2014; Kinkel et al., 2015; Mysliwiec et al., 2011). *Jarid2* is expressed throughout the murine embryonic heart and plays an important role in cardiac development through epigenetic regulation of *Notch1* expression (Mysliwiec et al., 2011). Additionally, *JARID2* is one of the many downstream target genes regulated by the well-characterised CHD-associated gene *NKX2.5* during outflow tract morphogenesis (Barth et al., 2010). Previous studies by Lee et al. (2000) investigated the role of *Jarid2* in normal heart development. The study found that homozygous null *Jarid2* mice presented with heart abnormalities, including VSD, DORV, and dilated atria (Lee et al., 2000). A possible role for *JARID2* in CHD pathogenesis has been suggested by WES analysis of 342 individuals with left-sided lesion disorders of the heart, in which one heterozygous LoF mutation involving *JARID2* was identified in an individual presenting with aortic and pulmonic stenosis (Li et al., 2017). Similarly, in a WES study of 182 individuals from 51 families with a history of CHD, Preuss et al. (2016) reported a missense mutation in *JARID2* (c.G1540A:p.G959A) which segregated with BAV in a proband and his father. Furthermore, a study by Liang et al. (2014) suggested a role for *JARID2* in the development of TOF. Liang et al. (2014) found that decreasing miRNA-940 interrupted cardiomyocyte proliferation and migration in the heart by targeting *JARID2* and may lead to cardiac anomalies

including TOF (Liang et al., 2014). The results of these studies support a role for deleterious mutations in *JARID2* as a cause of CHD (Li et al., 2017; Preuss et al., 2016). In our study, we identified a microduplication overlapping *JARID2* in an individual presenting with DORV. We were unable to validate this CNV using exome-sequencing data and further investigation into the variant is required. Although *Jarid2* mutations have been previously associated with DORV in murine models, because this gene has not been reported in many CHD patients as yet, this variant is best considered a VOUS at this stage (Lammer et al., 2009; Lee et al., 2000).

FSTL1 encodes a glycoprotein expressed in mesenchymal cells that plays a vital role during embryonic development (Chaly et al., 2014). The function of FSTL1 is poorly understood, however, previous studies have suggested a role for the protein in the regulation of cell proliferation, differentiation and organ development, and as a cardioprotective factor secreted by the heart (Oshima et al., 2008; Ouchi et al., 2008). Previous mouse studies found that *Fstl1*-deficient mice died at birth as a result of multiple developmental abnormalities of the lungs, skeletal, and respiratory systems (Sylva et al., 2011). Furthermore, conditional deletion of Fstl1 in cells from the endocardial lineage led to cardiac abnormalities including mitral and AV valve defects, which ultimately resulted in heart failure 2-4 weeks after birth (Prakash et al., 2017). In line with these findings, *FSTL1* has a high pLI score of 0,96, suggesting that LoF mutations involving this gene are not tolerated in humans (ExAC). This provides evidence that *FSTL1* may play a role in the development of congenital abnormalities (Prakash et al., 2019). A study by Prakash et al. (2019) investigated the role of *FSTL1* variations in 69 individuals presenting with congenital defects including CHD. However, no pathogenic variants in *FSTL1* were detected in these individuals, suggesting that additional research for the role of *FSTL1* in human disease is required (Prakash et al., 2019). In this study, we identified one microduplication overlapping *FSTL1* in a proband diagnosed with VSD, patent foramen ovale, and autism spectrum disorder. To the best of our knowledge, this is the first study to report a CNV involving *FSTL1* in a patient diagnosed with CHD. However, as this gene has been minimally explored in relation to CHD, this variant is best classified as a VOUS at this stage.

The human *KDM2A* gene is located on chromosome 11q13.2 and encodes the F-box protein FBXL11. FBXL11 is a lysine demethylase of H3k36 that plays a critical role in cell differentiation, proliferation, senescence, and apoptosis (Kawakami et al., 2015). A study by Kawakami et al. (2015) found that mice deficient in *Kdm2a* showed severe growth retardation that led to embryonic lethality during the mid-gestation period. Furthermore, delayed organogenesis was seen in the mutant mice embryos who presented with smaller hearts and cardiac looping abnormalities (Kawakami et al., 2015). The

candidate gene approach used in this study led to the identification of a microduplication overlapping *KDM2A* in an individual presenting with ASD. *KDM2A* is extremely intolerant of LoF mutations with a high pLI score of one (ExAC). Furthermore, *Kdm2a* is widely expressed in the murine heart (GXD) and displays a cardiac phenotype when mutated in mice (MGI), suggesting that pathogenic variants involving this gene may play a role in the development of CHD. To the best of our knowledge, this is the first study to associate a *KDM2A* mutation with CHD.

The inclusion of CMA in CHD management can provide diagnostic information that could alter the way in which CHD patients are managed, by enabling the prediction of recurrent risks and prognostic outcomes for patients before and after surgical interventions (Thomford et al., 2018). An objective of the PROTEA project is to explore the genetic basis of CHD on the African continent, which may improve our understanding of the genotype-phenotype correlations for CHD in Africa. This analysis is the first step in doing so and provides comprehensive genotypic information for a South African CHD cohort, which will contribute toward the establishment of a phenotypic/genotypic registry of African CHD patients in accordance with the aims of PROTEA.

The cohort for this study was made up of 47 males and 58 females aged between 0 and 55 years. While many CHD-genetic studies have focussed on patients of European descent, our study presented a unique CHD cohort comprised predominantly of individuals of Cape mixed ancestry (66,67%) and Black Africans (31,43%). To our knowledge, this is the first CMA analysis of CHD in Cape mixed ancestry and Black African patients. The study sites, RCWMCH and GSH, are public hospitals situated in the Western Cape. It is therefore not surprising that the majority of patients enrolled in this pilot study were of Cape mixed ancestry, as this population has the highest representation in the Western Cape (De Wit et al., 2010). The Cape mixed ancestry population is unique to South Africa and is one of the most admixed populations in the world (De Wit et al., 2010). Therefore, studying and sequencing individuals from this unique population is advantageous, as this may help identify benign mutations in other populations, as well as provide genetic information that can be extrapolated to all world populations. Of the twelve CNVs detected in this study, only two of the CNVs were identified in a Black African patient, and none were in Caucasian patients. This is possibly due to the small sample size of the cohort meaning that very few Black African and Caucasian patients were included in the analyses. However, this may indicate differences in the genetic causes of CHD in Black African patients. These results suggest that this population group should be studied further for potential new causes of disease.

This study has certain limitations that should be discussed. Firstly, as this is a pilot study, it is based on a small sample of participants. However, the results indicate that CMA can be performed locally in

South Africa to study individuals diagnosed with CHD. The analytical methods used in this study can be replicated in the future to investigate the genetics of other patients in the PROTEA registry, which at the time of writing, consisted of over 500 participants. Furthermore, these methods can be expanded to include CHD patients from other African countries to begin to explore the genetic underpinnings of CHD in these populations. At present, data on the genetics of CHD in Africa is scarce, and studies such as these could be important in addressing this imbalance. Secondly, due to a lack of genotype data from the parents of the study subjects, we were unable to determine whether the detected CNVs were *de novo* or inherited. Previous studies have prioritised *de novo* variants, as these CNVs are considered more likely to be pathogenic (Kearney et al., 2011). To overcome this, we employed very stringent filtering criteria to identify the best candidate disease-causing variants. This is another possible avenue for future work that could focus on family studies or functional work using animal models to further investigate the CNVs detected in this study. Finally, an ethnically-matched control group could not be included in the filtering process due to the minimal sequencing of healthy individuals from the Cape mixed ancestry and Black African populations to date. This adds an additional challenge to the already complicated interpretation of CNVs. To overcome this, our pilot study focussed on identifying well-characterised CHD-associated genes, where we could be fairly sure that a detected CNV in our cohort is a true susceptibility variant.

Notably, our study cohort was made up of individuals with a variety of heart defects and CHD phenotypes. Previous genetic studies have investigated homogenous cohorts comprising a single cardiac phenotype (Baban et al., 2014; Page et al., 2019) as well as heterozygous cohorts with various cardiac phenotypes, as seen in our study cohort (Goldmuntz et al., 2011; Soemedi et al., 2012b; Tomita-Mitchell et al., 2012; Xia et al., 2019). Though certain pathogenic variants have been associated with a particular spectrum of malformations, for example 22q11.2 deletion syndrome with outflow tract abnormalities, (Blue et al., 2012; Fahed et al., 2013) and Noonan syndrome with PS/TOF (Aburawi et al., 2015; Lepri et al., 2014; Pierpont et al., 2018), genotype-phenotype correlations in CHD are typically complex. Even the same genetic change can produce different cardiac phenotypes in the same family, highlighting the genetic heterogeneity of CHD (Øyen et al., 2009; Pierpont et al., 2007; Zaidi & Brueckner, 2017). While both genetic approaches have their benefits, previous studies from our collaborating laboratory at the University of Manchester demonstrated that collecting and analysing homogenous patient cohorts can take a considerable amount of time with associated costs. As one of the main aims of this pilot study was to establish that we were able to perform a genetic investigation on CHD patients in South Africa and establish a local bioinformatic pipeline, we decided to follow the

heterozygous cohort approach, enrolling participants into the study as they presented to the cardiac clinics, inclusive of all cardiac phenotypes.

The issue that not enough individuals of non-Western European ancestry have been genotyped in databases such as gnomAD and DGV is an acknowledged problem worldwide, and there is a global effort to increase the amount of sequencing data in other ethnic groups (Bentley, Callier, & Rotimi, 2017). The CMA of individuals of non-European ancestry in the current study may contribute towards the establishment of a South African genetic database that can be used in future genetic studies. A related, ongoing sub-study of PROTEA involves exome-sequencing of the same study cohort. This will indicate the role of single-gene mutations in CHD in the local patient population. Combining the exome and CNV data will give a better understanding of the overall role of genetics in the development of CHD in South African patients, and contribute toward the current body of knowledge surrounding CHD aetiology worldwide.

Our study led to the identification of likely causative CNVs in 5,6% of a South African cohort of probands diagnosed with various types of CHD. Furthermore, we identified four candidate genes that may play a role in the development of CHD and can be further investigated in future studies. This pilot investigation is the first step toward understanding the genetic basis of CHD in African patients and begins to expand upon our understanding of CHD. Studying the genetics of CHD in African populations could help us to identify risk factors specific to the SSA CHD population which could eventually translate to preventative measures to reduce the burden of CVD on the African continent (Thomford et al., 2018).

Chapter 5. Conclusion

The findings of this thesis indicate that CMA is a viable technique for identifying likely-causative variants and that the analytical approach used can be performed locally and produces results similar to those seen in international CHD genetic studies. We anticipate that this study will set the scene for future comprehensive genetic investigations of CHD in African populations while providing key datasets to improve genetic studies of individuals of non-European ancestry worldwide. This thesis underscores the growing importance of CHD genetic studies for both research and clinical purposes. Understanding the genetic causes of CHD will be of great importance to CHD patients and the increasing adult CHD population in terms of family planning, understanding recurrence risk, and exposure to genetic testing and genetic counselling. Finally, advancing our understanding of CHD aetiology will help define disease risk in South Africa and improve the way we care for and treat our cardiac patients.

References

Aburawi, E. H., Aburawi, H. E., Bagnall, K. M., & Bhuiyan, Z. A. (2015). Molecular insight into heart development and congenital heart disease: An update review from the Arab countries. *Trends in Cardiovascular Medicine*, *25*(4), 291–301. https://doi.org/10.1016/j.tcm.2014.11.007

Akhirome, E., Walton, N. A., Nogee, J. M., & Jay, P. Y. (2017). The complex genetic basis of congenital heart defects. *Circulation Journal*, *81*(5), 629–634. https://doi.org/10.1253/circj.CJ-16-1343

American Heart Association. (2018). Care and treatment for congenital heart defects. Retrieved February 29, 2020, from https://www.heart.org/en/health-topics/congenital-heart-defects/care-and-treatment-for-congenital-heart-defects

Ang, S. Y., Uebersohn, A., Ian Spencer, C., Huang, Y., Lee, J. E., Ge, K., & Bruneau, B. G. (2016). KMT2D regulates specific programs in heart development via histone H3 lysine 4 di-methylation. *Development (Cambridge)*, *143*(5), 810–821. https://doi.org/10.1242/dev.132688

Aoki, Y., Niihori, T., Narumi, Y., Kure, S., & Matsubara, Y. (2008). The RAS/MAPK syndromes: Novel roles of the RAS pathway in human genetic disorders. *Human Mutation*, *29*(8), 992–1006. https://doi.org/10.1002/humu.20748

Baasanjav, S., Al-Gazali, L., Hashiguchi, T., Mizumoto, S., Fischer, B., Horn, D., … Hoffmann, K. (2011). Faulty initiation of proteoglycan synthesis causes cardiac and joint defects. *American Journal of Human Genetics*, *89*(1), 15–27. https://doi.org/10.1016/j.ajhg.2011.05.021

Baban, A., Postma, A. V., Marini, M., Trocchio, G., Santilli, A., Pelegrini, M., … Pongiglione, G. (2014). Identification of TBX5 mutations in a series of 94 patients with Tetralogy of Fallot. *American Journal of Medical Genetics, Part A*, *164*(12), 3100–3107. https://doi.org/10.1002/ajmg.a.36783

Bailliard, F., & Anderson, R. H. (2009). Tetralogy of Fallot. *Orphanet Journal of Rare Diseases*, *4*(1). https://doi.org/10.1186/1750-1172-4-2

Barth, J. L., Clark, C. D., Fresco, V. M., Knoll, E. P., Lee, B., Argraves, W. S., & Lee, K.-H. (2010). Jarid2 is among a set of genes differentially regulated by Nkx2.5 during outflow tract morphogenesis. *Developmental Dynamics*, *239*(7), 2024–2033. https://doi.org/10.1002/dvdy.22341

Bassili, A., Mokhtar, S. A., Dabous, N. I., Zaher, S. R., Mokhtar, M. M., & Zaki, A. (2000). Risk factors for congenital heart diseases in Alexandria, Egypt. *European Journal of Epidemiology*, *16*(9), 805–814. https://doi.org/10.1023/A:1007601919164

Battaglia, A., Doccini, V., Bernardini, L., Novelli, A., Loddo, S., Capalbo, A., … Carey, J. C. (2013). Original article Confirmation of chromosomal microarray as a first- tier clinical diagnostic test for individuals with developmental delay , intellectual disability , autism spectrum disorders and dysmorphic features. *European Journal of Paediatric Neurology*, *17*(6), 589–599. https://doi.org/10.1016/j.ejpn.2013.04.010

Becker, S. M., Halees, Z. Al, Molina, C., & Paterson, R. M. (2001). Consanguinity and congenital heart disease in Saudi Arabia. *American Journal of Medical Genetics*, *99*(1), 8–13. https://doi.org/10.1002/1096-8628(20010215)99:1<8::AID-AJMG1116>3.0.CO;2-U

Benhaourech, S., Drighil, A., & El Hammiri, A. (2016). Congenital heart disease and down syndrome: Various aspects of a confirmed association. *Cardiovascular Journal of Africa*, *27*(5), 287–290. https://doi.org/10.5830/CVJA-2016-019

Benson, D. W., Silberbach, G. M., Kavanaugh-McHugh, A., Cottrill, C., Zhang, Y., Riggs, S., … Kugler, J. D. (1999). Mutations in the cardiac transcription factor NKX2.5 affect diverse cardiac developmental pathways. *Journal of Clinical Investigation*, *104*(11), 1567–1573. https://doi.org/10.1172/JCI8154

Bentley, A. R., Callier, S., & Rotimi, C. N. (2017). Diversity and inclusion in genomic research: why the uneven progress? *Journal of Community Genetics*, *8*(4), 255–266. https://doi.org/10.1007/s12687-017-0316-6

Bloor, S., Giri, D., Didi, M., & Senniappan, S. (2017). Novel splicing mutation in B3GAT3 associated with short stature, GH deficiency, hypoglycaemia, developmental delay and multiple congenital anomalies. *Endocrine Abstracts*, *2017*. https://doi.org/10.1530/endoabs.51.p021

Blue, G. M., Kirk, E. P., Sholler, G. F., Harvey, R. P., & Winlaw, D. S. (2012). Congenital heart disease: Current knowledge about causes and inheritance. *Medical Journal of Australia*, *197*(3), 155–159. https://doi.org/10.5694/mja12.10811

Blue, G. M., Kirk, E. P., Giannoulatou, E., Sholler, G. F., Dunwoodie, S. L., Harvey, R. P., & Winlaw, D. S. (2017). Advances in the genetics of congenital heart disease: A clinician's guide. *Journal of the American College of Cardiology*, *69*(7), 859–870. https://doi.org/10.1016/j.jacc.2016.11.060

Botto, L. D., Lin, A. E., Riehle-Colarusso, T., Malik, S., & Correa, A. (2007). Seeking causes: Classifying and evaluating congenital heart defects in etiologic studies. *Birth Defects Research Part A - Clinical and Molecular Teratology*, *79*(10), 714–727. https://doi.org/10.1002/bdra.20403

Butler, M. G., Dagenais, S. L., Rockson, S. G., & Glover, T. W. (2007). A novel VEGFR3 mutation causes Milroy disease. *American Journal of Medical Genetics, Part A*, *143A*(11), 1212-1217 https://doi.org/10.1002/ajmg.a.31703

Cai, M., Huang, H., Su, L., Lin, N., Wu, X., Xie, X., … Xu, L. (2018). Fetal congenital heart disease: Associated anomalies, identification of genetic anomalies by single-nucleotide polymorphism array analysis, and postnatal outcome. *Medicine (United States), 97*(50). https://doi.org/10.1097/MD.0000000000013617

Cavalcante, C. T. de M. B., de Souza, N. M. G., Pinto Júnior, V. C., Branco, K. M. P. C., Pompeu, R. G., Teles, A. C. de O., … de Andrade, G. V. (2016). Analysis of surgical mortality for congenital heart defects using RACHS-1 risk score in a Brazilian single center. *Brazilian Journal of Cardiovascular Surgery*, *31*(3), 219–225. https://doi.org/10.5935/1678-9741.20160022

Chaly, Y., Hostager, B., Smith, S., & Hirsch, R. (2014). Follistatin-like protein 1 and its role in inflammation and inflammatory diseases. *Immunologic Research*, *59*(1-3),266–272. https://doi.org/10.1007/s12026-014-8526-z

Chapnik, E., Sasson, V., Blelloch, R., & Hornstein, E. (2012). Dgcr8 controls neural crest cells survival in cardiovascular development. *Developmental Biology*, *362*(1), 50–56. https://doi.org/10.1016/j.ydbio.2011.11.008

Chen, C.-P., Lin, S.-P., Lin, C.-C., Chen, Y.-J., Chern, S.-R., Li, Y.-C., … Wang, W. (2006). Molecular cytogenetic analysis of de novo dup(5)(q35.2q35.3) and review of the literature of pure partial trisomy 5q. *American Journal of Medical Genetics Part A*, *140A*(14), 1594–1600. https://doi.org/10.1002/ajmg.a.31329

Chen, L., Zhou, W., Zhang, L., & Zhang, F. (2014). Genome Architecture and Its Roles in Human Copy Number Variation. *Genomics & Informatics*, *12*(4), 136. https://doi.org/10.5808/gi.2014.12.4.136

Chung, I. M., & Rajakumar, G. (2016). Genetics of congenital heart defects: The NKX2-5 gene, a key player. *Genes*, *7*(2). https://doi.org/10.3390/genes7020006

Cirstea, I. C., Kutsche, K., Dvorsky, R., Gremer, L., Carta, C., Horn, D., … Zenker, M. (2010). A restricted spectrum of NRAS mutations causes Noonan syndrome. *Nature Genetics*, *42*(1), 27–29. https://doi.org/10.1038/ng.497

Claeys, I., Holvoet, M., Eyskens, B., Adriaensens, P., Gewillig, M., Fryns, J. P., & Devriendt, K. (1997). A recognisable behavioural phenotype associated with terminal deletions of the short arm of chromosome 8. *American Journal of Medical Genetics - Neuropsychiatric Genetics*, *74*(5), 515–520. https://doi.org/10.1002/(SICI)1096-8628(19970919)74:5<515::AID-AJMG12>3.0.CO;2-F

Colman, M., Van Damme, T., Steichen-Gersdorf, E., Laccone, F., Nampoothiri, S., Syx, D., … Malfait, F. (2019). The clinical and mutational spectrum of B3GAT3 linkeropathy: Two case reports and literature review. *Orphanet Journal of Rare Diseases*, *14*(1), 1–10. https://doi.org/10.1186/s13023-019-1110-9

Cooper, G. M., Coe, B. P., Girirajan, S., Rosenfeld, J. a, Vu, T., Williams, C., … Abdel-hamid, H. (2012). Supplementary Information. A Copy Number Variation Morbidity Map of Developmental Delay. *Genetic Counseling*, *43*(9), 838–846. https://doi.org/10.1038/ng.909.A

Cordeddu, V., Di Schiavi, E., Pennacchio, L. A., Ma'ayan, A., Sarkozy, A., Fodale, V., … Tartaglia, M. (2009). Mutation of SHOC2 promotes aberrant protein N-myristoylation and causes Noonan-like syndrome with loose anagen hair. *Nature Genetics*, *41*(9), 1022–1026. https://doi.org/10.1038/ng.425

Costain, G., Silversides, C. K., & Bassett, A. S. (2016). The importance of copy number variation in congenital heart Disease. *Npj Genomic Medicine*, *1*(16031). https://doi.org/10.1038/npjgenmed.2016.31

Darnton-Hill, I., & Mkparu, U. C. (2015). Micronutrients in pregnancy in low- and middle-income countries. *Nutrients*, *7*(3), 1744–1768. https://doi.org/10.3390/nu7031744

De Decker, R., Bruwer, Z., Hendricks, L., Schoeman, M., Schutte, G., Lawrenson., J. (2016). Predicted v . real prevalence of the 22q11 . 2 deletion syndrome in children with congenital heart disease presenting to Red Cross War Memorial Children's Hospital, South Africa : A prospective study. *South African Medical Journal, 106*(6), 82–86. https://doi.org/10.7196/SAMJ.2016.v106i6.11003

de Ligt, J., Boone, P. M., Pfundt, R., Vissers, L. E. L. M., de Leeuw, N., Shaw, C., … Hehir-Kwa, J. Y. (2014). Platform comparison of detecting copy number variants with microarrays and whole-exome sequencing. *Genomics Data*, *2*, 144–146. https://doi.org/10.1016/j.gdata.2014.06.009

De Wit, E., Delport, W., Rugamika, C. E., Meintjes, A., Möller, M., Van Helden, P. D., … Hoal, E. G. (2010). Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape. *Human Genetics*, *128*(2), 145–153. https://doi.org/10.1007/s00439-010-0836-1

Dewan, P., & Gupta, P. (2012). Burden of Congenital Rubella Syndrome (CRS) in India: *Indian Pediatrics*, *49*(5), 377–399. https://doi.org/10.1007/s13312-012-0087-4

Digilio, M. C., Gnazzo, M., Lepri, F., Dentici, M. L., Pisaneschi, E., Baban, A., … Dallapiccola, B. (2017). Congenital heart defects in molecularly proven Kabuki syndrome patients. *American Journal of Medical Genetics, Part A*, *173*(11), 2912–2922. https://doi.org/10.1002/ajmg.a.38417

Dikow, N., Maas, B., Gaspar, H., Kreiss-Nachtsheim, M., Engels, H., Kuechler, A., … Moog, U. (2013). The phenotypic spectrum of duplication 5q35.2-q35.3 encompassing NSD1: Is it really a reversed sotos syndrome? *American Journal of Medical Genetics, Part A*, *161*(9), 2158–2166. https://doi.org/10.1002/ajmg.a.36046

Dixon, J. R., Xu, J., Dileep, V., Zhan, Y., Song, F., Le, V. T., … Yue, F. (2018). Integrative detection and analysis of structural variation in cancer genomes. *Nature Genetics*, *50*(10), 1388–1398. https://doi.org/10.1038/s41588-018-0195-8

Dumont, D. J., Jussila, L., Taipale, J., Lymboussaki, A., Pajusola, K., Breitman, M., … Taipale, J. (1998). Cardiovascular Failure in Mouse Embryos Deficient in VEGF Receptor-3. *American Association for the Advancement of Science, 282*(5390), 946–949. https://doi.org/10.1126/science.282.5390.946

Durocher, D., Charron, F., Warren, R., Schwartz, R. J., & Nemer, M. (1997). The cardiac transcription factors nkx2-5 and GATA-4 are mutual cofactors. *EMBO Journal*, *16*(18), 5687–5696. https://doi.org/10.1093/emboj/16.18.5687

Edwards, J. J., & Gelb, B. D. (2016). Genetics of congenital heart disease. *Current Opinion in Cardiology*, *31*(3), 235–241. https://doi.org/10.1097/HCO.0000000000000274

Erdogan, F., Larsen, L. A., Zhang, L., Tümer, Z., Tummerup, N., Chen, W., … Ullmann, R. (2008). High frequency of submicroscopic genomic aberrations detected by tiling path array comparative genome hybridisation in patients with isolated congenital heart disease. *Journal of Medical Genetics*, *45*(11), 704–709. https://doi.org/10.1136/jmg.2008.058776

Fahed, A. C., Gelb, B. D., Seidman, J. G., & Seidman, C. E. (2013). Genetics of congenital heart disease: The glass half empty. *Circulation Research*, *112*(4), 707–720. https://doi.org/10.1161/CIRCRESAHA.112.300853

Fotiou, E., Williams, S., Martin-Geary, A., Robertson, D. L., Tenin, G., Hentges, K. E., & Keavney, B. (2019). Integration of large-scale genomic data sources with evolutionary history reveals novel genetic loci for congenital heart disease. *Circulation. Genomic and Precision Medicine*, *12*(10), 442–451. https://doi.org/10.1161/CIRCGEN.119.002694

Fox, S. I. (2013). *Human Phsysiology* (13th ed.). New York, NY: McGraw Hill.

Fulcoli, F. G., Franzese, M., Liu, X., Zhang, Z., Angelini, C., & Baldini, A. (2016). Rebalancing gene haploinsufficiency in vivo by targeting chromatin. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms11688

Fuster, V. (2014). Global burden of cardiovascular disease: Time to implement feasible strategies and to monitor results. *Journal of the American College of Cardiology*, *64*(5), 520–522. https://doi.org/10.1016/j.jacc.2014.06.1151

Garg, V., Kathiriya, I. S., Barnes, R., Schluterman, M. K., King, I. N., Butler, C. A., … Srivastava, D. (2003). GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature*, *424*(6947), 443–447. https://doi.org/10.1038/nature01827

Garg, V., Muth, A. N., Ransom, J. F., Schluterman, M. K., Barnes, R., King, I. N., … Srivastava, D. (2005). Mutations in NOTCH1 cause aortic valve disease. *Nature*, *437*(7056), 270–274. https://doi.org/10.1038/nature03940

Gelb, J. J. E. and B. D. (2016). Genetics of congenital heart disease. *Physiology and Behavior*, *176*(5), 139–148. https://doi.org/10.1016/j.physbeh.2017.03.040

Geng, J., Picker, J., Zheng, Z., Zhang, X., Wang, J., Hisama, F., … Shen, Y. (2014). Chromosome microarray testing for patients with congenital heart defects reveals novel disease causing loci and high diagnostic yield. *BMC Genomics*, *15*(1), 1–10. https://doi.org/10.1186/1471-2164-15-1127

Ghorbel, M., Angelini, G., & Caputo, M. (2012). Gene expression profiling – A new approach in the study of congenital heart disease. *Front Lines of Thoracic Surgery, 178-198.* https://doi.org/10.5772/27706

Goldmuntz, E., Paluru, P., Glessner, J., Hakonarson, H., Biegel, J. A., White, P. S., … Shaikh, T. H. (2011). Microdeletions and microduplications in patients with congenital heart disease and multiple congenital anomalies. *Congenital Heart Disease*, *6*(6), 592–602. https://doi.org/10.1111/j.1747-0803.2011.00582.x

Griffin, H. R., Töpf, A., Glen, E., Zweier, C., Stuart, A. G., Parsons, J., … Goodship, J. A. (2010). Systematic survey of variants in TBX1 in non-syndromic tetralogy of Fallot identifies a novel 57 base pair deletion that reduces transcriptional activity but finds no evidence for association with common variants. *Heart*, *96*(20), 1651–1655. https://doi.org/10.1136/hrt.2010.200121

Grossfeld, P. D., Mattina, T., Lai, Z., Favier, R., Jones, K. L., Cotter, F., & Jones, C. (2004). The 11q terminal deletion disorder: A prospective study of 110 cases. *American Journal of Medical Genetics*, *129A*(1), 51–61. https://doi.org/10.1002/ajmg.a.30090

Guris, D. L., Fantes, J., Tara, D., Druker, B. J., & Imamoto, A. (2001). Mice lacking the homologue of the human 22q11.2 gene CRLK phenocopy neurocristopathies of DiGeorge syndrome. *Nature Genetics*, *27*(3), 293–298. https://doi.org/10.1038/85855

Guris, D. L., Duester, G., Papaioannou, V. E., & Imamoto, A. (2006). Dose-dependent interaction of Tbx1 and Crkl and locally aberrant RA signaling in a model of del22q11 syndrome. *Developmental Cell*, *10*(1), 81–92. https://doi.org/10.1016/j.devcel.2005.12.002

Hewitson, J., Brink, J., & Zilla, P. (2002). The challenge of pediatric cardiac services in the developing world. *Seminars in Thoracic and Cardiovascular Surgery*, *14*(4), 340–345. https://doi.org/10.1053/stcs.2002.35298

Hiroi, Y., Kudoh, S., Monzen, K., Ikeda, Y., Yazaki, Y., Nagai, R., & Komuro, I. (2001). Tbx5 associates with Nkx2-5 and synergistically promotes cardiomyocyte differentiation. *Nature Genetics*, *28*(3), 276–280. https://doi.org/10.1038/90123

Hoffman, J. I. E., & Kaplan, S. (2002). The incidence of congenital heart disease. *Journal of the American College of Cardiology*, *39*(12), 1890–1900. https://doi.org/10.1016/S0735-1097(02)01886-7

Holtzinger, A., & Evans, T. (2006). Erratum: Gata4 regulates the formation of multiple organs. *Development*, *132*(17), 4005-2014. https://doi.org/10.1242/dev.02227

Hoosen, E. G. M., Cilliers, A. M., HugoHamman, C. T., Brown, S. C., Lawrenson, J. B., Zuhlke, L., & Hewitson, J. (2011). Paediatric cardiac services in South Africa. *South African Medical Journal*, *101*(2), 106–107. https://doi.org/10.7196/samj.4684

Hu, Z., Yuan, X., Rao, K., Zheng, Z., & Hu, S. (2014). National trend in congenital heart disease mortality in China during 2003 to 2010: A population-based study. *Journal of Thoracic and Cardiovascular Surgery*, *148*(2), 596-602.e1. https://doi.org/10.1016/j.jtcvs.2013.08.067

Izumikawa, T., Sato, B., & Kitagawa, H. (2014). Chondroitin sulfate is indispensable for pluripotency and differentiation of mouse embryonic stem cells. *Scientific Reports*, *4*(1), 1–12. https://doi.org/10.1038/srep03701

Jenkins, K. J., Gauvreau, K., Newburger, J. W., Spray, T. L., Moller, J. H., & Iezzoni, L. I. (2002). Consensus-based method for risk adjustment for surgery for congenital heart disease. *Journal of Thoracic and Cardiovascular Surgery*, *123*(1), 110–118. https://doi.org/10.1067/mtc.2002.119064

Jenkins, K. J. (2004). Risk adjustment for congenital heart surgery: The RACHS-1 method. *Pediatric Cardiac Surgery Annual*, *7*(1), 180–184. https://doi.org/10.1053/j.pcsu.2004.02.009

Jin, S. C., Homsy, J., Zaidi, S., Lu, Q., Morton, S., DePalma, S. R., … Brueckner, M. (2017). Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nature Genetics*, *49*(11), 1593–1601. https://doi.org/10.1038/ng.3970

Jones, K. L., Schwarze, U., Adam, M. P., Byers, P. H., & Mefford, H. C. (2015). A homozygous B3GAT3 mutation causes a severe syndrome with multiple fractures, expanding the phenotype of linkeropathy syndromes. *American Journal of Medical Genetics, Part A*, *167*(11), 2691–2696. https://doi.org/10.1002/ajmg.a.37209

Kamath, B. M., Bauer, R. C., Loomes, K. M., Chao, G., Gerfen, J., Hutchinson, A., … Spinner, N. B. (2012). NOTCH2 mutations in Alagille syndrome. *Journal of Medical Genetics*, *49*(2), 138–144. https://doi.org/10.1136/jmedgenet-2011-100544

Karkkainen, M. J., Ferrell, R. E., Lawrence, E. C., Kimak, M. A., Levinson, K. L., McTigue, M. A., … Finegold, D. N. (2000). Missense mutations interfere with VEGFR-3 signalling in primary lymphoedema. *Nature Genetics*, *25*(2), 153–159. https://doi.org/10.1038/75997

Kawakami, E., Tokunaga, A., Ozawa, M., Sakamoto, R., & Yoshida, N. (2015). The histone demethylase Fbxl11/Kdm2a plays an essential role in embryonic development by repressing cell-cycle regulators. *Mechanisms of Development*, *135*, 31–42. https://doi.org/10.1016/j.mod.2014.10.001

Kearney, H. M., Thorland, E. C., Brown, K. K., Quintero-Rivera, F., & South, S. T. (2011). American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genetics in Medicine*, *13*(7), 680–685. https://doi.org/10.1097/GIM.0b013e3182217a3a

Keates, A. K., Mocumbi, A. O., Ntsekhe, M., Sliwa, K., & Stewart, S. (2017). Cardiovascular disease in Africa: Epidemiological profile and challenges. *Nature Reviews Cardiology*, *14*(5), 273–293. https://doi.org/10.1038/nrcardio.2017.19

Kim, D. S., Kim, J. H., Burt, A. A., Crosslin, D. R., Burnham, N., Kim, C. E., … Jarvik, G. P. (2016). Burden of potentially pathologic copy number variants is higher in children with isolated congenital heart disease and significantly impairs covariate-adjusted transplant-free survival. *Journal of Thoracic and Cardiovascular Surgery*, *151*(4), 1147-1151.e4. https://doi.org/10.1016/j.jtcvs.2015.09.136

Kinkel, S. A., Galeev, R., Flensburg, C., Keniry, A., Breslin, K., Gilan, O., … Blewitt, M. E. (2015). Jarid2 regulates hematopoietic stem cell function by acting with polycomb repressive complex 2. *Blood*, *125*(12), 1890–1900. https://doi.org/10.1182/blood-2014-10-603969

Kopf, P. G., & Walker, M. K. (2009). Overview of developmental heart defects by dioxins, PCBs, and pesticides. *Journal of Environmental Science and Health - Part C Environmental Carcinogenesis and Ecotoxicology Reviews*, *27*(4), 276–285. https://doi.org/10.1080/10590500903310195

Kuo, C. T., Morrisey, E. E., Anandappa, R., Sigrist, K., Lu, M. M., Parmacek, M. S., … Leiden, J. M. (1997). GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. *Genes and Development*, *11*(8), 1048–1060. https://doi.org/10.1101/gad.11.8.1048

Laing, N., Kraus, S. M., Shaboodien, G., & Ntusi, N. A. B. (2019). An overview of the genetic basis of cardiovascular disease. *South African Medical Journal*, *109*(6), 364–370. https://doi.org/10.7196/SAMJ.2019.v109i6.14069

Lammer, E. J., Chak, J. S., Iovannisci, D. M., Schultz, K., Osoegawa, K., Yang, W., … Shaw, G. M. (2009). Chromosomal abnormalities among children born with conotruncal cardiac defects. *Birth Defects Research Part A: Clinical and Molecular Teratology*, *85*(1), 30–35. https://doi.org/10.1002/bdra.20541

Lander, J., & Ware, S. M. (2014). Copy number variation in congenital heart defects. *Current Genetic Medicine Reports*, *2*(3), 168–178. https://doi.org/10.1007/s40142-014-0049-3

Lee, P. Y., Costumbrado, J., Hsu, C. Y., & Kim, Y. H. (2012). Agarose gel electrophoresis for the separation of DNA fragments. *Journal of Visualized Experiments*, (62), 1–5. https://doi.org/10.3791/3923

Lee, Y., Song, A. J., Baker, R., Micales, B., Conway, S. J., & Lyons, G. E. (2000). Jumonji, a nuclear rotein that Is necessary for normal heart development. *Circulation Research*, *86*(9), 932–938. https://doi.org/10.1161/01.RES.86.9.932

Lepri, F. R., Scavelli, R., Digilio, M. C., Gnazzo, M., Grotta, S., Dentici, M. L., … Dallapiccola, B. (2014). Diagnosis of Noonan syndrome and related disorders using target next generation sequencing. *BMC Medical Genetics*, *15*(1), 1–11. https://doi.org/10.1186/1471-2350-15-14

Li, A. H., Hanchard, N. A., Furthner, D., Fernbach, S., Azamian, M., Nicosia, A., … Belmont, J. W. (2017). Whole exome sequencing in 342 congenital cardiac left sided lesion cases reveals extensive genetic heterogeneity and complex inheritance patterns. *Genome Medicine*, *9*(1), 95. https://doi.org/10.1186/s13073-017-0482-5

Li, L., Krantz, I. D., Deng, Y., Genin, A., Banta, A. B., Collins, C. C., … Spinner, N. B. (1997). Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for notch1. *Nature Genetics*, *16*(3), 243–251. https://doi.org/10.1038/ng0797-243

Liang, D., Xu, X., Deng, F., Feng, J., Zhang, H., Liu, Y., … Chen, Y. H. (2014). MiRNA-940 reduction contributes to human Tetralogy of Fallot development. *Journal of Cellular and Molecular Medicine*, *18*(9), 1830–1839. https://doi.org/10.1111/jcmm.12309

Liao, C., Li, R., Fu, F., Xie, G., Zhang, Y., Pan, M., … Li, D. (2014). Prenatal diagnosis of congenital heart defect by genome-wide high-resolution SNP array. *Prenatal Diagnosis*, *34*(9), 858–863. https://doi.org/10.1002/pd.4383

Lindsay, E. A., Vitelli, F., Su, H., Morishima, M., Huynh, T., Pramparo, T., … Baldini, A. (2001). Tbx1 haploinsufficiency in the DiGeorge syndrome region causes aortic arch defects in mice. *Nature*, *410*(6824), 97–101. https://doi.org/10.1038/35065105

Liu, Y., Chen, S., Zühlke, L., Black, G. C., Choy, M. K., Li, N., & Keavney, B. D. (2019). Global birth prevalence of congenital heart defects 1970-2017: Updated systematic review and meta-analysis of 260 studies. *International Journal of Epidemiology*, *48*(2), 455–463. https://doi.org/10.1093/ije/dyz009

Lyu, G., Zhang, C., Ling, T., Liu, R., Zong, L., Guan, Y., … Tao, W. (2018). Genome and epigenome analysis of monozygotic twins discordant for congenital heart disease. *BMC Genomics*, *19*(1), 1-13. https://doi.org/10.1186/s12864-018-4814-7

Malhotra, D., & Sebat, J. (2012). CNVs: Harbingers of a rare variant revolution in psychiatric genetics. *Cell*, *148*(6), 1223–1241. https://doi.org/10.1016/j.cell.2012.02.039

Markbreiter, J., & Philippa, B. (2016). *CVD Advocacy Toolkit: The Road to 2018*. Retrieved November 6, 2019 from https://www.world-heart-federation.org/resources/cvd-advocacy-toolkit-road-2018/

McDaniell, R., Warthen, D. M., Sanchez-Lara, P. A., Pai, A., Krantz, I. D., Piccoli, D. A., & Spinner, N. B. (2006). NOTCH2 mutations cause Alagille syndrome, a heterogeneous disorder of the notch signaling pathway. *American Journal of Human Genetics*, *79*(1), 169–173. https://doi.org/10.1086/505332

McDonald-McGinn, D. M., Sullivan, K. E., Marino, B., Philip, N., Swillen, A., Vorstman, J. A. S., … Bassett, A. S. (2015). 22q11.2 deletion syndrome. *Nature Reviews Disease Primers*, *1*(1), 1–19. https://doi.org/10.1038/nrdp.2015.71

Mensah, G. A., Roth, G. A., & Fuster, V. (2019). The global burden of cardiovascular diseases and risk factors: 2020 and beyond. *Journal of the American College of Cardiology*, *74*(20), 2529–2532. https://doi.org/10.1016/j.jacc.2019.10.009

Merscher, S., Funke, B., Epstein, J. A., Heyer, J., Puech, A., Lu, M. M., … Kucherlapati, R. (2001). TBX1 is responsible for cardiovascular defects in velo-cardio-facial/DiGeorge syndrome. *Cell*, *104*(4), 619–629. https://doi.org/10.1016/S0092-8674(01)00247-1

Miller, D. T., Adam, M. P., Aradhya, S., Biesecker, L. G., Brothman, A. R., Carter, N. P., … Ledbetter, D. H. (2010). Consensus statement: Chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *American Journal of Human Genetics*, *86*(5), 749–764. https://doi.org/10.1016/j.ajhg.2010.04.006

Mlynarski, E. E., Xie, M., Taylor, D., Sheridan, M. B., Guo, T., Racedo, S. E., … Emanuel, B. S. (2016). Rare copy number variants and congenital heart defects in the 22q11.2 deletion syndrome. *Human Genetics*, *135*(3), 273–285. https://doi.org/10.1007/s00439-015-1623-9

Mocumbi, A. O. (2012). Lack of focus on cardiovascular disease in sub-Saharan Africa. *Cardiovascular Diagnosis and Therapy*, *2*(1), 74–747. https://doi.org/10.3978/j.issn.2223-3652.2012.01.03

Molkentin, J. D., Lin, Q., Duncan, S. A., & Olson, E. N. (1997). Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Genes and Development*, *11*(8), 1061–1072. https://doi.org/10.1101/gad.11.8.1061

Muñoz-Chápuli, R., & Pérez-Pomares, J. M. (2010). Cardiogenesis: An embryological perspective. *Journal of Cardiovascular Translational Research, 3*(1), 37-48. https://doi.org/10.1007/s12265-009-9146-1

Mysliwiec, M. R., Bresnick, E. H., & Lee, Y. (2011). Endothelial Jarid2/Jumonji is required for normal cardiac development and proper Notch1 expression. *Journal of Biological Chemistry*, *286*(19), 17193–17204. https://doi.org/10.1074/jbc.M110.205146

Okubo, A., Miyoshi, O., Baba, K., Takagi, M., Tsukamoto, K., Kinoshita, A., … Matsumoto, N. (2004). A novel GATA4 mutation completely segregated with atrial septal defect in a large Japanese family. *Journal of Medical Genetics*, *41*(7), 1–4. https://doi.org/10.1136/jmg.2004.018895

Oshima, Y., Ouchi, N., Sato, K., Izumiya, Y., Pimentel, D. R., & Walsh, K. (2008). Follistatin-Like 1 Is an Akt-regulated cardioprotective factor that Is secreted by the heart. *Circulation*, *117*(24), 3099–3108. https://doi.org/10.1161/CIRCULATIONAHA.108.767673

Ouchi, N., Oshima, Y., Ohashi, K., Higuchi, A., Ikegami, C., Izumiya, Y., & Walsh, K. (2008). Follistatin-like 1, a secreted muscle protein, promotes endothelial cell function and revascularization in ischemic tissue through a nitric-oxide synthase-dependent mechanism. *Journal of Biological Chemistry*, *283*(47), 32802–32811. https://doi.org/10.1074/jbc.M803440200

Øyen, N., Poulsen, G., Boyd, H. A., Wohlfahrt, J., Jensen, P. K. A., & Melbye, M. (2009). Recurrence of congenital heart defects in families. *Circulation*, *120*(4), 295–301. https://doi.org/10.1161/CIRCULATIONAHA.109.857987

Øyen, N., Poulsen, G., Wohlfahrt, J., Boyd, H. A., Jensen, P. K. A., & Melbye, M. (2010). Recurrence of discordant congenital heart defects in families. *Circulation: Cardiovascular Genetics*, *3*(2), 122–128. https://doi.org/10.1161/CIRCGENETICS.109.890103

Page, D. J., Miossec, M. J., Williams, S. G., Monaghan, R. M., Fotiou, E., Cordell, H. J., … Keavney, B. D. (2019). Whole exome sequencing reveals the major genetic contributors to nonsyndromic tetralogy of Fallot. *Circulation Research*, *124*(4), 553–563. https://doi.org/10.1161/CIRCRESAHA.118.313250

Pang, A. W., Macdonald, J. R., Pinto, D., Wei, J., Rafiq, M. A., Conrad, D. F., … Scherer, S. W. (2010). Towards a comprehensive structural variation map of an individual genome. *Genome Biology, 11*(R52). https://doi.org/10.1186/gb-2010-11-5-r52

Parrott, A., & Ware, S. M. (2012). The role of the geneticist and genetic counselor in an ACHD clinic. *Progress in Pediatric Cardiology*, *34*(1), 15–20. https://doi.org/10.1016/j.ppedcard.2012.05.004

Partanen, T. A., Arola, J., Saaristo, A., Jussila, L., Ora, A., Miettinen, M., … Alitalo, K. (2000). VEGF-C and VEGF-D expression in neuroendocrine cells and their receptor, VEGFR-3, in fenestrated blood vessels in human tissues. *The FASEB Journal*, *14*(13), 2087–2096. https://doi.org/10.1096/fj.99-1049com

Pierpont, M. E., Basson, C. T., Benson, D. W., Gelb, B. D., Giglia, T. M., Goldmuntz, E., … Webb, C. L. (2007). Genetic basis for congenital heart defects: Current knowledge - A scientific statement from the American Heart Association Congenital Cardiac Defects Committee, Council on Cardiovascular Disease in the Young. *Circulation*, *115*(23), 3015–3038. https://doi.org/10.1161/CIRCULATIONAHA.106.183056

Pierpont, M. E., Brueckner, M., Chung, W. K., Garg, V., Lacro, R. V., McGuire, A. L., … Russell, M. W. (2018). Genetic basis for congenital heart disease: Revisited: A scientific statement from the American Heart Association. *Circulation*, *138*(21), e653-e711. https://doi.org/10.1161/CIR.0000000000000606

Prakash, S., Borreguero, L. J. J., Sylva, M., Flores Ruiz, L., Rezai, F., Gunst, Q. D., … Van Den Hoff, M. J. B. (2017). Deletion of Fstl1 (Follistatin-Like 1) from the endocardial/endothelial lineage causes mitral valve disease. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *37*(9), e116–e130. https://doi.org/10.1161/ATVBAHA.117.309089

Prakash, S., Mattiotti, A., Sylva, M., Mulder, B. J. M., Postma, A. V., & van den Hoff, M. J. B. (2019). Identifying pathogenic variants in the Follistatin-like 1 gene (*FSTL1*) in patients with skeletal and atrioventricular valve disorders. *Molecular Genetics & Genomic Medicine*, *7*(4), e00567. https://doi.org/10.1002/mgg3.567

Pranckėnienė, L., Preikšaitienė, E., Gueneau, L., Reymond, A., & Kučinskas, V. (2019). De novo duplication in the CHD7 gene associated with severe CHARGE Syndrome . *Genomics Insights*, *12*, 117863101983901. https://doi.org/10.1177/1178631019839010

Preuss, C., Capredon, M., Wünnemann, F., Chetaille, P., Prince, A., Godard, B., … Andelfinger, G. (2016). Family based whole exome sequencing reveals the multifaceted role of Notch signaling in congenital heart disease. *PLOS Genetics*, *12*(10), e1006335. https://doi.org/10.1371/journal.pgen.1006335

Qiagen. (2014). (EN)- Gentra Puregene Handbook- QIAGEN. Retrieved from https://www.qiagen.com/us/resources/resourcedetail?id=a9e6a609-4600-4b03-afbd-974318590ce5&lang=en

Racedo, S. E., McDonald-Mcginn, D. M., Chung, J. H., Goldmuntz, E., Zackai, E., Emanuel, B. S., … Morrow, B. E. (2015). Mouse and human CRKL is dosage sensitive for cardiac outflow tract formation. *American Journal of Human Genetics*, *96*(2), 235–244. https://doi.org/10.1016/j.ajhg.2014.12.025

Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., George, H., Andrews, T. D., … Hurles, M.E. (2009). Group Global variation in copy number in the human genome. *Nature, 444*(7118), 444–454. https://doi.org/10.1038/nature05329

Richards, A. A., Santos, L. J., Nichols, H. A., Crider, B. P., Elder, F. F., Hauser, N. S., … Garg, V. (2008). Cryptic chromosomal abnormalities identified in children with congenital heart disease. *Pediatric Research*, *64*(4), 358–363. https://doi.org/10.1203/PDR.0b013e31818095d0

Robinson, W. M., & Newburger, J. W. (2003). Ethical issues concerning cardiac surgery in patients with syndromic abnormalities. *Seminars in Thoracic and Cardiovascular Surgery. Pediatric Cardiac Surgery Annual*, *6*(1), 147–151. https://doi.org/10.1053/pcsu.2003.50016

Rodriguez-Viciana, P., Tetsu, O., Tidyman, W. E., Estep, A. L., Conger, B. A., Cruz, M. S., … Rauen, K. A. (2006). Germline mutations in genes within the MAPK pathway cause cardio-facio-cutaneous syndrome. *Science*, *311*(5765), 1287–1290. https://doi.org/10.1126/science.1124642

Roos-Hesselink, J. W., Kerstjens-Frederikse, W. S., Meijboom, F. J., & Pieper, P. G. (2005). Inheritance of congenital heart disease. *Netherlands Heart Journal*, *13*(3), 88–91. https://doi.org/10.1007/978-3-319-38913-4_4

Rosenfeld, J. A., Kim, K. H., Angle, B., Troxell, R., Gorski, J. L., Westemeyer, M., … Shaffer, L. G. (2012). Further evidence of contrasting phenotypes caused by reciprocal deletions and duplications: Duplication of NSD1 causes growth retardation and microcephaly. *Molecular Syndromology*, *3*(6), 247–254. https://doi.org/10.1159/000345578

Schott, A. J., Benson, D. W., Basson, C. T., Pease, W., Silberbach, M., Moak, J. P., … Seidman, J. G. (1998). Congenital Heart Disease Caused by Mutations in the Transcription Factor NKX2-5*. American Association for the Advancement of Science, 281(*5373), 108-111*. https://www.jstor.org/stable/2895396 Linked references are available on JSTOR for this article*. *281*(5373), 108–111.

Sellier, C., Hwang, V. J., Dandekar, R., Durbin-Johnson, B., Charlet-Berguerand, N., Ander, B. P., … Tassone, F. (2014). Decreased DGCR8 expression and miRNA dysregulation in individuals with 22q11.2 deletion syndrome. *PLoS ONE*, *9*(8), 13–15. https://doi.org/10.1371/journal.pone.0103884

Shen, W., Paxton, C. N., Szankasi, P., Longhurst, M., Schumacher, J. A., Frizzell, K. A., … Xu, X. (2018). Detection of genome-wide copy number variants in myeloid malignancies using next-generation sequencing. *Journal of Clinical Pathology*, *71*(4), 372–378. https://doi.org/10.1136/jclinpath-2017-204823

Shieh, J. T. C., Bittles, A. H., & Hudgins, L. (2012). Consanguinity and the risk of congenital heart disease. *American Journal of Medical Genetics, Part A*, *158 A*(5), 1236–1241. https://doi.org/10.1002/ajmg.a.35272

Silberbach, M., Roos-Hesselink, J. W., Andersen, N. H., Braverman, A. C., Brown, N., Collins, R. T., ... & Kadian-Dodov, D. (2018). Cardiovascular health in Turner syndrome: a scientific statement from the American Heart Association. *Circulation: Genomic and Precision Medicine*, *11*(10), e000048.

Singh, M. K., Li, Y., Li, S., Cobb, R. M., Zhou, D., Lu, M. M., … Gruber, P. J. (2010). Gata4 and Gata5 cooperatively regulate cardiac myocyte proliferation in mice. *Journal of Biological Chemistry*, *285*(3), 1765–1772. https://doi.org/10.1074/jbc.M109.038539

Soemedi, R., Wilson, I. J., Bentham, J., Darlay, R., Töpf, A., Zelenika, D., … Keavney, B. D. (2012a). Contribution of global rare copy-number variants to the risk of sporadic congenital heart disease. *American Journal of Human Genetics*, *91*(3), 489–501. https://doi.org/10.1016/j.ajhg.2012.08.003

Soemedi, R., Topf, A., Wilson, I. J., Darlay, R., Rahman, T., Glen, E., … Keavney, B. D. (2012b). Phenotype-specific effect of chromosome 1q21.1 rearrangements and GJA5 duplications in 2436 congenital heart disease patients and 6760 controls. *Human Molecular Genetics*, *21*(7), 1513–1520. https://doi.org/10.1093/hmg/ddr589

Stark, K. L., Xu, B., Bagchi, A., Lai, W. S., Liu, H., Hsu, R., … Gogos, J. A. (2008). Altered brain microRNA biogenesis contributes to phenotypic deficits in a 22q11-deletion mouse model. *Nature Genetics*, *40*(6), 751–760. https://doi.org/10.1038/ng.138

Sylva, M., Li, V. S. W., Buffing, A. A. A., van Es, J. H., van den Born, M., van der Velden, S., … van den Hoff, M. J. B. (2011). The BMP Antagonist follistatin-like 1 is required for skeletal and lung organogenesis. *PLoS ONE*, *6*(8), 6–15. https://doi.org/10.1371/journal.pone.0022616

Syrmou, A., Tzetis, M., Fryssira, H., Kosma, K., Oikonomakis, V., Giannikou, K., … Kanavakis, E. (2013). Array comparative genomic hybridization as a clinical diagnostic tool in syndromic and nonsyndromic congenital heart disease. *Pediatric Research*, *73*(6), 772–776. https://doi.org/10.1038/pr.2013.41

Tammela, T., Zarkada, G., Wallgard, E., Murtomäki, A., Suchting, S., Wirzenius, M., … Alitalo, K. (2008). Blocking VEGFR-3 suppresses angiogenic sprouting and vascular network formation. *Nature*, *454*(7204), 656–660. https://doi.org/10.1038/nature07083

Tatton-Brown, K. (2014). *NSD 1* , *EZH 2* and *DNMT 3 A* overgrowth genes and their associated overgrowth syndromes. *eLS*. https://doi.org/10.1002/9780470015902.a0025369

Teteli, R., Uwineza, A., Butera, Y., Hitayezu, J., Murorunkwere, S., Umurerwa, L., … Mutesa, L. (2014). Pattern of congenital heart diseases in Rwandan children with genetic defects. *Pan African Medical Journal*, *19*, 1–10. https://doi.org/10.11604/pamj.2014.19.85.3428

Thienpont, B., Mertens, L., De Ravel, T., Eyskens, B., Boshoff, D., Maas, N., … Devriendt, K. (2007). Submicroscopic chromosomal imbalances detected by array-CGH are a frequent cause of congenital heart defects in selected patients. *European Heart Journal*, *28*(22), 2778–2784. https://doi.org/10.1093/eurheartj/ehl560

Thomford, N. E., Dzobo, K., Yao, N. A., Chimusa, E., Evans, J., Okai, E., … Dandara, C. (2018). Genomics and epigenomics of congenital heart defects : Expert review and lessons learned in Africa. *OMICS: A Journal of Integrative Biology, 22*(5), 301-321. https://doi.org/10.1089/omi.2018.0033

Tomita-Mitchell, A., Mahnke, D. K., Struble, C. A., Tuffnell, M. E., Stamm, K. D., Hidestrand, M., … Mitchell, M. E. (2012). Human gene copy number spectra analysis in congenital heart malformations. *Physiological Genomics*, *44*(9), 518–541. https://doi.org/10.1152/physiolgenomics.00013.2012

Tompson, S. W. J., Ruiz-Perez, V. L., Blair, H. J., Barton, S., Navarro, V., Robson, J. L., … Goodship, J. A. (2007). Sequencing EVC and EVC2 identifies mutations in two-thirds of Ellis-van Creveld syndrome patients. *Human Genetics*, *120*(5), 663–670. https://doi.org/10.1007/s00439-006-0237-7

Truty, R., Paul, J., Kennemer, M., Lincoln, S. E., Olivares, E., Nussbaum, R. L., & Aradhya, S. (2019). Prevalence and properties of intragenic copy-number variation in Mendelian disease genes. *Genetics in Medicine*, *21*(1), 114–123. https://doi.org/10.1038/s41436-018-0033-5

Van Der Linde, D., Konings, E. E. M., Slager, M. A., Witsenburg, M., Helbing, W. A., Takkenberg, J. J. M., & Roos-Hesselink, J. W. (2011). Birth prevalence of congenital heart disease worldwide: A systematic review and meta-analysis. *Journal of the American College of Cardiology*, *58*(21), 2241–2247. https://doi.org/10.1016/j.jacc.2011.08.025

Vermeesch, J. R., Balikova, I., Schrander-Stumpel, C., Fryns, J. P., & Devriendt, K. (2011). The causality of de novo copy number variants is overestimated. *European Journal of Human Genetics*, *19*(11), 1112–1113. https://doi.org/10.1038/ejhg.2011.83

von Oettingen, J. E., Tan, W.-H., & Dauber, A. (2014). Skeletal dysplasia, global developmental delay, and multiple congenital anomalies in a 5-year-old boy-Report of the second family with *B3GAT3* mutation and expansion of the phenotype. *American Journal of Medical Genetics Part A*, *164*(6), 1580–1586. https://doi.org/10.1002/ajmg.a.36487

Walsh, T., McClellan, J. M., McCarthy, S. E., Addington, A. M., Pierce, S. B., Cooper, G. M., … Sebat, J. (2008). Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science*, *320*(5875), 539–543. https://doi.org/10.1126/science.1155174

Wang, J. C., Ross, L., Mahon, L. W., Owen, R., Hemmat, M., Wang, B. T., … Sahoo, T. (2015). Regions of homozygosity identified by oligonucleotide SNP arrays: Evaluating the incidence and clinical utility. *European Journal of Human Genetics*, *23*(5), 663–671. https://doi.org/10.1038/ejhg.2014.153

Wang, X., Li, P., Chen, S., Xi, L., Guo, Y., Guo, A., & Sun, K. (2014). Influence of genes and the environment in familial congenital heart defects. *Molecular Medicine Reports*, *9*(2), 695–700. https://doi.org/10.3892/mmr.2013.1847

Wat, M. J., Shchelochkov, O. A., Holder, A. M., Breman, A. M., Dagli, A., Bacino, C., … Kang, S. H. L. (2009). Chromosome 8p23.1 deletions as a cause of complex congenital heart defects and diaphragmatic hernia. *American Journal of Medical Genetics, Part A*, *149*(8), 1661–1677. https://doi.org/10.1002/ajmg.a.32896

Wessels, M. W., & Willems, P. J. (2010). Genetic factors in non-syndromic congenital heart malformations. *Clinical Genetics*, *78*(2), 103–123. https://doi.org/10.1111/j.1399-0004.2010.01435.x

Wonkam, A., Toko, R., Chelo, D., Tekendo-Ngongang, C., Kingue, S., & Dahoun, S. (2017). The 22q11.2 deletion syndrome in congenital heart defects: Prevalence of microdeletion syndrome in Cameroon. *Global Heart*, *12*(2), 115–120. https://doi.org/10.1016/j.gheart.2017.01.003

World Heart Federation. (2017). Key facts – World Heart Federation. Retrieved December 8, 2019, from https://www.world-heart-federation.org/resources/key-facts/

Wu, X. li, Li, R., Fu, F., Pan, M., Han, J., Yang, X., … Liao, C. (2017). Chromosome microarray analysis in the investigation of children with congenital heart disease. *BMC Pediatrics*, *17*(1), 1–9. https://doi.org/10.1186/s12887-017-0863-3

Xia, Y., Yang, Y., Huang, S., Wu, Y., Li, P., & Zhuang, J. (2019). Corrigendum to: Clinical application of chromosomal microarray analysis for the prenatal diagnosis of chromosomal abnormalities and copy number variations in fetuses with congenital heart disease. *Prenatal Diagnosis*, *38*(6), 406-413. https://doi.org/10.1002/pd.5432

Yagi, H., Furutani, Y., Hamada, H., Sasaki, T., Asakawa, S., Minoshima, S., … Matsuoka, R. (2003). Role of TBX1 in human del22q11.2 syndrome. *Lancet*, *362*(9393), 1366–1373. https://doi.org/10.1016/S0140-6736(03)14632-6

Yu, Y. P., Michalopoulos, A., Ding, Y., Tseng, G., & Luo, J. H. (2014). High fidelity copy number analysis of formalin-fixed and paraffin-embedded tissues using affymetrix cytoscan HD chip. *PLoS ONE*, *9*(4), 1–9. https://doi.org/10.1371/journal.pone.0092820

Yuan, S. M., & Jing, H. (2009). Palliative procedures for congenital heart defects. *Archives of Cardiovascular Diseases*, *102*(6–7), 549–557. https://doi.org/10.1016/j.acvd.2009.04.011

Zaidi, S., Choi, M., Wakimoto, H., Jiang, J., John, D., Romano-Adesman, A., … Lifton,P. L. (2013). De novo mutations in histone modifying genes in congenital heart disease. *Nature, 498*(7453), 220–223. https://doi.org/10.1038/nature12141

Zaidi, S., & Brueckner, M. (2017). Genetics and genomics of congenital heart disease. *Circulation Research*, *120*(6), 923–940. https://doi.org/10.1161/CIRCRESAHA.116.309140

Zarrei, M., Fehlings, D. L., Mawjee, K., Switzer, L., Thiruvahindrapuram, B., Walker, S., … Scherer, S. W. (2018). De novo and rare inherited copy-number variations in the hemiplegic form of cerebral palsy. *Genetics in Medicine*, *20*(2), 172–180. https://doi.org/10.1038/gim.2017.83

Zhang, B., Liang, S., Zhao, J., Qian, Z., Bassig, B. A., Yang, R., … Yang, S. (2016). Maternal exposure to air pollutant PM2.5 and PM10 during pregnancy and risk of congenital heart defects. *Journal of Exposure Science and Environmental Epidemiology*, *26*(4), 422–427. https://doi.org/10.1038/jes.2016.1

Zhang, F., Carvalho, C. M. B., & Lupski, J. R. (2015). Complex human chromosomal and genomic rearrangements The nature of copy number variation and the link with human disease. *Trends in Genetics*. *25*(7), 298–307. https://doi.org/10.1016/j.tig.2009.05.005

Zhang, L., Zhong, T., Wang, Y., Jiang, Q., Song, H., & Gui, Y. (2006). TBX1, a DiGeorge syndrome candidate gene, is inhibited by retinoic acid. *International Journal of Developmental Biology*, *50*(1), 55–61. https://doi.org/10.1387/ijdb.052036lz

Zhao, F., Weismann, C. G., Satoda, M., Pierpont, M. E. M., Sweeney, E., Thompson, E. M., & Gelb, B. D. (2001). Novel TFAP2B mutations that cause Char syndrome provide a genotype-phenotype correlation. *American Journal of Human Genetics*, *69*(4), 695–703. https://doi.org/10.1086/323410

Zhu, H., Kartiko, S., & Finnell, R. H. (2009). Importance of gene - environment interactions in the etiology of selected birth defects. *Clinical Genetics*, *75*(5), 409–423. https://doi.org/10.1111/j.1399-0004.2009.01174.x

Zimmerman, M. S., Smith, A. G. C., Sable, C. A., Echko, M. M., Wilner, L. B., Olsen, H. E., … Kassebaum, N. J. (2020). Global, regional, and national burden of congenital heart disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet Child & Adolescent Health*, *4642*(19). https://doi.org/10.1016/s2352-4642(19)30402-x

Zühlke, L., Mirabel, M., & Marijon, E. (2013). Congenital heart disease and rheumatic heart disease in Africa: Recent advances and current priorities. *Heart*, *99*(21), 1554–1561. https://doi.org/10.1136/heartjnl-2013-303896

Zühlke, L., Lawrenson, J., Comitis, G., Decker, R. De, Brooks, A., Fourie, B., … Hugo-hamman, C. (2019). Congenital Heart Disease in Low- and Lower-Middle – Income Countries : Current Status and New Opportunities. *Current Cardiology Report, 21*(163). https://doi.org/10.1007/s11886-019-1248

Appendices

Appendix A. Ethics approval

UNIVERSITY OF CAPE TOWN
Faculty of Health Sciences
Human Research Ethics Committee

Room E53-46 Old Main Building
Groote Schuur Hospital
Observatory 7925
Telephone [021] 406 6626
Email: shuretta.thomas@uct.ac.za
Website: www.health.uct.ac.za/fhs/research/humanethics/forms

30 May 2019

HREC REF: 339/2019

A/Prof L Zuhlke
Paediatrics & Child Health
Room 3.16, 3rd floor
ICH Building
Red Cross War Memorial Children's Hospital

Dear A/Prof Zuhlke

PROJECT TITLE: INVESTIGATING THE GENETICS OF CONGENITAL HEART DISEASE IN SOUTH AFRICA (MSc Candidate - Ms N A Saacks) SUB-STUDY LINKED TO R017/2014

Thank you for submitting your study to the Faculty of Health Sciences Human Research Ethics Committee.

It is a pleasure to inform you that the HREC has **formally approved** the above-mentioned study.

- Please clarify how 466/2008 is linked to R017/2014.

Approval is granted for one year until 30 May 2020.

Please submit a progress form, using the standardised Annual Report Form if the study continues beyond the approval period. Please submit a Standard Closure form if the study is completed within the approval period.
(Forms can be found on our website: www.health.uct.ac.za/fhs/research/humanethics/forms)

Please quote the HREC REF in all your correspondence.

Please note that the ongoing ethical conduct of the study remains the responsibility of the principal investigator.

Please note that for all studies approved by the HREC, the principal investigator **must** obtain appropriate institutional approval, where necessary, before the research may occur.

The HREC acknowledge that the student, Nicole Aimee Saacks will also be involved in this study.

Yours sincerely

PROFESSOR M BLOCKMAN
CHAIRPERSON, FHS HUMAN RESEARCH ETHICS COMMITTEE

HREC 339/2019

Signature Removed

Federal Wide Assurance Number: FWA00001637.
Institutional Review Board (IRB) number: IRB00001938
NHREC-registration number: REC-210208-007

This serves to confirm that the University of Cape Town Human Research Ethics Committee complies to the Ethics Standards for Clinical Research with a new drug in patients, based on the Medical Research Council (MRC-SA), Food and Drug Administration (FDA-USA), International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use: Good Clinical Practice (ICH GCP), South African Good Clinical Practice Guidelines (DoH 2006), based on the Association of the British Pharmaceutical Industry Guidelines (ABPI), and Declaration of Helsinki (2013) guidelines. The Human Research Ethics Committee granting this approval is in compliance with the ICH Harmonised Tripartite Guidelines E6: Note for Guidance on Good Clinical Practice (CPMP/ICH/135/95) and FDA Code Federal Regulation Part 50, 56 and 312.

HUMAN RESEARCH ETHICS COMMITTEE
12 MAY 2020
HEALTH SCIENCES FACULTY
UNIVERSITY OF CAPE TOWN

UNIVERSITY OF CAPE TOWN | FACULTY OF HEALTH SCIENCES
Human Research Ethics Committee

FHS016: Annual Progress Report / Renewal

HREC office use only (FWA00001637; IRB00001938) This serves as notification of annual approval, including any documentation described below.			
☑ Approved	Annual progress report	Approved until/next renewal date	30.05.2021
☐ Not approved	See attached comments		
Signature Chairperson of the HREC/ Designee	Signature Removed	Date Signed	13/5/2020

Note: Please note that incomplete submissions will not be reviewed.
Please email this form and supporting documents (if applicable) in a combined pdf-file to hrec-enquiries@uct.ac.za.
Please clarify your plan for research-related activities during COVID-19 lockdown

Comments to PI from the HREC

Principal Investigator to complete the following:

1. Protocol Information

Date (when submitting this form)	05-05-2020		
HREC REF Number	339/2019	Current Ethics Approval was granted until	30 May 2020
Protocol title	Investigating the genetics of congenital heart disease in South Africa (MSc Candidate – Ms N A Saacks) Sub-study linked to R017/2014		
Protocol number (If applicable)			
Are there any sub-studies linked to this study?	☐ Yes X No		
If yes, could you please provide the HREC Ref's for all sub-studies? **Note:** A separate FHS016 must be submitted for each sub-study.	N/A		
Principal Investigator	Professor Liesl Zühlke		

(Note: Please complete the Closure form (FHS010) if the study is completed within the approval period)

Appendix B. Informed consent form

REQUEST FOR MOLECULAR STUDIES (DNA)

Molecular Laboratory
Division of Cardiovascular Genetics
4th Floor, Chris Barnard Building,
UCT Medical School, Observatory 7925
Tel: (021) 406 6615 Fax: (021) 447 8789

Blood should be drawn in plastic EDTA Tubes (Purple top) 20 ml of blood in total is required in adults (2-4 EDTA tubes). Please label blood tubes with patient's name and DOB. Place blood in fridge at 4 °C until able to send to laboratory.
Please DO NOT send specimens on ice or frozen

Patient details: (or hospital sticker here)

Surname:_____ First Name (S):_____
Hospital folder number:_____
Sex: M ☐ F ☐ Date of Birth (DD/MM/YYYY):_____
Patient address: _____
Contact numbers: _____
Email address: _____
Referring hospital and doctor: _____
Referring doctor's contact details: _____

Clinical information: (PLEASE COMPLETE A FAMILY PEDIGREE OVER THE PAGE)
New Family: Yes ☐ No ☐ Family name: _____
Ethnic Origin: Black African ☐ Mixed Race ☐ Caucasian ☐ Asian ☐ Other: _____
Provisional diagnosis:
Clinically affected ☐ At Risk (unaffected clinically) ☐ Spouse ☐ Query ☐

Morpho-functional phenotype	Specific diagnosis
Marfan Syndrome	☐ Confirmed clinically ☐ Suspected

Additional disorders (apparent or previously treated):_____
Have samples from this patient been send to a DNA lab before? ☐Yes ☐ No ☐ Don't Know
If yes, please specify lab:_____

For Laboratory use only:
DNA number:_____ Vol. Blood:_____ (ml) Other: _____
Date Taken: _____ Date Received (DD/MM/YYYY):_____

Informed Consent

I understand that I have agreed to participate in genetics research that will be conducted at the Cardiovascular Genetics Laboratory at the Hatter Institute, situated at the University of Cape Town in South Africa. This laboratory, in collaboration with University of Manchester genomics research, is dedicated to doing research related to genetic causes of congenital heart disease in people living in Africa.

I understand that a blood sample (5-25ml or 1-5 teaspoons of blood) or saliva in young children will be collected from me and my genetic material will be extracted from the sample for analysis. In some instances, other samples may be collected depending on the circumstances (please specify if collected): _____

I understand that a portion of my genetic material will be stored at the Cardiovascular Genetics Laboratory at the Hatter Institute, situated at the University of Cape Town in South Africa. My genetic material/information together with other relevant medical information, may be shared with other researchers and institutions involved in HREC approved studies but my personal identifying information will not be shared. I authorize that my doctors can provide relevant clinical information (medical records) to researchers.

I understand that the nature of research means that I may or may not receive a result from studies performed on my DNA. Although the laboratory will do its best to confirm that the findings relate to my condition, results received from a research laboratory should be confirmed diagnostically. If a genetic cause for my condition is found, the researchers from the Cardiovascular Genetics Laboratory will do their best to inform me of the results, either via my doctor, a genetics counsellor or in writing, depending on the available resources. If my contact details change, it is my responsibility to inform the laboratory. In the event that I am unavailable or incapacitated, I **DO / DO NOT** (*please delete where not applicable*) want my immediate family member to be informed of the results.

I understand that the Cardiovascular Genetics Laboratory is under obligation to respect my confidentiality. I understand that my participation in genetics research is entirely voluntary and that I may withdraw my consent at any time without it affecting my future medical care.

_____	_____	_____
Participant name	Signature	Date
_____	_____	_____
Doctor/nurse/genetics counsellor	Signature	Date

Study: PROTEA **HREC REF**: R017/2014 _____ (Study Number)

Appendix C. DNA extraction protocol

Note: exceptions to the protocols are written in **bold**

Gentra Puregene Blood Kit protocol – DNA purification from buffy coat

1. If the buffy coat preparation contains red blood cells, continue with step 2. Otherwise, pipet 3 ml Cell Lysis Solution into a 15 ml centrifuge tube, add 150–250 µl sample, and continue with step 8.
2. Dispense 3 volumes RBC Lysis Solution into a 15 ml centrifuge tube (e.g., if processing 250 µl buffy coat, dispense 750 µl RBC Lysis Solution). Add 150–250 µl buffy coat preparation.
3. Invert to mix and incubate for 10 min at room temperature (15–25°C). Invert again at least once during the incubation.
4. Centrifuge for 5 min at 2000 x g. **(Samples were centrifuged for 10 min)**
5. Carefully discard the supernatant by pipetting or pouring, leaving approximately 100–200 µl of the residual liquid and the pellet.
6. Vortex the tube vigorously to resuspend the pellet in the residual liquid.
7. Add 3 ml Cell Lysis Solution and pipet up and down or vortex vigorously to lyse the cells. Usually no incubation is required; however, if cell clumps are visible, incubate at 37°C until the solution is homogeneous. Samples are stable in Cell Lysis Solution for at least 2 years at room temperature.
8. Optional: If RNA-free DNA is required, add 15 µl RNase A Solution and mix by inverting 25 times. Incubate for 15 min at 37°C. Then incubate for 3 min on ice to quickly cool the sample. **(not done)**
9. Add 1 ml Protein Precipitation Solution and vortex vigorously for 20 s at high speed. **(1.5 ml Protein Precipitation Solution was added)**
10. Centrifuge for 5 min at 2000 x g. **(Samples were centrifuged for 10 min)**
11. Pipet 3 ml isopropanol into a clean 15 ml centrifuge tube and add the supernatant from the previous step by pouring carefully. **(6 ml isopropanol was used)**
12. Mix by inverting gently 50 times. **(samples were inverted until DNA was visible)**
13. Centrifuge for 3 min at 2000 x g. **(the visible DNA was removed by pipette and centrifuged for 1 min in a new 2 ml microcentrifuge tube)**
14. Carefully discard the supernatant and drain the tube by inverting on a clean piece of absorbent paper, taking care that the pellet remains in the tube.
15. Add 3 ml of 70% ethanol and invert several times to wash the DNA pellet. **(1 ml of 70% ethanol was added)**
16. Centrifuge for 1 min at 2000 x g.
17. Carefully discard the supernatant. Drain the tube on a clean piece of absorbent paper, taking care that the pellet remains in the tube. Allow to air dry for 5–10 min.
18. Add 300 µl DNA Hydration Solution and vortex for 5 s at medium speed to mix. **(200 µl DNA Hydration Solution was added)**
19. Incubate at 65°C for 1h to dissolve the DNA. **(not done)**
20. Incubate at room temperature overnight.

Appendix D. Buffers and reagents

DNA resuspension

1 M Tris-HCl (pH 8.0):

- 121.1 g Tris base (Glentham Life Sciences, Corsham, UK) in 800 ml distilled water
- Adjust the pH to 8.0 with concentrated HCl
- Make to a total volume of 1 L with sterile distilled water

0.5 M EDTA (pH 8.0):

- 186.1 g EDTA (Glentham Life Sciences) in 800 ml distilled water
- Adjust the pH to 8.0 with NaOH
- Make to a total volume of 1 L with sterile distilled water

1X Tris/EDTA (TE) buffer:

- 1 ml 1 M Tris-HCl (final concentration: 10 mM)
- 0.2 ml 0.5 M EDTA (final concentration: 1 mM)
- 98.8 ml sterile distilled water

Agarose gel electrophoresis

Tris/Borate/EDTA (TBE) buffers

10X TBE (stock):

- 121.1 g Tris (Glentham Life Sciences) (final concentration: 1 M)
- 61.8 g Boric acid (AMRESCO, Solon, OH USA) (final concentration: 1 M)
- 7.4 g EDTA (Glentham Life Sciences) (final concentration: 0.02 M)
- Made to a total volume of 1 L with sterile distilled water

1X TBE (working):

- Made by a 1:10 dilution of stock 10X TBE buffer, with sterile distilled water

Agarose gels

Loading dye:

- 1 ml 5X Green GoTaq® Flexi buffer (Promega)
- 5 µl GelRed® nucleic acid gel stain (Biotium)
- Mixed with DNA samples at a ratio of 3:5 (dye:sample) before loading

1% agarose gel:

- 1 g SeaKem® LE agarose (Lonza, Basel, Switzerland)
- 100 ml 1X TBE

DNA molecular weight marker

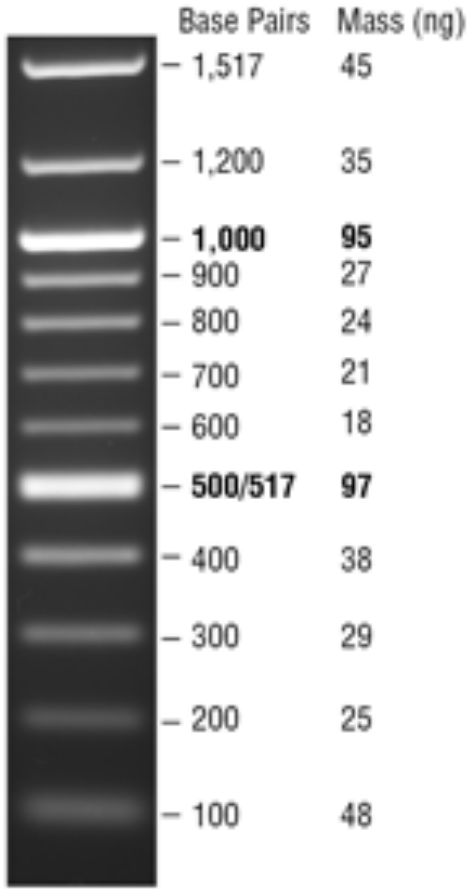

Figure A.1. The 100bp DNA molecular weight marker (New England Biolabs)

Appendix E. In-house list of CHD-associated genes

Table A.1. Cardiovascular genetics in-house list of CHD-associated genes

Gene symbols									
ABL1	*BVES*	*DISP1*	*FOXC1*	*HEY2*	*KRAS*	*NIPBL*	*PLXND1*	*SEMA3D*	*STK11*
ACTC1	*CACNA1C*	*DLL4*	*FOXC2*	*HIBCH*	*LEFTY2*	*NKX2-5*	*PQBP1*	*SEMA3E*	*STRA6*
ACVR1	*CDK13*	*DOCK6*	*FOXH1*	*HOXA1*	*MAP2K1*	*NKX2-6*	*PRDM6*	*SF3B4*	*TAB2*
ACVR2B	*CDKL5*	*EFTUD2*	*FOXL1*	*HRAS*	*MAP2K2*	*NODAL*	*PRKD1*	*SHOC2*	*TBX1*
ADAMTS10	*CFC1*	*EHMT1*	*GATA3*	*INVS*	*MED13L*	*NOTCH1*	*PTPN11*	*SHROOM3*	*TBX20*
ALDH1A2	*CHD4*	*ELN*	*GATA4*	*IRX4*	*MEIS2*	*NOTCH2*	*RAB23*	*SMAD3*	*TBX5*
ANKRD1	*CHD7*	*EVC*	*GATA5*	*JAG1*	*MESP1*	*NPHP3*	*RAD21*	*SMAD4*	*TDGF1*
ANKRD11	*CHUK*	*EVC2*	*GATA6*	*KANSL1*	*MMP21*	*NPHP4*	*RAF1*	*SMAD6*	*TFAP2B*
ARID1A	*CITED2*	*EYA1*	*GDF1*	*KAT6A*	*MYBPC3*	*NR2F2*	*RAI1*	*SMARCA4*	*TGFBR1*
ARID1B	*COX7B*	*FANCL*	*GJA1*	*KAT6B*	*MYH11*	*NRAS*	*RBM8A*	*SMARCB1*	*TLL1*
ATP7A	*CREBBP*	*FBN1*	*GJA5*	*KCNH2*	*MYH6*	*NRP1*	*RIT1*	*SMC1A*	*TMEM260*
B3GAT3	*CRELD1*	*FGFR2*	*GPC3*	*KDM6A*	*MYH7*	*NSD1*	*RPL11*	*SMC3*	*TTC37*
BCOR	*CRKL*	*FLNA*	*GPC5*	*KDR*	*MYOM2*	*PEX7*	*RPSA*	*SNX8*	*UBR1*
BMPR2	*DDX11*	*FLT4*	*HAND1*	*KMTA*	*NFATC1*	*PITX2*	*SALL1*	*SON*	*VEGFA*
BRAF	*DDX59*	*FOXA2*	*HAND2*	*KMT2S*	*NHS1*	*PLXNA2*	*SALL4*	*SOS1*	*ZFPM2*
									ZIC3

Appendix F. CN plots of samples presenting with a 22q11.21 deletion

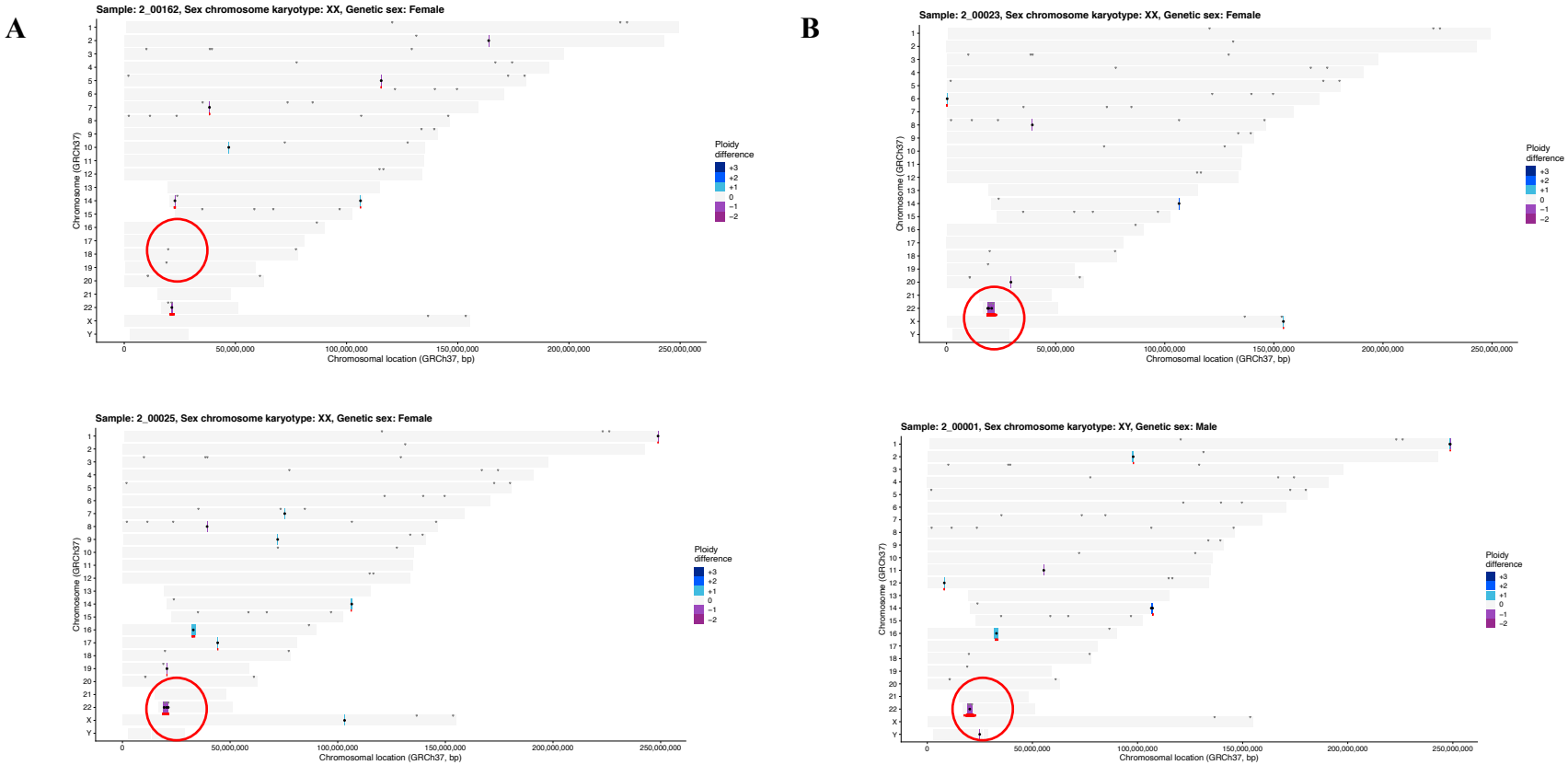

Figure A.2. CN plots of two samples with detected 22q11.21 deletions compared to two control samples clinically diagnosed with 22q11.2 deletion syndrome

A. A microdeletion of the 22q11.21 region overlapping CHD-associated genes was identified in two female probands, 2_00162 and 2_00025 diagnosed with non-syndromic VSD and TOF, respectively. B. An observed microdeletion of the 22q11.21 region in two control samples clinically diagnosed with 22q11.2 deletion syndrome. Each CNV is marked by a black circle to aid visibility and is denoted by a coloured region the size of the CNV. CNVs are coloured according to the difference from standard CN. The CNV region associated with 22q11.2 deletion syndrome is circled in red.

Appendix G. Identification of thirty-three candidate CHD-associated genes

Table A.2. Thirty-three protein-coding candidate genes identified in 89 South African non-syndromic CHD patients

Gene name	*Chromosomal co-ordinates*	*Event*	*pLI score*	*Expression in mouse embryonic heart*	*Cardiac phenotypes in mouse*	*Patient ID*	*Phenotype*
ACTR3B	7:152524844-152726381	duplication	1.0	Yes	No	2_00381	DORV + AVSD, right isomerism + PS
ADCY5	3:120032413-124764110	duplication	0.99	Yes	No	2_00132	VSD + patent foramen ovale, autism spectrum disorder
BCR	22:23258369-23650874	duplication	1.0	Yes	No	2_00198	PA + VSD, PDA, congenitally corrected TGA
C4orf3	4:120187119-120385754	duplication	0.87	Yes	No	2_00101	TA
CD86	3:120032413-124764110	duplication	0.96	Yes	No	2_00132	VSD + patent foramen ovale, autism spectrum disorder
CDKN2AIP	4:184303159-184858728	deletion	0.99	Yes	No	2_00013	AVSD + right club foot, left ptosis,
CYC1	8:145010006-145164257	duplication	0.98	Yes	No	2_00360	PDA + family history
CYFIP1	15:22770422-23282799	deletion	0.99	Yes	No	2_00192	PDA
DCP1A	3:53330701-53471338	duplication	0.94	Yes	No	2_00189	Functionally Univentricular Heart
DGCR8	22:19790229-21083936	deletion	1.0	Yes	Yes	2_00025	TOF
EXT1	8:119062671-119258258	duplication	1.0	Yes	No	2_00018	ASD
FBXO45	3:196190723-196368501	duplication	0.86	Yes	No	2_00027	TOF
FSTL1	3:120032413-124764110	duplication	0.96	Yes	Yes	2_00132	VSD + patent foramen ovale, autism spectrum disorder
GNAZ	22:23258369-23650874	duplication	0.87	Yes	No	2_00198	PA + VSD, PDA, congenitally corrected TGA
GTF2E1	3:120032413-124764110	duplication	0.90	Yes	No	2_00132	VSD + patent foramen ovale, autism spectrum disorder
JARID2	6:15283550-16036315	duplication	1.0	Yes	Yes	2_00039	DORV
KALRN	3:120032413-124764110	duplication	1.0	Yes	No	2_00132	VSD + patent foramen ovale, autism spectrum disorder
KDM2A	11:66832101-66967055	duplication	1.0	Yes	Yes	2_00036	ASD
KPNA1	3:120032413-124764110	duplication	0.99	Yes	No	2_00132	VSD + patent foramen ovale, autism spectrum disorder
MED15	22:19790229-21083936	deletion	0.96	Yes	No	2_00025	TOF

Table A.2. Thirty-three protein-coding candidate genes identified in 89 South African non-syndromic CHD patients continued…

Gene name	*Chromosomal co-ordinates*	*Event*	*pLI score*	*Expression in mouse embryonic heart*	*Cardiac phenotypes in mouse*	*Patient ID*	*Phenotype*
OSBPL11	3:124969022-126078890	duplication	0.90	Yes	No	2_00132	VSD + patent foramen ovale, autism spectrum disorder
PLCG2	16:81860283-81977626	deletion	1.0	Yes	No	2_00304	cleft mitral valve
PLXNA1	3:126200404-127237813	duplication	1.0	Yes	No	2_00132	VSD + patent foramen ovale, autism spectrum disorder
PTPLB	3:120032413-124764110	duplication	0.92	Yes	No	2_00132	VSD + patent foramen ovale, autism spectrum disorder
PUM1	1:31538533-31647947	duplication	1.0	Yes	No	2_00103	TOF
RPL9	4:39355275-39498530	duplication	0.92	Yes	No	2_00132	VSD + patent foramen ovale, autism spectrum disorder
SLC2A14	12:7976409-8118156	duplication	0.99	Yes	No	2_00046	VSD + PS
SNRNP40	1:31725962-31959820	duplication	0.83	Yes	No	2_00103	TOF
STOX2	4:184303159-184858728	deletion	0.96	Yes	No	2_00013	AVSD + right club foot, left ptosis,
STXBP5L	3:120032413-124764110	duplication	1.0	No	No	2_00132	VSD + patent foramen ovale, autism spectrum disorder
ZC3H3	8:144616532-144718976	duplication	0.96	Yes	No	2_00360	PDA + family history
ZDHHC8	22:19790229-21083936	deletion	0.99	Yes	No	2_00025	TOF
ZNF148	3:124764414-124968620	duplication	0.93	Yes	No	2_00132	VSD + patent foramen ovale, autism spectrum disorder

Note. Values are rounded off to two decimal points. Acronyms: ASD – atrial septal defect; AVSD – atrioventricular septal defect; DORV – double outlet right ventricle; ID – identifier; PA – pulmonary atresia; PDA – patent ductus arteriosus; PS – pulmonary valve stenosis; TA – tricuspid atresia; TGA – transposition of the great arteries; TOF – tetralogy of Fallot; VSD – ventricular septal defect; The pLI score shows the probability that the given gene falls into the haploinsufficiency category, and as a result id extremely intolerant of loss-of-function variation. A gene with a high pLI score (pLI ≥ 0.9) is extremely loss-of-function intolerant. A gene with a low pLI score (pLI ≤ 0.1) can tolerate loss-of-function variation.